scientific reports

OPEN

Abnormal events detection using spatio-temporal saliency descriptor and fuzzy representation analysis

R. Tino Merlin[1✉], R. Karthick[2], A. Aalan Babu[3], G. Vennira Selvi[4], D. Usha[5] & R. Nithya[6]

In recent years, the research on abnormal events detection is a significant work in surveillance video. Many researchers have been attracted by this work for the past two decades. As a result, several abnormal event detection approaches have been developed. Though several approaches have been used in the field still many problems remain to get the abnormal events detection accuracy. Moreover, many feature representations have limited capability to describe the content since several research works applied hand craft features, this type of feature can work in limited problems. To overcome this problem, this paper introduced the novel feature descriptor namely STS-D (Spatial and Temporal Saliency - Descriptor), which includes spatial and temporal information of the objects. This feature descriptor efficiently describes the shape and speed of the object. To find the anomaly score, fuzzy representation is modeled to efficiently differentiate the normal and abnormal events using fuzzy membership degree. The benchmark datasets UMN, UCSD Ped1 and Ped2 and real time roadway surveillance dataset are used to evaluate the performance of the proposed approach. Also, several existing abnormal events detection approaches are used to compare with the proposed method to evaluate the effectiveness of the proposed work.

Surveillance cameras have become essential in public spaces for safety and security, capturing human behavior[1], group activities[2], and specific targets[3–5], including humans, animals, and vehicles. These systems play a critical role in areas such as traffic monitoring[7], crime prevention[8], and crowd management[9], but identifying abnormal events within large volumes of video data is time-consuming and challenging due to the rarity and variability of such events. Abnormal events often lack a standard definition, with anomalies differing significantly from typical behaviors and requiring specific interpretations depending on the application.

Abnormal event detection is a very challenging problem in the video surveillance system, because of the fact that there is no common rule for the concept of detecting abnormal events[3,10]. An anomalous event might just be different from a normal event but not a suspicious event from the surveillance point of view[6]. Abnormal definition can be built based on three common assumptions for applying in research. They are: anomalous events infrequently occur in comparison to normal events, abnormal events have significantly different characteristics from normal events, and abnormal events are events which have a specific meaning[3]. The assumption or definition made varies with the target of the application.

In general, the surveillance camera naturally gives a large amount of video data, which is time-consuming[11] to analyze the abnormal events. Besides, most video contents are almost normal events, if one has to find any abnormal event from the surveillance video data, people who should watch almost all contents in the video. Therefore, there is a need to develop a system to find abnormal events quickly. Several researchers have focused on abnormal event detection for the past two decades. This research has been done in various ways viz. trajectory-based[12–14], pattern matching based[15–21], simulation-based[7,8] texture-based[4,5], etc. Though the research works have been applied in several directions, still some challenges needed attention by the researchers in the field of abnormal events detection. The examples of normal and abnormal events are shown in Fig. 1.

[1]Department of Computer Science and Engineering, Francis Xavier Engineering College, Tirunelveli, Tamilnadu, India. [2]Department of Computer Science and Engineering, Dr. Mahalingam College of Engineering and Technology, Pollachi, Tamilnadu, India. [3]Department of Computer Science and Engineering, School of Computing, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, Tamilnadu, India. [4]School of Computer Science and Engineering and Information Science, Presidency University, Bangalore 560064, Tamilnadu, India. [5]Department of Computer Science, Mother Teresa Women's University, Kodaikanal, India. [6]Department of Computer Science and Engineering, Bannari Amman Institute of Technology Sathyamangalam, Erode, Tamilnadu, India. ✉email: tinophd@gmail.com

Fig. 1. Examples of normal and abnormal events.

This paper considers abnormal events such as human running, vehicles appearing in pedestrian paths, people walking in the wrong direction, and heavy objects appearing in the scene. The abnormalities are analysed by introducing novel feature extraction approaches namely STS-D (Spatial and Temporal Saliency – Descriptor) and STS-C (Spatial and Temporal Saliency-Coordinates). To increase the accuracy of the abnormal events detection, *STS-D* is applied on the frame sequence to analyze appearance and speed of the objects. The STS-C is introduced to efficiently identify the moving objects' structures by analyzing global features on the frame sequence.

The proposed approach is divided into two phases: novel descriptors for video representation and build a degree of membership between normal and abnormal events using fuzzy representation. Our contributions of this paper are listed as follows:

- **STS-D (Spatial and Temporal Saliency - Descriptor)**: This novel feature descriptor incorporates spatial and temporal information to analyze the appearance and speed of objects, facilitating precise abnormal event detection.
- **STS-C (Spatial and Temporal Saliency - Coordinates)**: This descriptor identifies object structure by analyzing global features of moving and non-moving objects, enhancing recognition accuracy.
- **Fuzzy Representation**: We employ a fuzzy membership model to differentiate normal and abnormal events using the STS-D descriptor and membership degrees.

The rest of this paper is structured as follows: Sect. 2 deals with the relevant research works that have been carried out in the past on the detection of the abnormal event. The proposed video representation *STS-D is* explained in Sect. 3, fuzzy representation based abnormal events detection is explained in Sect. 4. Finally, the conclusion and future work are provided in Sect. 5.

Related works

Abnormal event detection in video surveillance relies heavily on pattern recognition to distinguish between normal and abnormal behavior. Pattern matching plays a key role in this area, identifying data attributes through machine learning algorithms based on key features. The patterns are often derived through techniques like blob detection, feature extraction, batch learning, and clustering. To find unusual events using pattern recognition clustering approaches, Duan-Yu and Po-Chung introduced the force field model[15]. Firstly, the pairs of pixels were clustered into groups based on orientations such as 45°, 90°, 135°, 180°, 225°, 270°, 315°, and 360°. Then, the corresponding pairs of clusters were compared in the consecutive frames. If the cluster size, orientation, and distance change suddenly, it is considered as an unusual event present in the scene. Yannick et al.[16] developed the pattern recognition technique using motion information. Initially, motion information was obtained from the objects and the label was assigned as '1' based on the motion information, which is denoting moving objects and '0' denoting static background. The corresponding motion label was matched with the motion label in the consecutive frame. By this action, the algorithm estimated how much mutual label transport from one frame to another frame. Based on the number of labels, the algorithm decides the selected frame whether normal event or abnormal event. Kimand Grauman[17] developed an unusual events pattern using principal component analysis by dividing the frame into spatio-temporal volumes. It is used to characterize the spatio-temporal information of spatio-temporal volumes.

To detect unusual events locally and globally, statistical information is derived on the spatio-temporal volume. Based on the information, unusual events patterns can be found. Mahadevan et al.[18] developed a normal events pattern using a mixture of dynamic texture. It is used to describe normal motion patterns. It is combined with a saliency detection method to detect unusual events. Guogang et al.[1] developed an energy model to find people running. Motion vectors are obtained by Lucas–Kanade optical flow approach for a series of images. Then, a mask is generated from the foreground objects in normal scenes. In the case of people running, the mask is not matching the objects; there may be suppressed objects entirely due to the object's force. In some occasions, this model is not apt due to intensity variation. Weiyao Lin et al.[19] created a normal activities pattern that uses the patch-based method. The input video sequence is divided into patches and blobs are extracted from each

batch. Blob is a representation of grouping similar intensity values. In the testing stage, the normal pattern blob is extracted from the images. If the overall normality likelihood value is different from the trained normal pattern, it is termed as abnormality. However, blob extraction is difficult to extract reliable blob sequence due to object occlusion in the videos. Though the pattern matching process correctly detected the abnormal events in many cases, this process cannot correctly identify the object or human moving in the wrong direction as an unusual moving object. Because the pattern does not fit if an object size highly varies.

Other approaches such as deep learning and fuzzed logic based methods are involved in object's detection and action recognition task. Davar Giveki[20] developed the human action recognition method using atanassov's intuitionistic adaptive 3D fuzzy histon roughness index texture features to describe slow and fast moving object's in the scene. To classify the actions, a fuzzy c-means clustering algorithm is applied to cluster the texture features. Then, the clustered features are classified using a fuzzy membership function. Fard et al.[21] developed based modeling for identifying human objects. Then, the fuzzy membership function is applied on the texture to connect each action of objects. Davar Giveki[22] developed a gated recurrent unit network for extracting action features. Also, optical flow gated recurrent unit network is introduced to compute the motion features and their relationship on the objects.

From the literature review, the following observations are made: (1) object recognition is essential for effective abnormal event detection; and (2) a reliable system must capture both the appearance and speed of an object's structure to detect unusual events accurately. To address these needs, this paper introduces a novel feature descriptor, STS-D, which describes object structure in terms of shape and speed, enabling efficient and accurate detection of abnormal events.

The STS-D video description methodology

This section describes how video content is represented in the frame sequence by using STS-D global descriptor. In traditional feature descriptors namely HOF[23], HOG[24], and MBH[17], the features such as gradient, direction, optical flow, and magnitude are derived using statistical measures. As a result, spatial information of each pixel is erased by aggregating features. In a real-world scenario, spatial information is important to represent the local and global activity pattern efficiently. In order to overcome the above challenge, the STS-D global descriptor is developed. To develop the STS-D descriptor, a semantic video analysis is carried out for extracting the object's shape and motion by deriving the spatial and temporal content on the moving objects. Also, STS-D can efficiently detect the foreground object from the background image. The STS-D has multiple steps such as feature extraction, STS-C formation, and foreground detection that are also discussed in the section. Finally, in order to detect the abnormal events on the frame sequence, fuzzy membership is defined based on the STS-D descriptors using object's edges and motion values. The Proposed work flow diagram is explained in the Fig. 2.

Preliminary work of STS-D

The surveillance video is taken as the input for the proposed methodology. The frame consists of foreground and background objects. To analyze the abnormal events, foreground objects only needed since foreground objects give more contribution in abnormal events. Therefore, to obtain the object appearance, edge information is extracted using gradient computation. The edge information can control the non-essential information of each
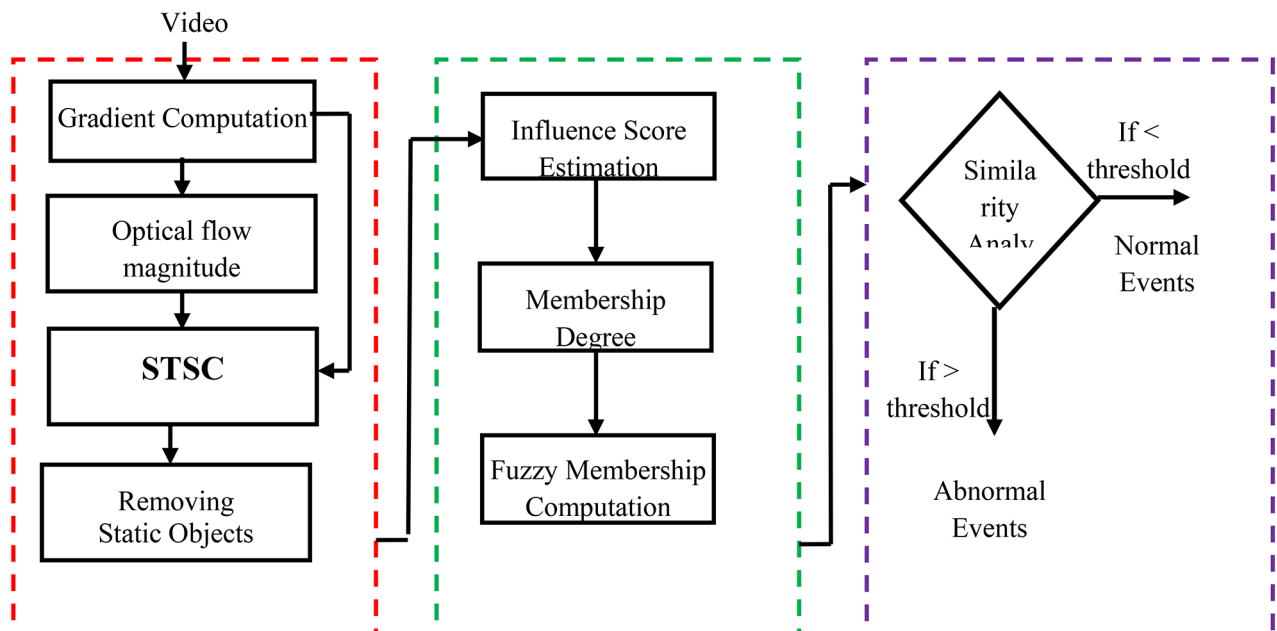

Fig. 2. The proposed work flow diagram.

object. The gradient of the frame is calculated by using the sobel filter[23,24]. This filter is applied on the frame to compute the local gradient values where the local gradient points are denoted *by g_i, i = 1, 2, 3, 4, 5, 6, 7, 8, 9.* Then, this process is repeated as a 3×3 window that is moved until it reaches the whole frame over the video. It is explained using the Eq. (1).

$$G_j = g_1, g_2, g_3, ..., g_n \qquad (1)$$

where G_j is the collection of gradient values of jth frame.

In order to achieve the motion features on the image, optical flow magnitude is extracted on the image to efficiently characterize the video content. In general, the image intensities have spatial and temporal variations at all pixels. The optical flow is estimated using gradients with x and y direction by tracking densely sampled points[20]. The pixels on the image can be assumed to have some moving objects. The object pixels are taken from each frame and then the object pixels are tracked based on displacement of objects by using optical flow technique[22–24]. In this process, consider a pixel $I(x, y, t)$ in the first frame, and it moves in *dx* and *dy* directions in the next frame after *t* time. If the pixels are the same intensity in the frames, the pixel intensity does not change where $I(x, y, t) = I(x + dx, y + dy, t + dt)$. The corresponding *t* value applies at every gradient *x* and *y* direction for obtaining velocity in *x* and *y* direction. Let v_x and v_y are the velocities in *x* and *y* direction. In order to reduce the constraint, two degrees v_x and v_y are estimated by measuring the magnitude of the optical flow field. The total magnitude at i^{th} position in the pixel is explained in the Eqs. (2) and (3).

$$m = \sqrt{v_x{}^2 + v_y{}^2} \qquad (2)$$

$$M_j = m_1, m_2, m_3, ...m_n \qquad (3)$$

where *j* is the collection of optical flow magnitude values in j^{th} frame.

Spatial and temporal saliency-coordinates (STS-C)

According to the object movement, the optical flow magnitude M_j can be correlated with the appearance of the object using gradient G_j. In case of people running or heavy objects appearing in the frames, the appearance of the objects and magnitude of the object changed simultaneously where the object may be suppressed entirely by appearance and motion due to the object's force. In this situation, the efficient features information cannot be extracted on the moving objects with its structure using the traditional feature descriptors[17,25–28]. In order to achieve the feature information with its structure, the *STS-C* is developed. The aim of the *STS-C* is to identify the appearance and motion of objects by maintaining the actual coordinates. Therefore, the multi dimensional feature sets are built using G_j and M_j with respective coordinates using ordered pairs of Cartesian products. A cartesian product of two features G_j and M_j will repeatedly be mapped for each element from two features in order to achieve the object's structure. To obtain the object structure, requires the function, it relates an input feature G_j and M_j to a multidimensional feature set. A function is an ordered triple $\langle p, R^n, G_j \rangle$ such that *p* is the function, *R* is the real number and G_j is the two dimensional input feature vector. Then, a function relates each element of a real number with exactly one element of the input feature. It denoted as $p: R^n \rightarrow G_j$ such that it can be written the following Eq. 4.

$$p(G) = \begin{bmatrix} g_{11} & g_{12} & ...... & g_{1n} \\ g_{21} & g_{22} & .... & g_{2n} \\ .... & .... & .... & ...... \\ g_{m1} & g_{n2} & .... & g_{mn} \end{bmatrix}, \qquad (4)$$

Similarly, the function is an ordered triple $\langle q, R^n, M_j \rangle$ such that *q* is the function and M_j is the two dimensional input feature vector. It is denoted as $q: R^n \rightarrow M_j$ and it is explained by using the Eq. 5.

$$q(M) = \begin{bmatrix} m_{11} & m_{12} & ...... & m_{1n} \\ m_{21} & m_{22} & .... & m_{2n} \\ .... & .... & .... & ...... \\ m_{m1} & m_{n2} & .... & m_{mn} \end{bmatrix}, \qquad (5)$$

The R^n is called the domain of *p* and *q* and G_j and M_j its co-domain of the respective functions. The STS-C is formed by computing $p \times q: R^n \times R^n \rightarrow G_j \times M_j$ such that $p \times q \ (G_j \times M_j)$, it is explained using in Eq. 6. The structure of STS-C is shown in Fig. 3.

$$STS-C = p \times q \left(\begin{bmatrix} (g_{11}, m_{11}) & (g_{12}, m_{12}) & ...... & (g_{1n}, m_{1n}) \\ (g_{21}, m_{21}) & (g_{22}, m_{22}) & ...... & (g_{2n}, m_{2n}) \\ .... & .... & .... & ...... \\ (g_{n1}, m_{n1}) & (g_{n2}, m_{n2}) & .... & (g_{nn}, m_{nn}) \end{bmatrix} \right) \qquad (6)$$

Foreground detection using STS-C

In case, describing the features of video frames directly as[29–36], the problem is that spatial and temporal information is unable to locate foreground objects accurately. Also, the foreground objects are completely destroyed by taking into account some unwanted information from the background image due to influence of background clutter, whether condition, illumination changes or occlusion. In this situation, false detection may occur. To detect the abnormality in an efficient manner, it is necessary to detect the foreground object. In
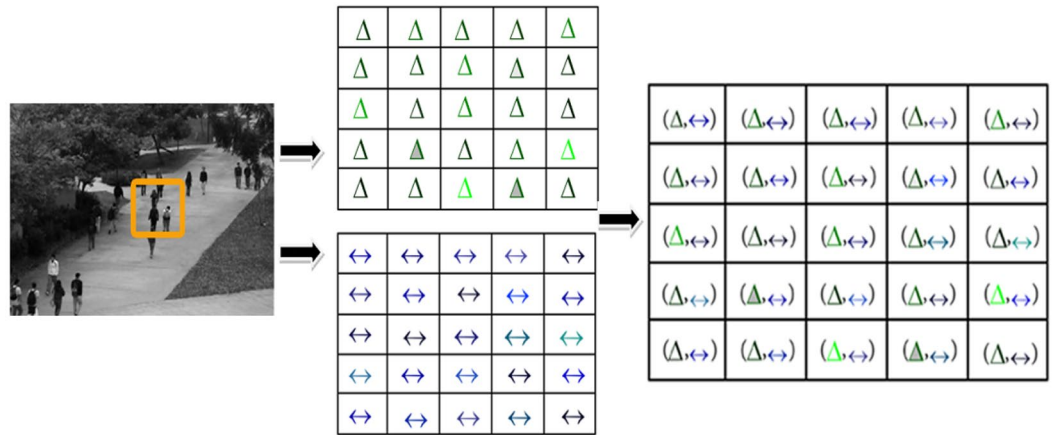
Fig. 3. An illustration of the STS-C via ordered pair of Cartesian product. (Δ denotes gradient values and ↔ denotes optical flow magnitude values where each coordinate gradient and magnitude are differentiated by different colours).

general, the feature of the foreground contributes well to detect anomalies[20]. To identify the foreground objects, the STS-C is used.

The STS-C can efficiently capture the local motion and spatial distribution on the frame. To identify the foreground object, the STS-C is analyzed between consecutive frames (f_{k+1}) and (f_k) where k is the frame number as $k = 1$ to n. By this process, the static object is obtained by connecting the set of corresponding pairs of pixels where located in the same spatial location that is indicated M_j is '0'. Then, M_j and G_j values are simultaneously removed from the respective coordinates. The remaining pair of pixels is represented as foreground objects. The resultant structure of the information is termed as STS-D global descriptor. This process is also used to reduce the memory storage.

Abnormal events detection

The proposed abnormal events detection system has two stages: (1) Influence Score Estimation, (2) Fuzzy Membership Computation.

Influence score estimation
To describe the abnormal events, the features on the objects are computed using influence score using the STS-D. There are three possible ways to determine the abnormal events: (1) the G_j and M_j of the foreground object are simultaneously influenced by external force, (2) the M_j feature is usually higher than the appearance feature due to object force, and (3) the G_j feature is higher than M_j values due to object appearance changes. On the other hand, the G_j and M_j values of the foreground object balanced to each other during normal events. In order to find the influence score, the spatio-temporal description is derived from the STS-D using the threshold values α and β. The α is applied on the G_j and the β is applied on the M_j and the influence score is found using the Eq. 7.

$$Influence\ Score = \begin{cases} 1, if\ G_j > \alpha\ and\ M_j > \beta \\ 2, if\ G_j < \alpha\ and\ M_j > \beta \\ 3, if\ G_j > \alpha\ and\ M_j < \beta \\ 0, \text{otherwise} \end{cases} \tag{7}$$

Fuzzy membership computation
In order to find the abnormal events on the frame, the influence score degree is computed using fuzzy membership function. Fuzzy sets are sets whose elements contain degree of membership[37], it is explained in the Eq. (8). Let every single frame f_k be the universal set U where $k = 1$ to n and then U is split into four categories u_r where $r = 1$ to 4. The categories are defined such as u_1, u_2, u_3, and u_4 where u_1 = Influence score as 1, u_2 = Influence score as 2, u_3 = Influence score as 3, and u_4 = Influence score as 0.

The influence scores are defined using the threshold value of the spatio-temporal features of STS-D. The range of threshold value determines the intensity of the abnormality. The optimal threshold value can be assigned by experimenting data. The threshold values are changed according to the nature of the video.

$$\mu_{u_r}^{STS-D} \rightarrow\ <0, 1> \tag{8}$$

To detect the abnormal events using fuzzy membership, firstly, the degree of influence score of u_1, u_2, u_3, and u_4 is computed and then fuzzy membership values are estimated in a frame f_k. The degree of membership value is calculated in the frame by using Eq. (9).

$$\mu_{u_k}^{STS-D}(f_k) = \frac{\left| u_k \cap STS - D_{\deg}(f_k) \right|}{\left| STS - D_{\deg}(f_k) \right|} \tag{9}$$

where $STS - D_{\deg}(f_k)$ is the total feature values included $u_1, u_2, u_3,$ and u_4. All influence scores $u_1, u_2, u_3,$ and u_4 are computed individually for each frame using the Eq. (10).

$$\mu_U^{STS-D}(f_k) = \frac{\left| u_1 \cap STS - D_{\deg}(f_k) \right|}{\left| STS - D_{\deg}(f_k) \right|}, \frac{\left| u_2 \cap STS - D_{\deg}(f_k) \right|}{\left| STS - D_{\deg}(f_k) \right|},$$

$$\frac{\left| u_3 \cap STS - D_{\deg}(f_k) \right|}{\left| STS - D_{\deg}(f_k) \right|}, \text{ and } \frac{\left| u_4 \cap STS - D_{\deg}(f_k) \right|}{\left| STS - D_{\deg}(f_k) \right|} \tag{10}$$

where u_1 and u_4 membership values are taken in a frame as abnormal and normal events respectively where u_1 exhibits high motion and appearance of the objects and u_4 exhibits less motion objects in a frame. However, u_2 and u_3 will decide whether the elements possibly belong to the u_1 or u_4 by considering the u_2 and u_3 membership values. It is explained using the Eq. 11.

$$fuzzy \text{ membership} = \begin{cases} u_1, if\ u_2\ or\ u_3 > 0.5 \\ \\ u_4, if\ u_2\ or\ u_3 < 0.5 \end{cases} \tag{11}$$

The u_1 and u_4 membership values decide the content of the frame f_k whether the normal or abnormal events.

Experiments
Experimental setup and evaluation
The proposed work was carried out using MATLAB 2013a software by comparing several existing anomaly detection approaches to insight the performance. The experiments are carried out on two benchmark datasets such as UMN[38], UCSD ped1 and ped2[39] and, real time dataset.

Dataset description
UMN dataset The UMN dataset is widely applied dataset in anomaly detection that contains normal human activities such as persons with normal walking and abnormal crowd activities such as people escaping, people gathering, and people running from the crowd in different directions. This dataset has 7740 frames with 240×320 resolution in which 1450, 4415, and 2145 frames for scene 1, 2, and 3 respectively. In each scene, the first 400 frames were used for training and the remaining frames.

UCSD dataset UCSD ped1 and ped2 pedestrian datasets are the most frequently-used datasets in video anomaly detection. UCSD ped1 and ped2 datasets contain both crowded and uncrowded scenes and the scenes were captured by different environments. Also, the scene includes the following actions such as human walking, skaters, bikers and vehicles. UCSD ped1 dataset contains 34 and 36 training and testing videos respectively where each contains 200 frames with a resolution of 158×238. UCSD ped2 dataset contains 16 and 12 testing videos respectively where each video contains 120 to 180 with a resolution 240×360.

Real-time dataset The dataset consists of five scenes that contain normal human activities such as persons with normal walking and abnormal running such as people escaping and people running on the roadway. This dataset has 1000 frames where each scene has 200 frames.

Performance evaluation
To evaluate the performance of the proposed method, this paper used True Positive (TP), False Positive (FP), False Negative (FN), True Positive Rate (TPS) and False Positive Rate (FPR) metrics. The experimental results are evaluated with the frame level ground truth annotation of each frame. The following measures are used such as Precision, Recall, F1-score, Receiver Operating Characteristic (ROC) curve of TPR versus FPR to estimate the performance. They are explained in the Eqs. 12, 13, 14, 15 and 16.

$$Precision = \frac{TP}{TP + FP}, \tag{12}$$

$$Recall = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{13}$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precesion\ +\ Recall}, \tag{14}$$

$$TPR = \frac{Number \text{ of True Possitve Frames}}{\text{Number of Possitive Frames}} \tag{15}$$

Scene No.	Precision (%)	Recall (%)	F1-score (%)
Scene 1	95.00	80.00	86.85
Scene 2	93.00	79.00	85.43
Scene 3	89.00	70.00	78.36

Table 1. Experimental results on UMN dataset.

Sequence No.	Frames	UCSD Ped1			UCSD Ped2		
		Precision	Recall	F1-score	Precision	Recall	F1-score
1	200	93.00	52.00	66.70	91.00	59.00	67.53
2	200	87.00	61.00	72.00	90.00	54.00	67.50
3	200	93.00	70.00	80.00	91.00	71.00	80.00
4	200	90.00	79.00	84.14	92.00	75.00	82.63
5	200	87.00	80.00	83.35	90.00	83.00	86.35

Table 2. Experimental results on UCSD Ped1 and Ped2 dataset

Methods	AUC
Optical Flow computation	84.00%
SF	94.00%
MDT	93.00%
Chaotic invariants	95.04%
Biswas	94.00%
GLCM	95.09%
OPLKT-EMEHO	96.00%
RpNet	95.00%
Proposed Method	97.00%

Table 3. Comparative analysis on UMN dataset.

Scene No.	Precision (%)	Recall (%)	F1-score (%)
Scene 1	70.00	79.00	74.00
Scene 2	83.00	81.00	81.00
Scene 3	84.00	80.00	81.00
Scene 4	80.00	82.00	80.00
Scene 5	86.00	87.00	86.00

Table 4. Experimental results on real-time dataset.

$$FPR = \frac{Number \text{ of False Possitve Frames}}{\text{Number of Negative Frames}} \qquad (16)$$

where TP indicates correctly detected anomaly frames where two frames matched between ground truth and the detected frame, FP indicates incorrectly detects the frame where two frames not matched between ground truth and the detected frame, and FN indicates the incorrectly detected normal event frame where the abnormal events detected frame not matched with ground truth frame. The results are explained by Tables 1, 2, 3 and 4; Fig. 4 and 5.

The threshold values are needed to efficiently differentiate normal and unusual events for both UCSD Ped1 and Ped2 and UMN while evaluating all measures. The proposed method when identifying an abnormal events frame if it is in the annotated range, the detected frame is counted as a correctly identified frame for TPR. On the other hand, the detected abnormal events frame if it is in the outside annotated range, the detected frame is counted as a wrongly detected abnormal event frame for FPR.

Table 1 shows the performance of proposed work on UMN using F1-Score where scene 3 obtains a lower F1-score compare than other scenes. Though the scene 3 consists of abnormal running of persons the abnormal speed of the persons is low. Therefore, the proposed method considers the abnormal events as normal events since the fuzzy membership degree belongs to normal events. On the other hand, sequence 1 and 2 achieved a

Fig. 4. The visual results of proposed method.

Fig. 5. The resultant real-time abnormal events frames.

low F1-score than other sequence in UCSD Ped1 and Ped2. The UCSD Ped1 and Ped2 start with a normal person walking on the pedestrian path and it is followed by an irregular cyclist moving. Though irregular cyclists in the pedestrian path are abnormal events, the proposed work is considered as the normal events since the cyclist is too small in size as the person. Therefore, the influence score was achieved as '0' by the gradient and magnitude of the cyclist and thus fuzzy membership degree obtained as '0'. Therefore, the algorithm mistakenly decided irregular cyclist abnormal frames as the normal events frame. The results are shown in Table 2.

Comparative analysis
The UMN and UCSD Ped1 and Ped2 are used to compare with proposed work to evaluate the effectiveness of the proposed work since normal and abnormal frequently occur in all the sequences for both datasets. The

following existing abnormal events detection algorithms such as optical flow computation[40], SF[41], MDT[18], Chaotic invariants[42] Biswas[43], GLCM[44], OPLKT-EMEHO[45], and rpNet[46] are used to compare with the proposed work using UMN dataset.

Table 4 summarizes the experimental results on a real-time dataset. Scene 1 achieves a Precision of 70%, Recall of 79%, and an F1-score of 74%, indicating a moderate balance between detecting relevant instances and minimizing false positives. Scene 2 shows improved performance with 83% Precision, 81% Recall, and an F1-score of 81%, while Scene 3 maintains similar effectiveness with 84% Precision, 80% Recall, and an F1-score of 81%. These results collectively indicate consistent, high performance across scenes, with Scene 5 demonstrating the model's optimal effectiveness. Table 5 provides a quantitative analysis of abnormal event detection on a real-time dataset across five scenes. For each scene, the table displays the total number of abnormal events, alongside the number of events that were correctly and incorrectly detected by the model.

Figure 4 illustrates the performance of the proposed method on the UMN and UCSD Ped1 and Ped2 datasets through a comparative visual representation. The figure is organized into three main sections for each dataset: Normal Frames, Abnormal Frames, and Incorrectly Detected Normal Frames as Abnormal Frames. The Normal Frames column displays frames where typical activities occur, which the model correctly identifies as normal. The Abnormal Frames column includes scenes with unusual events that the model successfully detects as abnormal, such as people running in the UMN dataset and vehicles moving along pedestrian paths in the UCSD dataset. Lastly, the Incorrectly Detected Normal Frames as Abnormal Frames column illustrates instances where the proposed method misclassifies typical activities as anomalies. For the UMN dataset, this includes scenarios such as normal walking activities that were flagged as abnormal due to subtle variations in crowd density or motion. In the UCSD Ped1 and Ped2 datasets, misclassifications often occur in scenes where small movements or minor irregularities—such as slight deviations in pedestrian flow—are mistaken for abnormal events. These false positives highlight the model's sensitivity to minor deviations, reflecting its challenge in accurately distinguishing low-level variations in typical behavior from genuinely unusual events. Figure 5 illustrates the proposed method's performance on a real-time dataset by showcasing both correctly and incorrectly detected events. In the Correctly Detected Abnormal Events column, the frames highlight instances where the model successfully identifies genuine anomalies, such as unusual object appearances. Conversely, the Incorrectly Detected Normal Events as Abnormal Events column presents frames depicting normal activities, like standard pedestrian flow, that are mistakenly classified as abnormal. These false positives indicate the model's tendency to misinterpret certain normal activities under specific conditions, possibly due to variations in lighting.

Tables 3 and 6 show the performance on the UMN and UCSD dataset which consists of comparison results between proposed methods and existing abnormal events detection methods using AUC and EER. From these analyses, the results revealed that the proposed method obtained better results for both UMN and UCSD Ped1 and Ped2 since the video frames are applied to capture the gradient and magnitude of the foreground objects with the help of STS-D global descriptors as well as compute fuzzy membership degree. Therefore, the proposed method correctly detected the abnormal events in many frames where people were running, vehicles in pedestrian paths, carts, and cars. In the case of people running with low speed, the descriptor is not working well since the influence score will be less to classify the abnormal event. In some occasions, this descriptor is not clearly described the nature of the content due to intensity variation. In real-time dataset, the lighting variation is highly varies between frames. Therefore, the performance degrades due to insufficient details of objects. Because, the objects edges are highly needed to describe the objects appearance and motion. Therefore, the comparative results are high for benchmark dataset when compared real-time dataset.

The performance is also affected by lighting variations and intensity changes within the frames of the real-time dataset. The descriptor occasionally fails to accurately capture the nature of the content when lighting conditions fluctuate significantly between frames, leading to insufficient detail in object edges, which are essential for identifying movement and appearance. This issue is particularly evident in scenes with rapidly changing lighting conditions.

Conclusions

This paper introduced the novel STS-D global feature descriptor to describe the spatial and temporal information which can efficiently describe the shape and speed of the foreground objects. To identify the moving object's structure by STS-D, the STS-C was derived on the foreground objects to differentiate moving and non-moving foreground objects in the frame sequence. Also, fuzzy representation was modelled to efficiently differentiate the normal and abnormal events using fuzzy membership degree. The experimental results on two benchmark datasets revealed the effectiveness of the proposed work. Although the proposed method gives better performance in all abnormal events cases the descriptor does not give proper information for the small abnormal objects. Therefore, the proposed method could not identify the abnormal events if any small size abnormal objects were

Scene No.	No of Abnormal Events	Correctly Detected	Incorrectly Detected
Scene 1	5	3	2
Scene 2	7	6	1
Scene 3	6	6	-
Scene 4	3	2	1
Scene 5	5	4	1

Table 5. Quantitative analysis on real-time dataset.

Methods	UCSD (Ped1)		UCSD (Ped2)	
	AUC	EER	AUC	EER
Adam et al. [47]	77.10%	40.01	-	40.11%
Kim et al. [17]	59.00	-	68.03%	-
Mahran et al. [25]	67.13%	29.11%	55.10%	40.20%
Mahadevan et al. [18]	76.10%	32.42%	59.00%	36.00%
Wang et al. [48]	72.20%	33.22%	85.46%	18.00%
Xu et al. [49]	92.28%	16.10%	92.00%	15.10%
Hasan et al. [50]	81.00%	26.00%	91.10%	20. 90%
Lonescu et al. [51]	70.00	-	81.00%	-
Chong et al. [52]	87.00%	11.00%	85.00%	11.00%
Liu et al. [53]	82.00	-	94.00%	-
Wang et al. [54]	77.00%	31.25%	96.00%	7.00%
Song et al. [55]	91.40%	14.00%	89.33%	14.00%
Rajasekaran and Raja Sekar [45]	91.00%	14.00%	90.00%	15.00%
Haidar et al. [46]	88.00%	13.00%	93.00%	11.00%
Proposed Method	94.00%	11.00%	96.00%	10.00%

Table 6. Comparative analysis on UCSD dataset.

involved in the frame. Therefore, the efficient feature is still needed to describe the shape and speed of the small size objects. The proposed method can be extended in various video analyses in the future.

Data availability
The datasets used in this research is publically available.

References
1. Xiong, G. et al. An energy model approach to people counting for abnormal crowd behaviour detection, Elsevier, *Neurocomputing*. **83**, 121–135 (2011).
2. Xu, D. et al. Video Anomaly Detection Based on a Hierarchical Activity Discovery within spatio -temporal Contexts. (Elsevier, *Neurocomputing*, **143**, 144–152 2014).
3. Angela, A., Sodemann, M. P., Ross & Borghetti, B. J. A review of anomaly detection in automated surveillance. *IEEE Trans. on Systems. Man, and Cybernetics*. **42**, 1257–1271 (2012).
4. Yuan, Y., Wang, D. & Wang, Q. Anomaly detection in traffic scenes via spatial-aware motion reconstruction. *IEEE Trans. on Intelligent Transportation Systems*.18, 1–12 (2016).
5. Kaelon Lloyd, P. L., Rosin, D., Marshall, Simon, C. & Moore March, Detecting violent and abnormal crowd activity using temporal analysis of grey level co-occurrence matrix (GLCM)-based texture measures. *Springer, Machine Vision and Applications*. 28,1–11 (2017).
6. Colque, R. M., Caetano, C., Toledo, M. & Schwartz, W. R. Histograms of optical flow organizations and magnitude and entropy to detect anomalous events in videos. *IEEE Trans. Circuits Syst. Video Technol*. **27**, 1–10 (2016).
7. Vallejo, D., Albusac, J., Jimenez, L., Gonzalez, C. & Moreno, J. A cognitive surveillance system for detecting incorrect traffic behaviours. *Elsevier, Experts System with Applications*. **36**, 10503–10511 (2009).
8. Hao, Y. et al. A graphical simulator for modelling complex crowd behaviors. *2018 22nd International Conference Information Visualisation*. 1–6 (2018).
9. Wilbert, G. et al. Statistical abnormal crowd behaviour detection and simulation for real-time applications. *Springer, 10th International Conference*. **10463**, 671–682 (2017).
10. Chu, W., Xue, H., Yao, C. & Cai, D. Sparse coding guided spatiotemporal feature learning for abnormal event detection in large videos. *IEEE Trans. Multimedia*. **14**, 1–14 (2017).
11. Vennila, T. J. & Balamurugan, V. A stochastic framework for keyframe extraction. *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, 1–5 (2020).
12. Fozia Mehboob, M. et al. Trajectory based vehicle counting and anomalous event visualization in smart cities. *Springer, Cluster Comput*. 21,1–10 (2017).
13. Jiang, F., Yuan, J., Sotirios, A., Tsaftaris, Aggelos, K. & Katsaggelos Anomalous video event detection using spatiotemporal context. *Elsevier, Computer Vision and Image Understanding*. **115**, 323–333 (2010).
14. Asha Paul, M. K., Kavitha, J. & Jansi Rani, P. A. Keyframe extraction techniques: a review. *Recent. Pat. Comput. Sci.* **11** (1), 3–16 (2018).
15. Huang, P. C. Duan-Yu Chen and Motion - based unusual event detection in human crowds. *Elsevier, J. Vis. Commun. Image Representation*. **22**, 178–186 (2011).
16. Yannick Benezeth, P. M., Jodoi & Saligrama, V. Abnormality detection using low-level co-occurring events. *Elsevier, Pattern Recognition Letters*. **32**, 423–431 (2011).
17. Kim, J. K.Grauman, Observe locally, infer globally: a space-time mrf for detecting abnormal activities with incremental updates, in: IEEE Conference on Computer Vision and Pattern Recognition, pp.2921–2928 (2009).
18. Mahadevan, V. & Bhalodia, L. W. V. N.Vasconcelos, Anomaly detection in crowded scenes, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.1975–1981 (2010).
19. Lin, W. et al. Summarizing surveillance videos with local – patch – learning - based abnormality detection, blob sequence optimization, and type - based synopsis. *Elsevier Neurocomputing*. **155**, 84–98 (2015).

10

20. Fard, M. G., Montazer, G. A. & Giveki, D. A Novel Fuzzy Logic-based Method for Modeling and Recognizing Yoga Pose, 2023 9th International Conference on Web Research (ICWR), Tehran, Iran, Islamic Republic of, pp. 1–6 (2023).

21. Giveki, D. Robust moving object detection based on fusing Atanassov's intuitionistic 3D fuzzy histon roughness index and texture features. *Int. J. Approximate Reasoning.* **135**, 1–20 (2021).

22. Giveki, D. Human action recognition using an optical flow-gated recurrent neural network. *Int. J. Multimedia Inform. Retr.* 13,1–18 (2024).

23. Song, M., Tao, D. & Maybank, S. J. Sparse camera network for visual surveillance— a comprehensive survey. *ar Xiv preprint.* **arXiv**, 1302.0446 (2013).

24. Vennila, T. J. & Balamurugan, V. A Rough Set Framework for Multihuman Tracking in Surveillance Video. in *IEEE Sens. J..* **23** (8), 8753–8760 (2023).

25. Mehran, R. A.Oyama, M.Shah, Abnormal crowd behaviour detection using social force model, in: IEEE Conference on Computer Vision and Pattern Recognition, pp.935–942 (2009).

26. Wang, S., Zhu, E., Yin, J. & Porikli, F. Video anomaly detection and localization by local motion based joint video representation and OCELM. *Elsevier, Neurocomputing.* **277**, 161–175 (2018).

27. Yang Xian, X., Rong, X., Yang & Tian, Y. Evaluation of low-level features for real-world surveillance detection. *IEEE Trans. Circuits Syst. Video Technol.* **27**, 1–11 (2017).

28. Navneet & Dalal Bill Triggs, and Cordelia Schmid Human detection using oriented histograms of flow and appearance. *European Conference on Computer Vision (ECCV '06), Graz, Austria.* 428–441 (2006).

29. Wang, H., Klser, A., Schmid, C. & Liu, C. L. Action recognition by dense trajectory. *CVPR*, (2011).

30. Chen, C. & Shao, Y. Anomalous crowd behavior detection and localization in video surveillance. *2014 IEEE International Conference on Control Science and Systems Engineering, Yantai.* 190–194 (2014).

31. Roberto Leyva, V., Sanchez & Li, C. T. Video anomaly detection with compact feature sets for online performance. *IEEE Trans. on image processing.* **26**, 3463–3478 (2017).

32. Asha Paul, M., Sampath Kumar, K., Sagar, S. & Sreeji, S. LWDS: lightweight DeepSeagrass technique for classifying seagrass from underwater images. *Environ. Monit. Assess.* **195** (5), 1–11 (2023).

33. Dheeraj Kumar, J. C., Bezdek, S., Rajasegarar, C., Leckie & Palaniswami, M. *A visual-numeric Approach to Clustering and Anomaly Detection for Trajectory data* pp. 265–281 (Springer, Dec. 2015).

34. Tian, W. A. N. G., Qiao, M., Zhu, A., Shan, G. & Snoussi, H. Abnormal event detection via the analysis of multi-frame optical flow information. **14**, 304–313 (2019).

35. Zhang, Y., Lu, H., Zhang, L. & Xiang, R. Combining motion and appearance cues for anomaly detection. *Elsevier Pattern Recognit.* **51**, 443–452 (2016).

36. Gall, J., Rosenhahn, B. & Seidel, H. Drift-free tracking of rigid and articulated objects, 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, pp. 1–8, (2008).

37. Sasikumar, R. & Sheik Abdullah, A. Stock market forecasting using Time invariant, fuzzy Time Series Model. *Res. Reviews: J. Stat.* **7** (1), 104s–111sp (2018).

38. https://paperswithcode.com/dataset/crowd11

39. http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm

40. Ryan, D., Denman, S., Fookes, C., Clinton, B. & Sridharan, S. Textures of optical flow for real-time anomaly detection in crowds. In: 2011 8th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 230–235 Doi: (2011). https://doi.org/10.1109/AVSS.2011.6027327

41. Mehran, R., Oyama, A. & Shah, M. Abnormal crowd behaviour detection using social force model. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 935–942 (2009).

42. Wu, S., Moore, B. E. & Shah, M. Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes. In: Proceedings of the IEEE Computer Society Conference on Com Vision and Pattern Recognition, pp. 2054–2060 (2010).

43. Biswas, S. & Gupta, V. Abnormality detection in crowd videos by tracking sparse components. *Mach. Vis. Appl.* **28**, 35–48 (2016). https://doi.org/10.1007/s00138-016-0800-8

44. Lloyd, K. et al. Detecting violent and abnormal crowd activity using temporal analysis of grey level co-occurrence matrix (GLCM)-based texture measures. *Mach. Vis. Appl.* **28**, 361–371 (2017). 2017.

45. Rajasekaran, G. & Sekar, J. R. *Abnormal Crowd Behavior Detection Using Optimized Pyramidal Lucas-Kanade Technique* pp. 2439–2412 (Intelligent Automation & Soft Computing, 2023).

46. Md, H., Sharif, L., Jiao & Omlin, C. W. Deep Crowd Anomaly Detection by Fusing Reconstruction and Prediction Networks, MDPI, electronics, pp. 1–41, (2023).

47. Adam, E., Rivlin, I., Shimshoni & Reinitz, D. Robust real-time unusual event detection using multiple fixed-location monitors. *IEEE Trans. Pattern Anal. Mach. Intell.* **30** (3), 555 (2008).

48. Wang, T. & Snoussi, H. Histograms of optical flow orientation for abnormal events detection, in IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS). IEEE, pp. 45–52 (2013).

49. Xu, Y., Yan, E., Ricci & Sebe, N. Detecting anomalous events in videos by learning deep representations of appearance and motion. *Comput. Vis. Image Underst.* **156**, 117–127 (2016).

50. Hasan, J., Choi, J., Neumann, A. K., Roychowdhury & Davis, L. S. Learning temporal regularity in video sequences, in IEEE Conference on Computer Vision and Pattern Recognition, pp. 733–742 (2016).

51. Ionescu, R. T., Smeureanu, S., Alexe, B. & Popescu, M. Unmasking the abnormal events in video, in IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22–29, 2017, pp. 2914–2922 (2017).

52. Chong, Y. S. & Tay, Y. H. Abnormal event detection in videos using spatiotemporal autoencoder. pp. 189–196, (2017).

53. Liu, W., Luo, W., Lian, D., Gao, S. & Recognition IEEE Conference on Computer Vision and Pattern Future frame prediction for anomaly detection - A new baseline, in, CVPR 2018, Salt Lake City, 2018. (2018).

54. Wang, S. et al. Detecting abnormality without knowing normality: A two-stage approach for unsupervised video abnormal event detection, in ACM Multimedia Conference on Multimedia Conference, MM 2018, Seoul, Republic of Korea, October 22–26, 2018, 2018, pp. 636–644. (2018).

55. Song, H., Sun, C., Wu, X., Chen, M. & Jia, Y. Learning Normal Patterns via Adversarial Attention-Based Autoencoder for Abnormal Event Detection in Videos, in *IEEE Transactions on Multimedia.* **22** (8), 2138–2148 (2020). https://doi.org/10.1109/TMM.2019.2950530

Author contributions

Funding
Not Applicable.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to R.T.M.

Reprints and permissions information is available at www.nature.com/reprints.

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