Classification of human activity detection based on an intelligent regression model in video sequences

Natarajan Kumaran¹ | Uyyala Srinivasulu Reddy²

¹ Research Scholar, Department of Computer Applications, National Institute of Technology, Trichy, Tamil Nadu, India
² Assistant Professor, Department of Computer Applications, National Institute of Technology, Trichy, Tamil Nadu, India

Correspondence
Natarajan Kumaran, Research Scholar, Department of Computer Applications, National Institute of Technology, Trichy, Tamil Nadu 620015, India.
Email: kmshre19@gmail.com

Abstract
The most critical objective in security surveillance is abnormal event detection in public scenarios. A scheme is presented for detecting abnormal behaviours in the activities of human groups based on social behaviour analysis. This approach efficiently models group activities than some of the previous strategies that use independent local features. This paper presents a feature descriptor method to signify the movement by implementing the optical flow through covariance matrix coding. The multi-RoI (region of interest) covariance matrix has some frames or patches which could represent the movement in high accuracy. Normal samples are plentiful in public surveillance videos, while there are only a few abnormal samples. For that, the model of a hybridised optical flow covariance matrix is represented in this paper. Optical flow (OF) in the temporal domain is measured as a critical feature of video streams. The logistic regression method is used to detect abnormal activities in a crowded scene. Finally, the behaviours of human crowds can be predicted using benchmark datasets such as UMN, UCSD as well as BEHAVE. The obtained experimental results show that the proposed approach can effectively detect abnormal events from the abandoned environment of surveillance videos.

1 | INTRODUCTION

Human action recognition has drawn an expanding consideration, and its applications are relevant in the extreme mechanism of video surveillance and security, video annotation and recovery, behavioural biometrics and human–computer connection, and so forth [1]. Vision-based human activity recognition is the way towards perceiving human activities in video sequences by using computer vision strategies. The term activity refers to basic movement patterns normally executed by solitary individuals and ordinarily extend for a small duration of time [2]. The objective of action recognition is to investigate progressing actions from an unknown video, consequently [3]. Accordingly, one of an adequate obstacle admired among human action recognition originate from tremendous conflicts such as great intra-class difference, scaling, impediment, and clutter. Video-based human activity detection has numerous applications in human–computer collaboration, observation, video ordering, and recovery. Activities or movements create some of the patterns which are fluctuating in terms of spatio-temporal forms in recordings, which could be utilised as feature descriptors for the detection of actions [4, 5].

In the earlier developments, action recognition has included utilising video successions caught by camcorders. Sequentially spatio-temporal highlights are broadly utilised for perceiving human actions and contrasted with some of the images which are conventional. In lighting conditions, the depth maps are very insensitive to changes and can give 3D data for recognising activities that are hard to portray the utilisation of traditional images [6, 7]. Space-time based strategies, for example, the highlights of local spatio-temporal are famous systems for video representation also demonstrates promising execution in action recognition. The intensity-based video images, which restrict the intensity of detecting an action [8]. Convolutional Neural Networks (CNNs) is a sort of deep models in which trainable channels and nearby neighbourhood pooling tasks are connected alternatingly on the raw input images, bringing about a chain of command of progressively complex highlights [9]. When prepared with suitable regularisation, CNNs can accomplish superior execution on visual object recognition tasks.
Furthermore, CNNs have been appeared to be invariant to specific varieties, for example, posture, lighting, and encompassing clutter. As a class of deep models for highlight development, CNNs have been applied to 2D pictures \cite{10, 11}. In the most recent decade, the progressions in computational capacities and the introduction of powerful methodologies with preparing deep neural system designs need to prompt their wide use to address different computer vision challenges \cite{12}.

The convolutional kernels mutually in space and time demonstrate with the 3D CNN because of a direct expansion of the well-known 2D CNN profound structures to the 3D spatio-temporal domain \cite{13}. The techniques go for adapting long-extended movement by taking in a chain of command comprising of different layers of 3D spatio-temporal convolution parts by early fusion, late fusion, or slow fusion. The CNN architecture of two-stream functions acquires movement patterns utilizing an extra CNN, and input given is the optical flow, which has been obtained from the progressive frames of video sequence \cite{14}.

In order to detect the abnormalities in the behaviour of the huge body with normal data, it is relevant to enhance the utilisation of normal data rather than the acknowledgment \cite{15}. Automatically detecting abnormal exercises or occasions from long length video sequences is important for intelligent surveillance, behaviour analysis, and security applications. Specifically, abnormal behaviour detection in queued scenes is a problem because of the substantial number of people on foot nearby, the instability of individual appearance, the continuous incomplete impediments that deliver, and the irregular motion pattern of the group \cite{16, 17}. Moreover, there are potential risky exercises in swarmed conditions, for example, crowd panic, stampedes, and accidents, including an extensive number of people, which makes computerised scene investigation in the most need \cite{18}.

Among numerous video detection works, identifying anomalous activities in video streams is fundamental significance as it is identified with certain topics in computer vision, which are fascinating, for example, visual saliency, imperative behaviour recognition, and interestingness prediction \cite{19}.

Tragically, in the video sequences, the anomalous action recognition is a troublesome test because of the instability of the classifications mutually for normality and abnormality. An illustrative method to address the detection problems of irregularity is to separate local features. After that handle, an exception identification errand where typical exercises in video scenes are displayed, also abnormal occasions are distinguished as fundamentally different as of the model \cite{20}. In a public scene concerning social security, the detection of an abnormal event is important. The public scene could be considered a significant place where common and daily activities often occur. In the public scene, security surveillance concerns are connected to livelihood subjects. The security guards check the surveillance videos, and for them, it is very simple to visual and intellectual fatigue. Hence, it is not possible to perform rapid action towards unpredicted or abnormal events. Because in public scenarios, most monitors are placed for safety purposes, the automatically set alarm system helps the security guard to observe the events which are in an emergency from a large database of videos, which is critical. The optical flow, as well as the consequent partial derivatives, are fused with the help of the covariance matrix descriptor and are projected into a lower range of dimensional feature space.

In the temporal domain, optical flow (OF) is measured using the Lucas Kanade method, and it is considered as a critical feature of video streams. The covariance is classified using logistic regression, and it probably patterns the dealings of the optical flow in the ROI. Although, as a linear method, the logistic regression is considered in which the predictions are renovated through logistic function. The abnormal event is detected using logistic regression with the proposed covariance matrix descriptor. Experimentations conducted in the standard datasets confirm that the proposed model does better than the existing methods. This research article designs the model of anomalous activity issues present in the complicated public scenarios, and advanced technologies are presented for the intelligent security surveillance system. They are remaining part of the article pre-arranged, as listed below.

In Section 2, the related works are reviewed. Section 3 explains the proposed method with the subsections mentioned below. In Section 3.1, foreground detection by using the Gaussian mixture model (GMM) in which each pixel is patterned as a mixture of Gaussian. The classification is based on the mean and variance of both foreground and back ground. In Section 3.2, for optical flow estimation, the Lucas kanade approach is proposed, and in Section 3.3 the motion descriptor presents the movement details, Section 3.4 introduces the logistic regression analysis for abnormal behaviour detection with a stochastic gradient descent algorithm for cost optimisation is discussed. In Section 4, the experimental evaluations conducted in the datasets of BEHAVE, UMN, and UCSD are explained. At last, Section 5 concludes with the prospect of future work.

2 | RELATED WORKS

The humans action recognition has been applied in some domains like human–computer communication, game control, and smart surveillance.

Liu et al. (2017) \cite{21} proposed an improved visualisation strategy of skeleton meant for human activity detection based on the view-invariant sequence.

At first, based on the sequences, a view-invariant change was made to devastate the impact of skeleton joints view variations from the spatio-temporal areas. Next, the skeletons changed were visualised as a sequence of colour images. Third, a model based on the convolutional neural system for colour images was received to isolate the strong as well as discriminative highlights. Using the choice level fusion of deep features, the final action class scores were produced. Ten CNNs hard determination and logistic regression with the proposed covariance matrix descriptor was performed, and in Section 3.3 the motion descriptor presents the movement details, Section 3.4 introduces the logistic regression analysis for abnormal behaviour detection with a stochastic gradient descent algorithm for cost optimisation is discussed. In Section 4, the experimental evaluations conducted in the datasets of BEHAVE, UMN, and UCSD are explained. At last, Section 5 concludes with the prospect of future work.

The optical flow, as well as the consequent partial derivatives, are fused with the help of the covariance matrix descriptor and are projected into a lower range of dimensional feature space.

In the temporal domain, optical flow (OF) is measured using the Lucas Kanade method, and it is considered as a critical feature of video streams. The covariance is classified using logistic regression, and it probably patterns the dealings of the optical flow in the ROI. Although, as a linear method, the logistic regression is considered in which the predictions are renovated through logistic function. The abnormal event is detected using logistic regression with the proposed covariance matrix descriptor. Experimentations conducted in the standard datasets confirm that the proposed model does better than the existing methods. This research article designs the model of anomalous activity issues present in the complicated public scenarios, and advanced technologies are presented for the intelligent security surveillance system. They are remaining part of the article pre-arranged, as listed below.
approximately a gain of 10%. The effectiveness of the scheme was shown by this result associated with the modern LSTM-based approaches.

In human behaviour subjects, the investigation means to comprehend the subject behaviour over time utilising the motion information. Ijjina et al. (2017) [22] specified an approach for perceiving human action because of RGB-D video based on the motion sequence information utilised deep learning. This technique does not perceive human behaviour, like hand movements and activities of a group. Multi-modal information usage, as well as the ConvNet features noise tolerance, provided the strength also the compliance of this method compared with other detection tasks. A combination of facts among several classes implied that best actions could be accomplished through the fusing facts across models highlighting certain temporal areas. Then the prevailing methods, the proposed approach is faster owing to the un-complicated calculation in figuring out the new representation (which could be parallelised) as well as the implementation is correspondently done for the ConvNet feature extraction. This approach’s efficiency was verified on SBU Kinect interaction, MINIA action NATOPS gesture, and Weizmann datasets.

Skeletons were usually accessible as a contribution for human action recognition. Hou et al. (2018) [23] depicted a strategy to encode the spatio-temporal data of a skeleton arrangement into colour texture images, referred to as skeleton optical spectra. It utilised a Convolutional Neural Networks (ConvNets) to attain the discriminative highlights for activity recognition. Such range portrayal made it conceivable to utilise a standard ConvNets design to take inappropriate powerful highlights from skeleton arrangements without preparing a huge number of parameters. In this technique, the ConvNets based end to end encoding of skeleton sequence was not available. To increase the accuracy rate, the Late score fusion was implemented, which involves the three orthogonal planes which were complementary to one another. On three standard used datasets, the results were obtained.

In intelligent video surveillance, anomalous action recognition was also a challenging research concern.

Sun et al. (2017) [24] presented Deep One-Class (DOC) by representing an end-to-end model that synchronised with the Convolutional Neural Network (CNN) using the one-class Support Vector Machine (SVM). The proposed model, related to the hierarchical models, finds the optimal global solution even though it simplifies the complication involved in the entire process. Approving DOC showed an easily accessible dataset, and this model has incredible execution; also, it is powerful for anomalous action recognition from surveillance videos. The video anomaly detection system should have been enhanced to acquire greater outcomes.

Separating discriminative and powerful highlights in action recognition of humans from the video sequence is the first and most basic advance. Liu et al. (2016) [25] determined a genetic programming (GP) strategy; this strategy enhances the motion feature descriptor on a population of crude 3D operators. The optimal arrangement chosen by GP was considered as the ideal action descriptor which has been attained. Feature learning was influenced by the proposed method, which lets a computer to routinely collect the holistic feature extraction using a pool of primitive operators, and these operators were formulated based on the basic idea of feature extraction. As an optimisation concern, the feature extraction done based on the GP approach was reviewed on four standard action datasets, specifically KTH, HMDB51, UCF YouTube, and Hollywood2. In all four datasets, the Experimental results showed that the proposed feature learning method based on GP accomplished improved performance in detection associated with modern hand-crafted and machine-learned methods.

3 | THE PROPOSED METHOD

Herein the article intends a new anomaly detection technique for congested ones with a covariance matrix description model for Local and Global Abnormal Detection. In Figure 1, the plan of the proposed approach is encapsulated. Initially using the Gaussian Mixture Model algorithm (GMM), the extraction of the foreground objects is done.

To model the distribution of image intensities, the GMM is used per image location, then the foreground pixels in one locality were categorised into a region of interest (ROI). Then the optical flow is measured using Lucas–Kanade method as this method is easy to compare to another method, very fast calculation with accurate time derivatives. Then, a feature descriptor is proposed to perform as per the movement based on the usage of covariance matrix coding and equivalent partial derivatives of several frames in a connective state, otherwise the patches of the frames. The movements in a high range of accuracy is represented by the covariance matrix of multi-RoI (region of interest). The covariance matrix descriptor is combined with the optical flow and equivalent partial derivatives, then prompted into feature space of lesser dimension. The issues of common binary classification varied for the abnormal event detection, which contains some samples which are normal besides very less or null abnormal samples used for training. Thus, instead of using

![FIGURE 1 Overview of proposed method](Image)
other classification approaches, the logistic regression method is used in this paper.

3.1 | Foreground detection by using Gaussian mixture model

The procedure involved in Foreground detection is extracting the object in motion (called foreground object) from the information, which is in static form termed background. In multistage computer vision systems, it remains the backbone and as a principal pace in various computer vision fields, namely video surveillance systems. So, the overall system performance is based on the result of foreground detection. GMM, among several algorithms, is one of the well-known ones because of its adaptive nature, good accuracy as well as less computational cost [26]. Each pixel in GMM [27] is modelled as a Gaussian mixture, then based on the mean and variance, it is categorised as the foreground or background. For identifying a specific pixel value, the probability is

\[ P(X_i) = \sum_{i=1}^{K} \omega_i, \eta \left( X_i; \mu_{i,t}, \sigma_{i,t}^2 \right) \]  

From (1) the probability density function \( \eta \) is represented as:

\[ \eta(X_i; \mu_{i,t}, \sigma_{i,t}^2) = \frac{1}{\sigma_{i,t} \sqrt{2\pi}} \exp\left(-\frac{(X_i - \mu_{i,t})^2}{2\sigma_{i,t}^2}\right) \]  

in which, \( K \), \( \omega_{i,t} \), \( \mu_{i,t} \), \( \sigma_{i,t}^2 \), are the number of Gaussian, weight estimation, the \( i \) th Gaussians mean and variance in the mixture at time \( t \). The decision standards at time \( t \), so the specific pixel as background or foreground is:

\[ |X_i - \mu_{i,t}| > \lambda \sigma_{i,t} \]  

In (3), \( \lambda \) is a perpetual threshold equivalent to 2.5, with one of the K Gaussian modules, if a match is noticed, the pixel is grouped as background. Moreover its factors are updated as:

\[ \omega_{i,t+1} = (1 - \alpha)\omega_{i,t} + \alpha \]  

\[ \mu_{i,t+1} = (1 - \beta)\mu_{i,t} + \beta X_{i,t+1} \]  

\[ \sigma_{i,t+1}^2 = (1 - \beta)\sigma_{i,t+1}^2 + \beta(X_{i,t+1} - \mu_{i,t})(X_{i,t+1} - \mu_{i,t})^T \]  

\[ \beta = \alpha \eta \left( X_{i,t+1}; \mu_{i,t+1}, \sigma_{i,t+1}^2 \right) \]  

where \( \alpha \) and \( \beta \) are the learning rate, which is constant also second learning rate, respectively. The pixel is sorted as a foreground, if none of any match is observed among any of the K Gaussian, besides only weight is updated:

\[ \omega_{i,t+1} = (1 - \alpha)\omega_{i,t} \]  

\[ \mu_{i,t+1} = \mu_{i,t} \]  

\[ \sigma_{i,t+1}^2 = \sigma_{i,t}^2 \]  

Gaussian Mixture Model is applied for extracting the foreground based on shape. Next to the extraction process, for separating skeleton shape from persisting objects this analysis is used. Foreground region of interest (ROI) is termed to feature extraction process. For that, we use feature extraction based on optical flow process. Optical flow feature set is termed to denote the segmented object(s). The optical flow based feature vectors are calculated beside the boundary and the feature set includes the shape and instantaneous velocity data mined through the action performers boundaries. The extracted optical flow based features are fed to a logistic regression.

3.2 | Lucas–Kanade method for optical flow measurement

The Lucas–Kanade (LK) method is generally utilised for estimating optical flow as well as computer vision in a differential method. The simple optical flow equations are solved by this approach for every pixel in that neighbourhood, using the minimum squares principle. This approach is not able to provide the details of flow in the identical internal regions of the image; hence it is a local method. In this approach, the flow is supposed to be constant in a pixel of the local neighbourhood under consideration [28]. The equation based on the optical flow can be made to hold for every pixel within a window positioned at \( P \). The LK method states that the movement of the image contents among two adjacent frames are in-significant and almost constant within the point P’s neighbourhood [29].

\[ I_x (p_1) \ V_x + I_y (p_1) \ V_y = -I_t (p_1) \]  

\[ I_x (p_2) \ V_x + I_y (p_2) \ V_y = -I_t (p_2) \]  

\[ I_x (p_n) \ V_x + I_y (p_n) \ V_y = -I_t (p_n) \]  

Inside the window \( P_1, P_2, \ldots, P_n \) are the pixels, and the partial derivatives of the image \( I \) are \( I_x(p), I_y(p), \) and \( I_t(p) \) concerning the position \( (x, y) \) and time \( t \), estimated at the point \( ‘P_n’ \) at the current time. In matrix form, the equations can be denoted:

\[ Av = b \]
where
\[
A = \begin{bmatrix}
I_x(p_1)I_y(p_1) \\
I_x(p_2)I_y(p_2) \\
\vdots \\
I_x(p_n)I_y(p_n)
\end{bmatrix},
\quad \nu = \begin{bmatrix}
V_x \\
V_y
\end{bmatrix},
\quad b = \begin{bmatrix}
-I_x(p_1) \\
-I_x(p_2) \\
\vdots \\
-I_x(p_n)
\end{bmatrix}
\]  

(14)

By means of the least squares standard, the solution of LK is determined. Now \(A^TA\nu = A^Tb\) can be expressed in another form and obtain
\[
V = (A^TA)^{-1}A^Tb
\]
where \(A^T\) is the transpose matrix which can be written as:
\[
\begin{bmatrix}
V_x \\
V_y
\end{bmatrix} = \left( \begin{bmatrix}
\sum I_x(p) & \sum I_x(p)I_y(p) & \sum I_x(p)^2 \\
\sum I_x(p)I_y(p) & \sum I_y(p) & \sum I_y(p)^2
\end{bmatrix} \right)^{-1} \times \begin{bmatrix}
\sum I_Ix(p)I_y(p) \\
\sum I_y(p) \\
\vdots \\
\sum I_y(p)^2
\end{bmatrix}
\]

(15)

Here, \(n = N\) to 1
The structure tensor of the image is characterised by the matrix \(A^TA\) at the point \(P\).

### 3.2.1 Weighted window

The former presented plain least squares clarification provides a similar prominence to all \(N\) pixels \(P_n\) in the window. In the procedure, it is generally suitable to provide additional weight to the pixels that are adjacent to the central pixel \(P\). For this purpose, the least squares equation’s weighted version is used [30].

\[A^TWAV = A^Tb,\]

this equation can be written so that to obtain \(V\).

In which \('W'\) is diagonal matrix of \((n \times n)\) with the weights \(W_i = \nu_i\) which is to be assigned to the equation of the pixel \(P_i\), which computes the value of \(V\):
\[
\begin{bmatrix}
V_x \\
V_y
\end{bmatrix} = \left( \begin{bmatrix}
\sum w_i I_x(p) & \sum w_i I_x(p)I_y(p) & \sum w_i I_x(p)^2 \\
\sum w_i I_x(p)I_y(p) & \sum w_i I_y(p) & \sum w_i I_y(p)^2
\end{bmatrix} \right)^{-1} \times \begin{bmatrix}
\sum w_i I_x(p)I_y(p) \\
\sum w_i I_y(p) \\
\vdots \\
\sum w_i I_y(p)^2
\end{bmatrix}
\]

(16)

The weight \('w'\) is generally allocated to a Gaussian function of the distance concerning \(P_i\) and \(P\). A weighted least squares (LS) fit of local first-order constraints has been applied by certain approaches such as Lucas and Kanade and others to model for \('r'\) which is a constant in every small spatial neighbourhood by minimising the below equation.
\[
V = \sum w^2(\nabla_x, \nabla_y)[(\nabla I(\nabla_x, \nabla_y, t))^2]
\]

(17)

In (17), \(w(x, y)\) represents a window function that provides considerable influence to constrain at the neighbourhood’s centre.

### 3.3 Motion descriptor

The goal of finding the anomalous action present in a surveillance video, a feasible descriptor is used for describing the information of the movement, which could be processed through a machine learning approach. Optical flow is a form of object’s motion visual formed based on the comparative motion concerning an observer and a scene. Using a basic feature, the optical flow is preferred to signify the movement among two image frames. As an objective function, the optical flow approach is optimised, which using a spatial term modelling, fuses with the brightness constancy limit, in whatever approach the flow probably varies across the image. Using a comprehensive constraint of smoothness, the aperture concerns could be solved. The Optical flow is formulated in Section 3.2. The Feature descriptor based on a covariance matrix dependent optical flow [31] combines the information of the spatio-temporal motion of some of the consecutive RoI (Region of Interest).

Initially, the video clips are divided into classes comprising of \(n\) number of frames, then for each of the two sequential frames, the optical flow is computed. The RoI features of the pixels in every group are arranged from \(1^{st}\) to \(n^{th}\). As an example, from the \(k^{th}\) group, the position of the pixel from \((1, 1)\) to \((b, w)\) are experimented using the height \(‘b’\) also width \(‘w’\) in a RoI. The coding related to the information of the movements are done using the matrix with \((2 + l)\) columns and \((b \times w)\) rows, the optical flow relied on features length is denoted as \(‘f’\); also the height and width of the RoI are denoted as ‘\(b’\) and ‘\(w’\’. One RoI’s (R) feature matrix is represented, as shown in (18):

\[
R = \begin{bmatrix}
11... feature \\
11... feature \\
...... \\
...... \\
...... \\
...... \\
bw feature
\end{bmatrix}
\]

(18)

Every RoI present in a group is classified with the serial numbers from \(1^{st}\) to \(n^{th}\) in a multi-RoI covariance feature, then one RoI groups feature matrix is represented as in Equation (19), Where \(I\) is a vector with all the elements 1. One group matrix’s dimension is \((n \times b \times w) \times (3 + l)\). The initial format of the column in the group is the RoI position, and further columns are the position of the pixel in one RoI. Based on the objective, the
TABLE 1  The constructed covariance descriptor based on the optical flow feature

| Feature Vector F | Description                                                                 |
|------------------|-----------------------------------------------------------------------------|
| F₁(4*4)          | [y,x,U,V]                                                                   |
| F₂(8*8)          | [y,x,U,Uₓ,Uᵧ,Vₓ,Vᵧ]                                                        |
| F₃(14*14)        | [y,x,U,Uₓ,Uᵧ,Vₓ,Vᵧ,Uₓₓ,Uᵧᵧ,Vₓₓ,Vᵧᵧ,Uₓᵧ,Vₓᵧ]                               |

FIGURE 2  The multi-RoI covariance matrix features calculation in progress. The optical flow of RoI is represented based on arrows. Blue, green and red colours denote different RoI's features based on the optical flow is determined. For the problems related to the detection of an anomalous event, the feature based on the optical flow is presented in Table 1.

The issues present in the feature selected for abnormal detection is described as:

\[ F(X,Y) = \phi(X,Y,U,V) \]  \hspace{1cm} (19)

In (20), the horizontal and vertical optical flow are denoted as \( U \), \( V \). \( \phi \) is the \( j \)th feature's optical flow. In the dimension \((h \times w) \times \) \((3 + l)\), \( 'F' \) denotes the feature gained for a single group of RoI, which is described in Equation (18). With the feature \( F \), the covariance matrix of dimension \((3 + l) \times \) \((3 + l)\) is presented in (20) [32]:

\[ c = \frac{1}{M-1} \sum_{i=1}^{m} (Z_i - \mu)(Z_i - \mu)^T \]  \hspace{1cm} (20)

In (20) the number of the pixels sampled is denoted as ‘\( M \)’. For a group of RoI it is signified as \( n \times h \times w \). The mean of the ‘\( m \)’ feature vector is ‘\( \mu \)’ and the \( j \)th point’s feature vector is ‘\( Z_j \)’.

Therefore, the covariance matrix unites frequent features. The proposed feature’s calculation is shown in Figure 2. A feature matrix ‘\( F' \)’ is represented in a video stream ‘\( V' \)’, with group ‘\( G \)’ having \( n \) number of frames, besides the multi-frame covariance matrix ‘\( \epsilon \)’ is computed. The intra-RoI information is defined by the pixels position, optical flow as well as the equivalent partial derivatives. Furthermore, the inter-RoI knowledge stamps by the RoI position in one group. Hence, the RoI movement property combined by the proposed covariance descriptor exposes the spatial-temporal feature. Due to symmetry, based on the covariance matrix ‘\( \epsilon \)’ that combines the ‘\( F' \)’ features, it has \( \frac{l^2 + l}{2} \) dissimilar values with the support of the integral image, the complications in creating a rectangular region covariance which is \( O(l^2) \). When the raw values are represented in a similar region, the feature dimension is \( n \times l \). Specifically, the number of the pixels are denoted by ‘\( n \)’. The feature’s dimension is considered as ‘\( B' \)’, if histograms of the joint features are used, where ‘\( B' \)’ denotes the number of histogram bins. Hence, when compared to other region descriptors, the covariance matrices are effective and is of low-dimensional.

3.4  |  Logistic regression analysis for abnormal behaviour detection

The logistic regression analysis for the detection of abnormal behaviour is introduced in this section. Logistic regression, as it is known, might be bi- or multinomial. Binary otherwise binomial logistic regression is the case wherein the identified result possess merely two conceivable forms (e.g. ‘alive’ vs. ‘dead’, ‘success’ vs. ‘failure’, or ‘yes’ vs. ‘no’) [33]. As commonly, the logistic regression analysis approaches such as binary logistic regression and multinomial logistic regression approaches are applied to tackle the concerns regarding classification problems since the probability of occurrence of an incidence cannot be computed directly.

Logistic regression (LR), in practice, is generally used as a classifier, specifically meant for probabilistic binary or multinomial classification [34]. For the category selection, the principle in logistic regression is based on the probability which is highest and is generated using a logistic function. The entire procedure flow chart is shown in Figure 3.

In LR, the user is provided with the probabilities which are explicit for classification also for class information. LR [35] is...
an arithmetical approach for examining a dataset wherein the determining variable is dichotomous (binary).

LR finds its application to find the relations concerning a single determined variable or a few self-determining variables. Every independent variable with weights are multiplied and then added. To obtain the result between 0 and 1, sigmoid function are added. The values more than 0.5 are noted as 1, and the values less than 0.5 are noted as 0.

It is essential to obtain the optimal one among the weights or regression coefficients, and for this purpose, the optimisation techniques is used.

Definition of logistic regression can be made in Equation (21):

\[
V = -(A_i \in R^d, B_i \in R)^n = 1, \quad (21)
\]

where,

\[
A_i \in R^d \text{ is } d \text{ dimensional self-determining variables}
\]

\[
B_i \in R \text{ is the determined variables}
\]

Here, the matrix A and B lies from -1 to 1.

The probability of ‘B’ turns to 1 which is given in Equation (22).

\[
P(B = 1|a, \theta) = \frac{1}{1 + e^{-\theta a}} \quad (22)
\]

The possibility of ‘B’ turns to -1 is

\[
P(B = 0|A, \theta) = 1 - P(B = 1|A, \theta) \quad (23)
\]

where

\[
\theta \in R^d \text{ weights}
\]

The appropriate weight vector ‘\theta’ is to be found for the negative log likelihood for reducing the error function (35),

\[
F(\theta) = \sum_{i=1}^{n} - \log \left( p(b_i|a_i, \theta) + \lambda \| \theta \|_1 \right) \quad (24)
\]

In which ‘\lambda’ is the regularisation term to correct the large weight parameters.

3.5 Stochastic gradient descent to minimise an error function

In logistic regression to obtain the best regression weights, two optimisation functions are generally used, that is, gradient descent and stochastic gradient descent (SGD). To minimise an error function, a set of parameters are updated by both the algorithms in an iterative manner. In the training set, the complete samples are trained by the SGD algorithm in a specific iteration to ensure a single parameter update.

The SGD algorithm uses only a single training sample in a specific iteration of the training set to do the update for a parameter.

Therefore, for big data analytics, the gradient descent algorithm does not suit. In the LR to update the weights, the SGD algorithm is considered, and it is updating the classifier, as new data approaches in considerably at the same time.

The typical SGD algorithm with the objective ‘\[F(\theta)\]’ updates the parameters ‘\theta’ as,

\[
\theta = \theta - \alpha \nabla F(\theta) \quad (25)
\]

In (25), the probability ‘\[E[\theta]\]’ is described by calculating the full training sets cost as well as a gradient. SGD uses a particular or small number of training samples, merely carry out using the probability in the update then figures the gradient of the parameters.

The new update is specified by

\[
\theta = \theta - \alpha \nabla F(\theta, A(i), B(i)) \quad (26)
\]

with a pair (\(A(i), B(i)\)) from the training set used.

4 RESULTS OF EVENT ANALYSIS

The experimental results of the approach proposed for anomalous event detection is evaluated in this section provide experiments with standard datasets like BEHAVE [36], UMN [37], and UCSD [38]. In order to develop a multi-RoI covariance feature, consecutive four frames RoI is used. For global range abnormal event detection, the entire frame is chosen as RoI in UMN datasets. For local abnormal event detection, every 16 × 12 blocks is carefully chosen as RoI in the UCSD dataset.

4.1 UMN dataset

This dataset is gathered from UMN, the situation in the panic state is identified as the abnormal event which contains the lawn, indoor and plaza scenes. There are 11 series of videos in the dataset. In every dataset, the samples found to be normal or abnormal are separated. The frames in which the walking people are considered as the training samples as well as normal testing samples. The frames in which people running are considered panic scenes, which is regarded as abnormal samples.

The ROC (Receiver operating characteristics) curve of lawn, indoor and plaza scenes in the UMN dataset is presented in the following Figures 4, 5, and 6, respectively. Table 2 provides the detection performance.

In Table 3, the performance evaluations compared with the current methods are shown. From the normal training frames,
FIGURE 4 shows the lawn scene’s ROC with dissimilar feature.

FIGURE 5 shows the indoor scene’s ROC with dissimilar feature.

FIGURE 6 shows the plaza scene’s ROC with dissimilar feature.

TABLE 2 The AUC obtained based on various features for lawn, indoor and plaza experiments

| Feature   | Area Under ROC |
|-----------|----------------|
|           | Lawn | Indore | Plaza |
| F1(4*4)   | 0.9955 | 0.7035 | 0.9532 |
| F2(8*8)   | 0.9998 | 0.9621 | 0.9870 |
| F3(16*16) | 0.9997 | 0.9567 | 0.9899 |

the proposed approach for abnormal detection in the UMN dataset has to prove that it can differentiate the global abnormal event. Furthermore, in Figure 4, 5, and 6 it is seen that the F1 (4 × 4) feature descriptor comes out with the results of a low

TABLE 3 The performance estimation of multi-RoI covariance matrix of optical flow based method and the current methods of the UMN dataset

| Techniques  | Area Under ROC |
|-------------|----------------|
| Social Force [39] | 0.96 |
| Optical Flow [39] | 0.84 |
| NN [40] | 0.93 |
| SRC [40] | 0.995 | 0.975 | 0.964 |
| STCOG [41] | 0.9362 | 0.7759 | 0.9661 |
| HOFO [42] | 0.9845 | 0.9037 | 0.9815 |
| Proposed | 0.9998 | 0.9621 | 0.9870 |
FIGURE 7  UMN dataset results (a) Train lawn (b) Train indore (c)Train Plaza (d) Test lawn (e) Test Lawn (f) Test Plaza

TABLE 4  Quantifiable evaluation of proposed approach with the existing approaches

| Method      | Evaluation Conditions |
|-------------|-----------------------|
|             | EER       | RD       | AUC       |
| Sparse      | 19 %      | 46 %     | 46.1 %    |
| Adam        | 38 %      | 24 %     | 13.3 %    |
| MPPCA+SF    | 32 %      | 27 %     | 21.3 %    |
| SF          | 31 %      | 21 %     | 17.9 %    |
| MPPCA       | 40 %      | 18 %     | 20.5 %    |
| MDT         | 25 %      | 45 %     | 44.1 %    |
| Proposed    | 24 %      | 64 %     | 67.8 %    |

range of accuracy, besides the high dimensional features gain more accurate results.

Figure 7 shows the results of the UMN dataset. a, b and c are the training of the normal samples from lawn, indoor and plaza scenes in which people are walking all over the place. d, e, f are the abnormal sample testing, which shows the persons are in motion.

Generally, the classification based detection method can be achieved with higher accuracy when the covariance matrix fuses more features. However, discrimination could also be weakened by more partial derivatives. Thus, for the detection of abnormal events, appropriate feature components are essential.

Also, some of the existing algorithms, like Adam [43], SF [44], MPPCA [44], MDT [45], are compared in Table 4.

4.2  UCSD dataset

To find local abnormal events, the UCSD dataset is used, likewise, global abnormal events along with the proposed approach in UMN datasets have been detected. For training, Normal events are used, which contain only pedestrians. In contrast, the abnormal events resemble uncharacteristic motion patterns of an individual or the movement of ordinary objects, including motorcycle otherwise van in the path. Twelve testing clips and 16 training clips and are incorporated in Ped2. Each clip has frames of 150 to 180 and a size of 360 × 240 pixels. Likewise, every clip is presented using a binary mask of pixel-level generated manually, could recognise the region having the anomalies [46]. For cell size calculation, 10 clips other than, per clip 20 frames were selected randomly. The appropriate cell size is revealed to be 30 × 17 pixels.

(i) Frame-level: In this aspect, a frame is said to be anomalous if no less than one pixel in the frame is identified as anomalous. These results of frame-level detection are matched to each frame’s frame-level ground truth. Then to find the number of true positive numbers as well as false positive numbers, the detection results are implemented. Though, this frame-level measurement cannot make sure that the detected anomalous pixel coincides with the actual anomalous position. For the reason that some portion of false pixel detection may cause a true positive frame.

(ii) Pixel-level: In this aspect, the results of detection are compared to each frame pixel-level ground truth. If 40% (or more) of anomalous ground truth pixels are identified as true positives, then the frame is considered to be an anomaly. A normal frame, instead, is recognised as false positive in case any normal pixel is found as anomalous. The pixel-level magnitude compared with the frame-level magnitude is much severe and concentrates more on the correct localisation of an anomaly event. The frame-level and pixel-level conditions are utilised for the experimental performance calculation. The frame level approach identifies which frame contains abnormal activities. The abnormal detection is related along with the remarks of the frame level ground truth.

In Figure 8, (a) cycling, (b) cycling of two persons and (c) EER, RD, and AUC are used for performance evaluation. EER (equal error rate) is the frames misclassified part, while the false positive rate is equivalent to the frame-level miss rate for the standard. The detection rate at the equal error is known as RD (rate of detection). The AUC is a field of area under the ROC curve. The accurate objects are not extracted, and the exact outlines of the moving objects are either not obtained in this proposed approach. Every image of size 16 × 12 is divided into local patches. In the training dataset for local abnormal event detection, the descriptor is extracted with the training sample of feature F2 (8 × 8). The images feature descriptor at the equivalent position is classified in the testing dataset. As height and width dimension, the patches are shifted along with both two pixel strides, and by logistic regression, each 2 × 2 block is classified several times. In Figures 8 and 9, the performance of the local abnormal detection are shown.

In the detection of the frames, the approach of frame-level especially concentrates on which having the abnormal event, whereas pixel-level approach particularly concentrates on the
FIGURE 8 Examples for the analysis of local abnormal detection based on the proposed algorithm

The results are compared with various evaluation standards of the existing approaches are mentioned in Table 4. The proposed approach, as well as the algorithm for anomalous detection, performs better than existing algorithms.

Some more example of detected normal/abnormal behaviour in BEHAVE videos is shown in Figure 10.

Figure 10(a–f) represented the people walking behind the van.

5 | CONCLUSION

In this paper, the logistic regression based approach is presented for detecting abnormal behaviours in a crowded scene. A multi-RoI covariance (MCOV) descriptor is used for combining the optical flow based feature of multiple fields of interest. In video streams, the Optical flow (OF) is considered as a critical feature. Based on normal samples training, the predictions are transformed with the logistic regression through logistic function to find abnormal events. To prove this proposed methods competitive property, the experiments are conducted using the benchmark datasets, that is, BEHAVE, UMN, and UCSD. Numerous algorithms deal with the anomaly detection with highly structured scenes. This is due to the shortage and nearly infinite variety of abnormal actions in real life. Though, attaching the vast quantity of footage that is caught by the omnipresent CCTV cameras in several public places around the world can offer a source for benchmark datasets, which in different contexts of interests are used. Certain datasets would facilitate researchers to assess how fine an abnormal-detection algorithm executes in two significant tasks, specifically abnormality detection. In the dataset, the ground truths for each frame should be provided to evaluate performance on this task. The other is the localisation of anomaly. In any anomaly detection system, it is vital not only to do well in identifying the presence of an abnormality in the scene but to find where it is taking place. In unstructured situations, there is a requirement to test the performance of these approaches. More research attention needs to be dedicated to the real world of messy environments, which has many moving objects and activities and for the progress of agendas that will especially and effectively deal with the problem of scalability of video analysis. This remains an open challenge. Based on current research, the forthcoming research scopes are action recognition in multimedia event detection problems. In future work, the possibly outspreads current approach into a network

location of the abnormal pixel. The performance result demonstrates that the frames can be detected by using a sparse reconstruction method where the abnormal event happens, but in ground truth it does not identify the real abnormal pixels. Compared to the sparse reconstruction method, the accuracy of the proposed method based on the location is higher.
system with multi-camera to upturn the detection performance of abnormal behaviour in several conditions.

REFERENCES

1. Varol, G., et al.: Long-term temporal convolutions for action recognition. IEEE Trans. Pattern Anal. Mach. Intell. 40(6), 1510–1517 (2018)
2. Akula, A., et al.: Deep learning approach for human action recognition in infrared images. Cogn. Syst. Res. 50, 146–154 (2018)
3. Wang, Y., et al.: Visual attribute-augmented three-dimensional convolutional neural network for enhanced human action recognition. arXiv preprint, arXiv: 1805.02860, 2018
4. Rahmani, H., et al.: Learning a deep model for human action recognition from novel viewpoints. IEEE Trans. Pattern Anal. Mach. Intell. 40(3), 667–681 (2018)
5. Talha, S.A.W., et al.: Features and classification schemes for view-invariant and real-time human action recognition. IEEE Trans. Cogn. Developmental Syst. 10, 894–902 (2018)
6. He, D., et al.: Exploiting spatial-temporal modelling and multi-modal fusion for human action recognition. arXiv preprint arXiv:1806.10319, 2018
7. Zhou, S., et al.: Spatial–temporal convolutional neural networks for anomaly detection and localization in crowded scenes. Signal Processing. Image Comm. 47, 358–368 (2016)
8. Zhang, H, Parker, L.E.: Code4d: color-depth local spatio-temporal features for human activity recognition from rgb-d videos. IEEE Trans. Circuits Syst. Video Technol. 26(5), 541–555 (2016)
9. Liu, H., et al.: Computational model based on neural network of visual cortex for human action recognition. IEEE Trans. Neural Netw. Learn. Syst. 29(5), 1427–1440 (2018)
10. Núñez, J.C., et al.: Convolutional neural networks and long short-term memory for skeleton-based human activity and hand gesture recognition. Pattern Recognit. 76, 80–94 (2018)
11. Wang, P., et al.: Action recognition based on joint trajectory maps with convolutional neural networks. Knowledge Based Syst. 158, 43–53 (2018)
12. Wang, P., et al.: Depth pooling based large-scale 3-D action recognition with convolutional neural networks. IEEE Trans. Multimedia 20(5), 1051–1061 (2018)
13. Minh, T.L., et al.: A fine-to-coarse convolutional neural network for 3D human action recognition. arXiv preprint, arXiv:1805.11790, 2018
14. Katiccioğlu, L., et al.: Learning latent representations of 3D human pose with deep neural networks. Int. J. Comput. Vision 126, 1326–1341 (2018)
15. Coyar, S., et al.: Toward abnormal trajectory and event detection in video surveillance. IEEE Trans. Circuits Syst. Video Technol. 27(3), 683–695 (2017)
16. Singh, D, Mohan, C.K.: Graph formulation of video activities for abnormal activity recognition. Pattern Recognit. 65, 265–272 (2017)
17. Mousavi, H., et al.: Detecting abnormal behavioral patterns in crowd scenarios. In: Toward Robotic Socially Believable Behaving Systems, Vol. II, pp. 185–205. Springer, Cham (2016)
18. Sabokrou, M., et al.: Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes. Comput. Vision Image Understanding 172, 88–97 (2018)
19. Fang, Z., et al.: Abnormal event detection in crowded scenes based on deep learning. Multimed. Tools Appl. 75(22), 14617–14639 (2016)
20. Chaker, R., et al.: Social network model for crowd anomaly detection and localization. Pattern Recognit. 61, 266–281 (2017)
21. Liu, M., et al.: Enhanced skeleton visualization for view invariant human action recognition. Pattern Recognit. 68, 346–362 (2017)
22. Ijjina, P.,Chalavadi, K.M.: Human action recognition in RGB-D videos using motion sequence information and deep learning. Pattern Recognit. 72, 504–516 (2017)
23. Hou, Y., et al.: Skeleton optical spectra-based action recognition using convolutional neural networks. IEEE Trans. Circuits Syst. Video Technol. 28(3), 807–811 (2018)
24. Sun, J., et al.: Abnormal event detection for video surveillance using deep one-class learning. Multimed. Tools Appl. 78, 3633–3647 (2017)
25. Liu, L., et al.: Learning spatio-temporal representations for action recognition: A genetic programming approach. IEEE Trans. Cybernetics 46(1), 158–170 (2016)
26. Reynolds, D.: Gaussian mixture models. In: Encyclopedia of Biometrics, pp. 827–832, Springer, New York (2015)
27. Bouwmans, T.: Traditional and recent approaches in background modeling for foreground detection: An overview. Comput. Sci. Rev. 11–12, 31–66 (2014)
28. Plyer, A., et al.: Massively parallel Lucas Kanade optical flow for real-time video processing applications. J. Real-Time Image Process. 11(4), 713–730 (2016)
29. Patel, D., Upadhyay, S.: Optical flow measurement using lucas kanade method. Int. J. Comput. Appl 61(10), 6–10 (2013)
30. Barron, J. T., Thacker, N.A.: Tutorial: Computing 2D and 3D optical flow, Vol. 1. Imaging Science and Biomedical Engineering Division, Medical School, University of Manchester (2005)
31. Wang, T., et al.: Abnormal event detection via covariance matrix for optical flow based feature. Multim. Tools Appl., 77(13), 17375–17395 (2018)
32. Tuzel, O., et al.: Region covariance: A fast descriptor for detection and classification. In: European Conference On Computer Vision, pp. 589–600. Springer, Berlin, Heidelberg (2006)
33. Liu, D., et al.: Incorporating logistic regression to decision-theoretic rough sets for classifications. Int. J. Approx. Reason. 55(1), 197–210 (2014)
34. Liu, W., et al.: Multiview Hessian regularized logistic regression for action recognition. Signal Process. 110, 101–107 (2015)
35. Kang, D., et al.: Data/feature distributed stochastic coordinate descent for logistic regression. In: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, pp. 1269–1278. ACM, New York (2014)
36. University of Edinburgh: The BEHA VE dataset. http://groups.inf.ed.ac.uk/vision/BEHAVEDATA/INTERACTIONS/. Accessed 11 May 2009
37. Mehran, R., et al.: Abnormal crowd behaviour detection using social force model. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 935–942 (2009)
38. UCSD: Anomaly detection dataset. http://www.svcl.ucsd.edu/projects/anomaly/dataset.html (2010)
39. Mehran, R., et al.: Abnormal crowd behavior detection using social force model. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 935–942. IEEE, Piscataway (2009). Anomaly detection dataset, 2010
40. Cong, Y., et al.: Sparse reconstruction cost for abnormal event detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3449–3456. IEEE, Piscataway (2011)
41. Shi, Y., et al.: Real-time abnormal event detection in complicated scenes. In: Proceedings of International Conference on Pattern Recognition (ICPR), pp. 3653–3656. IEEE, Piscataway (2010)
42. Wang, T., Snoussi, H.: Detection of abnormal visual events via global optical flow orientation histogram. IEEE Trans. Inf. Forensic Secur. 9(6), 988–998 (2014)
43. Adam, A., et al.: Robust real-time unusual event detection using multiple fixed-location monitors. IEEE Trans. Pattern Anal. Mach. Intell. 30(3), 555–560 (2008)
44. Mahadevan, V., et al.: Anomaly detection in crowded scenes. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1975–1981. IEEE, Piscataway (2010)
45. Li, W., et al.: Anomaly detection and localization in crowded scenes. IEEE Trans. Pattern Anal. Mach. Intell. 36(1), 18–32 (2014)
46. Roshthikari, M.J., Levine, M.D.: An on-line, real-time learning method for detecting anomalies in videos using spatio-temporal compositions. Comput. Vis. Image Underst. 117(10), 1436–1452 (2013)

How to cite this article: Kumaran N, Reddy US. Classification of human activity detection based on an intelligent regression model in video sequences. *IET Image Process*. 2021;15:65–76. https://doi.org/10.1049/ipr2.12006