IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received December 31, 2020, accepted January 16, 2021, date of publication January 25, 2021, date of current version February 2, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3054040

Anomaly Detection With Particle Filtering for Online Video Surveillance

ATA-UR-REHMAN[1], SAMEEMA TARIQ[2], HAROON FAROOQ[1], ABDUL JALEEL[3], AND SYED MUHAMMAD WASIF[4]

[1]Department of Electrical Engineering (RCET), University of Engineering and Technology, Lahore 39161, Pakistan
[2]Department of Electrical Engineering, University of Engineering and Technology, Lahore 39161, Pakistan
[3]Department of Computer Science (RCET), University of Engineering and Technology, Lahore 39161, Pakistan
[4]Department of Electrical Engineering, University of Gujrat, Gujrat 50700, Pakistan

Corresponding author: Ata-Ur-Rehman (a.ur.rehman@uet.edu.pk)

This work was supported by the Higher Education Commission of Pakistan under Grant 9827.

ABSTRACT With growing security threats, many online and offine frameworks have been proposed for anomaly detection in video sequences. However, existing online anomaly detection techniques are either computationally very expensive or lack desirable accuracy. This research work proposes a novel particle filtering based framework for online anomaly detection which detects video frames with anomalous activities based upon the posterior probability of activities in a video sequence. The proposed method also detects anomalous regions in anomalous video frames. We propose novel prediction and measurement models to accurately detect anomalous video frames and anomalous regions in video frames. Novel prediction model for particle prediction and likelihood model for assigning weights to these particles are proposed. These models efficiently utilise variable sized cell structure which creates variable sized sub-regions of scenes in video frames. Furthermore, they efficiently extract and utilise information from the video frame in the form of size, motion and location features. The proposed framework is tested on UCSD and LIVE datasets and compared with the existing state-of-the-art algorithms in the literature. The proposed anomaly detection algorithm outperforms the state-of-the art algorithms in terms of reduced Equal Error Rate (EER) with comparatively lesser processing time.

INDEX TERMS Video anomaly detection, online framework, particle filtering, inference mechanism.

I. INTRODUCTION

Automatic computer vision based video analysis for surveillance and security purposes is a challenging task. One of the applications of automatic video analysis is to detect anomalous activities in surveillance videos, which is also know as anomaly detection. Ideal video anomaly detection is defined as a process of detecting and tracking of abnormal events in real-time [1]. Machine learning based anomaly detection algorithms have achieved high accuracy in automatically detecting anomalous activities such as falling of objects, anomalous access in restricted areas, traffic accidents, traffic laws violations, criminal activities, and many more [2].

Anomaly detection is a challenging task because a complete list of possible anomalies are not known. Because of the rarity of anomalous activities, it is very difficult to collect data related to anomalies. Anomalies usually vary with the type of application and scenario. Hence, the definition of the anomalous event varies concerning the application. For instance, people running on a road define an anomalous event, while people running on football ground define a normal event. Therefore, anomaly detection algorithms usually use normal events as training data for training of models, and then apply model on online data to detect anomalies.

Anomaly detection algorithms can have a number of applications such as in: transportation systems, security systems, intrusion systems, and fault detection in industry. However, the scope of the this paper is limited to security systems only.

This paper presents a novel particle filtering based framework for online anomaly detection which detects video frames with anomalous activities based upon the posterior probability of activities in a video sequence. The proposed algorithm is based on particle filtering, where novel method is proposed for particle sampling in the context of anomaly

The associate editor coordinating the review of this manuscript and approving it for publication was Li He.

detection. Furthermore, a new technique to calculate likelihood of observations is proposed which helps to assign weights to particles in particle filtering.

The proposed algorithm divides video frames into variable sized cells. In the proposed particle filtering based technique, we first predict possible activities in a video frame. These predictions are further refined with the update step where, motion, location and size features extracted from video frames along with a clustering algorithm are used. This update step evaluates the posterior distribution of possible activities in video sequences which helps to classify video frames into two classes: anomalous and normal. Block diagram of the proposed particle filtering based anomaly detection framework is given in Fig. 1.

The proposed anomaly detection framework, efficiently detects anomalous video frames and further identifies regions in the video frames where possible anomalous activities are taking place. Proposed particle filtering based anomaly detection algorithm detects anomalous behaviours in video sequences by using minimum training data and can produce highly accurate results even when measurements are noisy.

Different publicly available data sets pertaining to real world scenarios are used to test the proposed framework. These videos are taken from the surveillance cameras. The proposed framework demonstrates an outstanding performance when compared with other state-of-the art online and offline anomaly detection techniques available in the literature.

II. RELATED LITERATURE

Anomaly detection is a challenging task, and many methods have been developed and applied in the literature that includes real-time detection, online detection and offline detection algorithms. Each method has its own merits and demerits. However, most of the anomaly detection applications require online and real time processing of video frames to detect anomalous activities.

Different deep neural network based approaches have been proposed in the literature for anomaly detection on abstraction feature learning [2]–[4] and on video prediction learning [5]. For anomaly detection, some normal patterns are modelled by sparse reconstruction method [6]–[9], which develops a dictionary on the basis of normal pattern and then patterns with high reconstruction error are detected as anomalies. The researchers have also employed agents approach for anomaly detection. One of the approaches is hybrid agent [10]. In this approach, the anomalous behavior is divided into group and single model, static and dynamic agents discriminate between these models.

Ryotaet et. al. [11] integrated convolution neural network and environment-dependent anomaly detector for detection and recounting of joint abnormal events. The proposed model was trained with generic environment knowledge at first and then with current experience it was trained on the environment-specific knowledge. In [12], a deep GMM model was developed and which employed PCANet for simultaneous appearance and motion features extraction. The deep model effectively detected the abnormal events in surveillance video. In [13]–[15] denoising auto-encoders and GANs were used to adversely learn latent representations for one-class novelty detection. A deep convolution neural network while utilizing ImageNet for feature extraction was used with transfer-level learning for an unsupervised anomaly detection in medical images [16]. In [17] a framework was proposed utilizing a deep auto-encoder as a parametric density estimator and through autoregression it learned the probability distribution of its underlying latent representations without any prior assumption about the nature of novelties. In a recent work [18], authors proposed two models to improve the performance of anomaly detection to overcome the limitations of model bias and domain prior. First model was based on log likelihood ratio of two identical models and in the second approach a Glow like multi-scale model was used.

Work on anomaly detection is mostly done on the normal videos motion model, but the work on both normal and abnormal models is done by [19]–[21]. The motion pattern modelling through weakly supervised method used an approach of multiple instance learning [20], [21], while anomaly scores are predicted by deep anomaly ranking model. Convolutional Neural Network (CNN) applied on the input sequence to predict state in [22], where, three learning policies to deal with intrinsic fuzzy predictions are proposed.

Most of the deep learning based anomaly detection methods provide hight detection accuracy, however, these techniques require large amount of training data and they are extremely dependent on large amount of computation resources. In [23] a deep learning based technique know as incremental spatiotemporal learner (ISTL) is proposed to address challenges and limitations of anomaly detection and localization for real-time video surveillance. However, this methods method gives relatively smaller detection accuracy

Many researchers adopt conventional techniques, they work on the regular pattern and when an irregular pattern occurs they call it as anomaly. For instance, the mixture of dynamic models on texture is proposed in [24], Gaussian process modeling in [25], [26], the Markov random field upon spatial-temporal domain in [27], the social force model based approach in [28] and Hidden Markov Models on video volumes in [29], [30].

Some of the authors provide new approach for anomaly detection. They used the information of motion and appearance of prior sequence to predict whole frame through Generative Adversarial Networks (GAN) and U-net as predictor [22].

The work presented in [31] and [32] are related to the proposed techniques. In [31], video frames are divided into variable sized cells and then background subtraction is performed on each video frame. It extracts features such as foreground occupancy, optical flow and Histograms of Optical Flow (HOF) descriptor which are analysed to construct

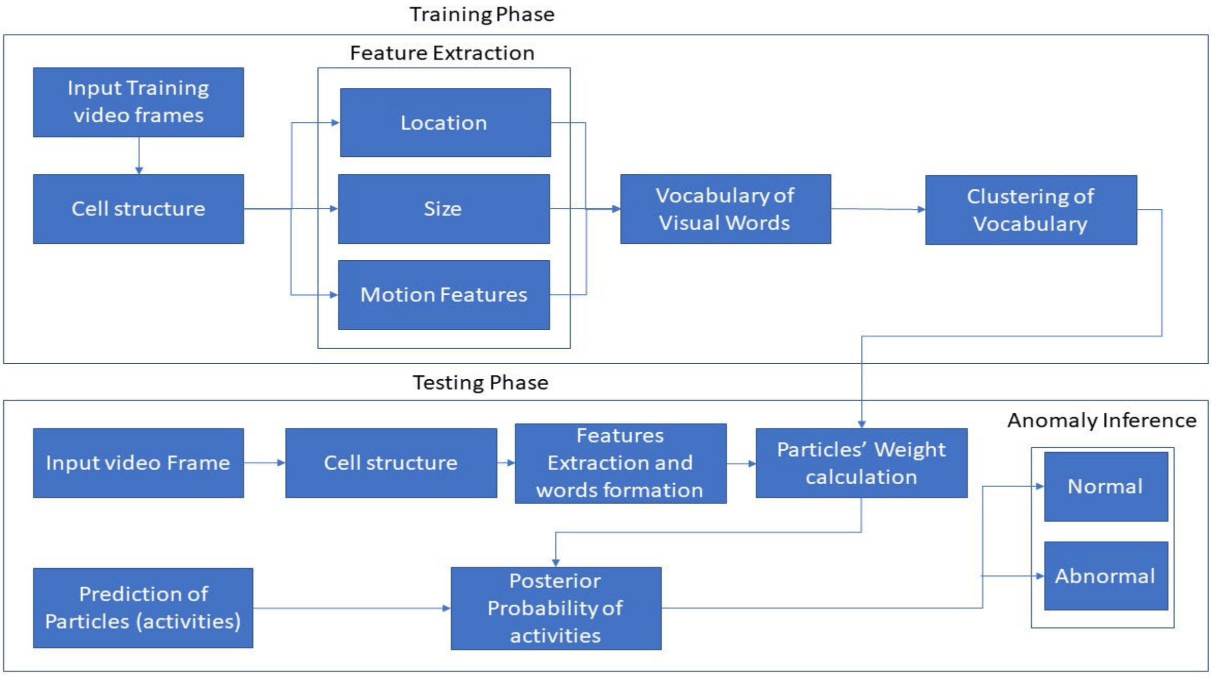


FIGURE 1. Block diagram of the proposed particle filtering based anomaly detection framework.

various models. These models are used to detect anomalous regions in video sequences.

We propose a particle filtering based framework which evaluates posterior distribution of activities in video sequences, and uses this posterior distribution to detect anomalous behaviours in videos. The proposed method is more efficient in terms of accuracy and computational cost.

A particle filtering based solution is also proposed in [32], which detects anomalies by using particle filter. The anomalies are decided on the basis of estimation loss. This loss is provided by applying optical flow on image sequences and then particle filter is applied. However, this techniques can only detect anomalous video frames but cannot find out anomalous regions within anomalous video frames. We propose a complete particle filtering based tracking algorithm which predicts activities with a novel prediction model, calculate likelihood of measurements by using a novel method, and finally estimates posterior distribution of activities.

The remainder of the paper is organized as follows: Section III describes the methodology of the proposed algorithm, extensive experimental validation is presented in Section IV, while Section V presents conclusions.

III. METHODOLOGY

The proposed method for the anomaly detection is mainly built on a probabilistic inference model. The proposed model analyses regions of video scene automatically to detect abnormal activities in it. The proposed method uses the basic structure of topic modeling [33] and develops a probabilistic model to analyse the activities happening in a video scene. The idea of topic modeling is not new, and it has been used for anomaly detection. However, in the proposed framework the concepts of topic modeling are used for feature extraction only. Actual contribution of this paper lies in presenting a novel particle filtering based anomaly detection framework which provides a more accurate solution for anomaly detection with lesser computational complexity.

In the context of topic modeling, the video sequence under consideration is divided into short non-overlapping clips which are named as visual documents. Each frame within a video clip is further divided into small non-overlapping cells of pixels. In each video clip which we call a document, there can be single or multiple activities happening. For instance, in a video recording of a public place, pedestrians walking on a footpath is an activity, while vehicles moving on a road can be another activity, which can be seen in Figs. 2 [33]. In the context of topic modeling, we call these activities topics. More precisely, topics correspond to the activities that are frequently occurring in a video scene. Whereas, topics are created due to the co-occurrence of certain visual features. These visual features are called visual words. A complete description of visual words in the context of particle filtering based anomaly detection algorithm is given in Section III-A. Meaning of an activity strictly depends on these visual words which have been used in the process of building a document.

A. VISUAL WORDS

In the proposed anomaly detection method, words are defined with the help of four different types of features: location features, two motion features; optical flow energy and HOF, and size of foreground objects. Therefore, we define an i^{th} word in a document as $\{\mathbf{w}_i^r\}_{r=1}^{N_w}$, where N_w is the dimension of a word, which is 4 in our case.

FIGURE 2. A usual scene recorded by a surveillance camera [33].

FIGURE 3. The cell structure of a scene. Different cell sizes are shown, where, regions closest to the camera are represented by larger cells, while regions away from the camera are represented by smaller cells.

1) CELL STRUCTURE AND LOCATION FEATURE

In anomaly detection, it is important to identify the location where an activity is taking place. Similar activities on different locations can have different meanings, for instance, a moving vehicle on a road is a normal activity, however, its movement on a footpath is an anomalous activity. To describe the location where an activity is taking place, we divide video frames into non-overlapping cells. We use the fact that detailed information about a scene in a video frame is given by the regions closer to the surveillance cameras while region away from it are less descriptive. Furthermore, objects closer to the camera look bigger compared to the objects away from the camera. This means that position of camera is very important in anomaly detection [34]–[36]. Therefore, as described in [31], we divide video frames into non-overlapping and variable size cells. Most of the surveillance cameras are installed at higher position to capture the scenes downward. This helps to capture clear view of a scene. In the proposed anomaly detection algorithm, similar position of a camera is considered. Therefore, variable size cells are developed in such a way that larger cells are created in regions close to camera. Such regions are found in the lower portions of a video frame. Similarly, cells of relatively smaller size are created in regions away from the camera. Such regions are found at upper portion of a video frame.

To create variable sized cells we start from the top left corner of a video frame. Vertical dimension of the k^{th} cell is y_k and the vertically adjacent $(k+1)^{th}$ cell is defined as [31]

$$y_{k+1} = \alpha y_k, \tag{1}$$

where $\alpha > 1$ is the growing rate. It will keep creating $(k+1)^{th}$ cell greater than the $(k)^{th}$ cell. Therefore, the vertical dimension Y_f of a video frame can be defined as

$$Y_f = \sum_{k=0}^{N_v} \alpha^k y_0 \tag{2}$$

where N_v is the number of cells along the vertical dimension and y_0 is the vertical dimension of the smallest cell. In order to get N_v, we initially set y_0 to a predefined starting value. Equation (2) can simply be converted into the form of geometric series as

$$\frac{Y_f}{y_0} = \frac{\alpha^{(N_v+1)} - 1}{\alpha - 1}. \tag{3}$$

The calculation for value of N_v can then be defined as [31]

$$N_v = log_\alpha \left(Y_f / y(\alpha - 1) + 1) - 1\right). \tag{4}$$

Vertical dimension of the smallest is adjusted by using equation (4). This adjusted dimension is denoted as $\hat{y}$

$$\hat{y} = \frac{(\alpha - 1)}{(\alpha^{(N_v+1)} - 1)} Y_f. \tag{5}$$

Similar procedure is employed to obtain the horizontal dimensions of cells in a video frame. We denote the horizontal dimension of a frame as X_f. To evaluate horizontal dimensions of cells in a video frame, we start from the top boarder of the frame, at position X/2, i.e. the medium-section of the frame, and create cells of variable dimension in horizontal direction. The growing rate α is same as used for defining vertical dimensions of cells, however, horizontal dimension is not regulated because we observe that most of the changes in objects' sizes are along the vertical dimension. An instance of this cell structure is shown in Fig. 3.

As already mentioned at the start of this section that a word is defined by its location, motion and size features. The location of a word $\mathbf{w}_i$ is defined by the index of the cell where a word lies, i.e. represented as l_{w_i}. This is a two dimensional index which shows the horizontal and vertical location of the centre of a cell.

2) SIZE FEATURE

To capture sizes of objects in a video frame, we employ background subtraction on a video frame. Background subtraction algorithm results in blobs of foreground pixels in a video frame. However, we are interested in calculating the foreground pixels contained by a cell in a video frame. Size of word $\mathbf{w}_i$ located in cell c is then defined as $s^c_{w_i}$. This is calculated as

$$s^{c}_{w_i} = \frac{1}{N}\sum_{n=1}^{N_p} u^{(n)} \tag{6}$$

where N_p is total number of pixels in a cell and $u^{(n)}$ is defined as

$$u^{(n)} = \begin{cases} 1 & \text{if } u^{(n)} \text{ is a foreground pixel} \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

A word with cell location l_{w_i} is considered as an active cell if more than 10% of its pixels are foreground pixels. Only active cells are considered in the proposed algorithm for further processing to detect anomalous activities.

3) MOTION FEATURES

To detect the foreground pixels in a video frame we have already employed background subtraction and have obtained foreground pixels. To extract motion features of a word, we first apply optical flow on all the obtained foreground pixels. For the proposed algorithm we have used Lucas-Kanade algorithm [37] to obtain horizontal and vertical motion of foreground pixels. Once we have optical flow information of all the foreground pixels, we calculate the optical flow energy of a word $\mathbf{w}_i$ located in cell c as [38]

$$O^{c}_{w_i} = \frac{1}{N_f}\sum_{n=1}^{N_f} \left|\left|[v^n_x, v^n_y]\right|\right|_2 \tag{8}$$

where v_x and v_y correspond to the horizontal and vertical optical flow components, respectively, obtained through Lucas-Kanade algorithm and N_f is the no of foreground pixels in cell c.

To obtain the HOF descriptor H_{w_i} of word $\mathbf{w}_i$ located in cell c we use a 5-bin optical flow histogram calculated in the range $[0, \frac{4}{5\pi}]$. The histograms are normalized using L1 normalization.

Optical flow energy feature $O^c_{w_i}$ and HOF descriptor H_{w_i} jointly define motion features of the visual word $\mathbf{w}_i$.

B. VOCABULARY OF WORDS

Size of the vocabulary of words in the proposed anomaly detection algorithm is defined as the cartesian product of the location, motion, and size features in word spaces. Vocabulary of words is defined as $V = \{\mathbf{W}_i\}_{i=1}^{N_v}$, where N_v is the number of words in the vocabulary, and as described earlier each word is four dimensional, and the *ith* word in the vocabulary is defined as $\mathbf{W}_i = \{\mathbf{w}^r_i\}_{r=1}^{N_w}$.

C. MODEL FORMATION

In the proposed algorithm, we define $\mathbf{D} = \{\mathbf{d}_j\}_{j=1}^{N_d}$ as a set of all the documents (video clips), where N_d is the number of documents in a complete video sequence. Each document $\mathbf{d}_j$ is an observation in the proposed algorithm, which is a collection of certain visual words. Set of topics are represented as $\boldsymbol{\tau} = \{\tau_k\}_{k=1}^{N_z}$, where, N_z is the total number of topics. The state vector at time step t which represents state of these topics is defined as $\mathbf{z}_t = \{z^k_t\}_{k=1}^{N_z}$, where, $z^k_t \in \{0, 1\}$, it is equal to 1 if an activity is observed. These are the topics which are being observed in the training data set. The probability over $\mathbf{z}_t$ and the observed document d is defined as

$$\begin{aligned} p(\mathbf{z}_t, d) &= p(\mathbf{d})p(\mathbf{z}_t|\mathbf{d}) \\ &= p(\mathbf{d}) \sum_{\mathbf{w}_l \in \mathbf{W}_d} p(\mathbf{z}_t, \mathbf{w}_l|\mathbf{d}) \\ &= p(\mathbf{d}) \sum_{\mathbf{w}_l \in \mathbf{W}_d} p(\mathbf{z}_t|\mathbf{w}_l, \mathbf{d})p(\mathbf{w}_l|\mathbf{d}) \end{aligned} \tag{9}$$

where, d is the *ith* document in a complete video sequence, however we have removed the subscript i to improve readability. Similarly, $\mathbf{W}_d$ is the set of words observed in *ith* document. In the proposed algorithm our objective is to estimate $p(\mathbf{z}_t|\mathbf{W}_d, \mathbf{d})$ which is the probability of the state vector $\mathbf{z}_t$ given a set of words $\mathbf{W}_d$ which are being observed in document d, whereas, other probabilities that are $p(\mathbf{d})$ and $p(\mathbf{w}_l|\mathbf{d})$ can easily be calculated by using a training data. In the proposed anomaly detection algorithm, we propose a particle filtering based novel technique to estimate the posterior probability $p(\mathbf{z}_t|\mathbf{W}_d, \mathbf{d})$ given a video document and observed visual words. In the next section we explain that how a particle filtering based algorithm is developed to estimate this posterior distribution. However, from now onwards, we would write this probability as $p(\mathbf{z}_t|\mathbf{W}_d)$, where we assume that the document $\mathbf{d}$ is known and observed words $\mathbf{W}_d$ are from the document $\mathbf{d}$.

D. PARTICLE FILTERING BASED ANOMALY DETECTION ALGORITHM

Particle filtering [39] based anomaly detection method is proposed to approximate the posterior probability of current state $\mathbf{z}_t$ given a clip of recorded video. For a video frame in a document $\mathbf{d}$ at time t. Particle filter estimates this current state of the state vector with the help of given past state $\mathbf{z}_{t-1}$ and current observed words $\mathbf{w}_t$ in frame at time t in a document $\mathbf{d}$. A particle filter basically estimates a posterior distribution of the current state of state vector given its previous state at time $t-1$ and current observations at time t. Therefore, in the proposed algorithm, we are interested in estimating the posterior distribution denoted as $p(\mathbf{z}_t|\mathbf{z}_{t-1}, \mathbf{w}_t)$, on the basis of the defined state transition model and an observations model. The state transition model is generally represented as [40], [41].

$$\mathbf{z}_t = f_t(\mathbf{z}_{t-1}, \mathbf{v}_{t-1}), \qquad p(\mathbf{z}_t|\mathbf{z}_{t-1}) \tag{10}$$

where, f_t is the nonlinear function of the state $\mathbf{z}_{t-1}$ and $\mathbf{v}_{t-1}$ is system noise at time t, whereas, the measurement model can typically be represented as [40], [41]

$$\mathbf{w}_t = h_t(\mathbf{z}_t, \mathbf{n}_t), \qquad p(\mathbf{w}_t|\mathbf{z}_t) \tag{11}$$

where h_t is a nonlinear function of the state $\mathbf{z}_t$ and $\mathbf{n}_t$ is measurement noise at time t. The basic idea in particle filtering is that this filtering technique represents the posterior distribution of a variable with the help of a set of random samples known as particles and their associated weights. Particle filtering technique estimates the posterior distribution based on these samples and their associated weights [41]. In particle filtering, when the number of samples become very large, the estimated posterior distribution becomes equivalent to the usual functional description of the posterior distribution [41]. Initially, all the particles are randomly drawn from the prior density and then passed through the state transition model represented as (10). Weights to these particles are then assigned with the help of a measurement model as described in (11). These assigned weights define the worth of particles. Higher the worth of particle, larger the assigned weight it has, and lower the worth of particle, smaller the assigned weight it has [40]. Particle filter approach mainly comprises of two steps: Prediction and update.

At every time step t the prediction step predicts N_p number of particles with the help of previous state at time $t-1$ and a state transition model. This is an approximation of prior probability density function. The update step assigns weights to these particles by using latest measurements and a measurement model. Further, at the update stage, posterior distribution of the state of unknown variable is calculated by using predicted particles and their associated weights. In the proposed algorithm the posterior density of variable $\mathbf{z}_t$ at every time step t is calculated as [41]

$$p(\mathbf{z}_t|\mathbf{w}_t) \approx \frac{1}{N_p}\sum_{i=1}^{N_p} \boldsymbol{\eta}_t^i \delta(\mathbf{z}_t - \mathbf{z}_t^i) \tag{12}$$

where, $\mathbf{z}_t^i$ is the ith particle at time t, $\boldsymbol{\eta}_t^i$ is the weight associated with the ith particle at time t and δ is the dirac delta function. As already mentioned that particles $\{\mathbf{z}_t^i\}_{i=1}^{N_p}$ are predicted by using a prediction model and associated weights $\{\boldsymbol{\eta}_t^i\}_{i=1}^{N_p}$ are calculated by using a measurement model. Next we explain the proposed prediction and measurement models.

1) PREDICTION MODEL

Particle filtering is a Markov model which estimates next state of a state variable at time t with help of its state at time $t-1$. At every time step t the prediction step predicts N_p number of particles with the help of previous state at time $t-1$ and a state transition model. This process can be viewed as taking N_p samples from an approximation of the prior probability density function. The process can be represented as

$$\mathbf{z}_t^i \sim q(\mathbf{z}_t|\mathbf{z}_{t-1}^i), \tag{13}$$

where $q(\cdot)$ is an approximation of the prior probability of state variable $\mathbf{z}_t$, which is evaluated by our prediction model explained below.

The standard particle filter uses the state dynamic model to sample particles for the current state, it then utilizes current observations to calculate the importance weight of these samples. However, in problems such as anomaly detection, the state dynamics are not strong and the current observations give significant information to estimate the current state. Therefore, sampling based on current observations give better sampling results compared to sampling through a state dynamic model. [42].

In this paper the proposed particle filtering based anomaly detection framework uses a variant of the PF which draws samples of the current state from a sampling distribution which is assumed to be conditionally independent of the previous states.

Most of the normal activities are area specific in a video scene, for instance, in a scene shown in Fig. 2 activities related to pedestrians can only be in the regions which represent a footpath, while activities related to vehicles can only be found on road. Therefore, every activity is directly related to the location in a scene. We exploit this information to evaluate prior probability of activity.

After background subtraction we identify active cells, a cell is marked as active if more than 10% of its pixels are foreground pixels. Given the location of active cell, we evaluate prior probability of all the activities i.e. $p(\tau_k|C_l)$, which is calculated as [33]

$$p(\tau_k|C_l) = \sum_{w\in V_c} p(w|\tau_k) \tag{14}$$

where, V_c a set of visual words that appear in cell C_k in the dictionary of training data set, and therefore, $p(w|\tau_k)$ is evaluated by using a training data set. Prior probability of all other active cells is evaluated in a similar fashion and overall prior probability of an activity is a product of these probabilities

$$p(\tau_k) = \prod_{C_l\in\boldsymbol{\xi}} p(\tau_k|C_l), \tag{15}$$

where ξ is a set of all the active cells. The proposed prediction model takes into account this prior probability of activity and its weight in the previous sate at time $t-1$ to evaluate the probability of the activity as a product of both. We predict an activity τ_k will occur at t if its prior probability is greater than a threshold. In the proposed work we concatenate state of all the activity to estimate a particles $\mathbf{z}_{t-1}^i$.

2) MEASUREMENT MODEL

To evaluate weights $\boldsymbol{\eta}$ of all the particles, we use training data as well as current observations at time t. As a first step, we create a dictionary of all the words found in a training data set. In the video based anomaly detection problem under discussion, we know that every activity can generates multiple words. Therefore, in the proposed anomaly detection algorithm we present a technique that groups set of all the words in the training data set into clusters with the help of a clustering technique. After the clustering process we get κ clusters, where κ is equal to the number of activities. We assume that the number of clusters is predefined and known. Clusters are regions represented as $\{\boldsymbol{\vartheta}_i\}_{i=1}^{\kappa}$, where each cluster is a group of words originated from same activity (topic). The clustering process in the proposed research aims to assign words in the training data to the respective clusters and gives as output a correspondence matrix $\mathbf{B} = \{[(\mathbf{b}^j)^T)]^T\}_{j=1}^{\tilde{N}_w}$, where $\mathbf{b}^j = [b^{j,1}, b^{j,2}, \ldots, b^{j,q}, \ldots, b^{j,\kappa}]$ indicates, to which cluster, the word vector $\tilde{\mathbf{w}}^j$ corresponds, where $\{\tilde{\mathbf{w}}^j\}_{j=1}^{\tilde{N}_w}$ is the set of words in a training data set. All but one of the elements of $\mathbf{b}^j$ is zero, if e.g. $\tilde{\mathbf{w}}^j$ belongs to the q^{th} cluster then the correspondence vector will be, $\mathbf{b}^j = [0, \ldots, 0, 1, 0, \ldots, 0]$, which shows only the q^{th} element of $\mathbf{b}^j$ is nonzero.

The clustering process aims to group $\tilde{N}_w$ words in the training data into κ clusters. Clusters are represented by their centers $\{\boldsymbol{\mu}^q\}_{q=1}^{\kappa}$, where a binary indicator $b^{j,q} \in \{0, 1\}$ represents to which of the q^{th} clusters word $\tilde{\mathbf{w}}^j$ is assigned; for example, when the data point $\tilde{\mathbf{w}}^j$ is assigned to the q^{th} cluster then $b^{j,q} = 1$, and $b^{j,l} = 0$ for $l \neq q$.

We assume that the appearance of words in a dictionary are modeled by a mixture of Gaussian distributions. Therefore, the clustering task can be viewed as fitting mixtures of Gaussian distributions [43] to the visual words in a dictionary. Every cluster $\boldsymbol{\vartheta}^q$ of visual words is assumed to be modeled by a Gaussian distribution, $\mathcal{N}(\tilde{\mathbf{w}}^j|\boldsymbol{\mu}^q, \Sigma^q)$ which is one component of the mixture of Gaussian distributions. Each cluster has a mean vector $\boldsymbol{\mu}^q$ and a covariance matrix $\boldsymbol{\Sigma}^q$. Therefore, the probability of a visual word $\tilde{\mathbf{w}}^j$ can be represented as [44]

$$p(\tilde{\mathbf{w}}^j|\boldsymbol{\pi}, \boldsymbol{\mu}, \Sigma) = \sum_{q=1}^{\kappa} \boldsymbol{\pi}^q \mathcal{N}(\tilde{\mathbf{w}}^j|\boldsymbol{\mu}^q, \Sigma^q), \tag{16}$$

where $\boldsymbol{\mu} = \{\boldsymbol{\mu}^q\}_{q=1}^{\kappa}$ and $\boldsymbol{\Sigma} = \{\boldsymbol{\Sigma}^q\}_{q=1}^{\kappa}$. The mixing coefficient vector is defined as $\boldsymbol{\pi} = \{[\boldsymbol{\pi}^q]^T\}_{q=1}^{\kappa}$, where $\boldsymbol{\pi}^q$ represents the probability of selecting the q^{th} component of the mixture which is the probability of assigning the j^{th} visual word to the q^{th} cluster. If we assume that all the visual words are independent, then the joint log likelihood function of all the visual words in a dictionary $\tilde{\mathbf{W}}$ becomes [43]

$$\ln p(\tilde{\mathbf{W}}|\boldsymbol{\pi}, \boldsymbol{\mu}, \Sigma) = \sum_{j=1}^{\tilde{N}_w} \ln\left\{\sum_{q=1}^{\kappa} \boldsymbol{\pi}^q \mathcal{N}(\tilde{\mathbf{w}}^j|\boldsymbol{\mu}^q, \Sigma^q)\right\}. \tag{17}$$

There are a number of techniques in the literature to evaluate the optimum values of the parameters to solve this clustering problem, such as [44]. However, in the proposed anomaly detection algorithm we have used expectation maximization (EM) optimization to evaluate these parameters to perform clustering in dictionary of visual words. The EM algorithm returns optimum values of these parameters which maximises the log likelihood function given in (17).

Once we obtain clusters from the dictionary of visual words, the next step in the proposed work is to assign clusters to the visual words in the test data. Every visual word $\mathbf{w}^j$ in the test data can be assigned to cluster $\boldsymbol{\vartheta}^q$ with probability

$$p(\mathbf{w}^j \in \boldsymbol{\vartheta}^q|\boldsymbol{\mu}^q, \Sigma^q) = \mathcal{N}(\mathbf{w}^j|\boldsymbol{\mu}^q, \Sigma^q). \tag{18}$$

However, we assign every visual word $\mathbf{w}^j$ to a cluster $\mathbf{C}^q$ which has a maximum probability of assignment according to (18), this is described as

$$p_{max}(\mathbf{w}^j \in \boldsymbol{\vartheta}^q|\boldsymbol{\mu}^q, \Sigma^q) = \max_q \{p(\mathbf{w}^j \in \boldsymbol{\vartheta}^q|\boldsymbol{\mu}^q, \Sigma^q)\}. \tag{19}$$

We assign word $\mathbf{w}^j$ to a cluster $\boldsymbol{\vartheta}^q$ with probability $p_{max}(\mathbf{w}^j \in \boldsymbol{\vartheta}^q|\boldsymbol{\mu}^q, \Sigma^q)$. We repeat this process for all the visual words appearing in a document in a test data set. As we have already mentioned that a cluster of visual words basically represent an activity, therefore, probability $p(\mathbf{w}^j|\mathbf{z}_t^i)$ is calculated as

$$p(\mathbf{w}^j|\mathbf{z}_t^i) = p_{max}(\mathbf{w}^j \in \boldsymbol{\vartheta}^q|\boldsymbol{\mu}^q, \Sigma^q). \tag{20}$$

However, to evaluate this probability we consider only those activities which are predicted to be happening at time t by a particle $\mathbf{z}_t^i$. Hence, the joint probability of all the words given particle $\mathbf{z}_t^i$ is calculated as

$$p(\mathbf{W}|\mathbf{z}_t^i) = \prod_{j=1}^{N_w} p(\mathbf{w}^j|\mathbf{z}_t^i) \tag{21}$$

where, we assume that all the visual words are independent and identically distributed. This information is used to evaluate weights of particles as described below.

To evaluate the weight $\boldsymbol{\eta}_t^i$ of ith particle $\mathbf{z}_t^i$ at time step t in particle filtering, we use (21), such that

$$\boldsymbol{\eta}_t^i = p(\mathbf{W}|\mathbf{z}_t^i) \tag{22}$$

3) UPDATE STEP

Once we have all the particles $\{\mathbf{z}_t^i\}_{i=1}^{N_p}$ and their associated weights $\{\boldsymbol{\eta}_t^i\}_{i=1}^{N_p}$, we use (12) to evaluate the posterior distribution of the state at time t i.e. $p(\mathbf{z}_t|\mathbf{w}_t)$.

E. ANOMALY INFERENCE

To detect anomalous video frames we keep a check on the posterior probability $p(\mathbf{z}_t|\mathbf{w}_t)$. If he posterior probability of $\mathbf{z}_t$ is less than a empirically calculated threshold ϵ_o we mark that video frame as an anomalous video frame. In the video frame detected as anomalous, we further evaluate the anomalous regions. Regions in a video frame related to visual words which has likelihood less than a empirically calculated threshold ϵ_r are marked as anomalous regions in anomalous video frame. Likelihood of visual words is calculated by using (20).

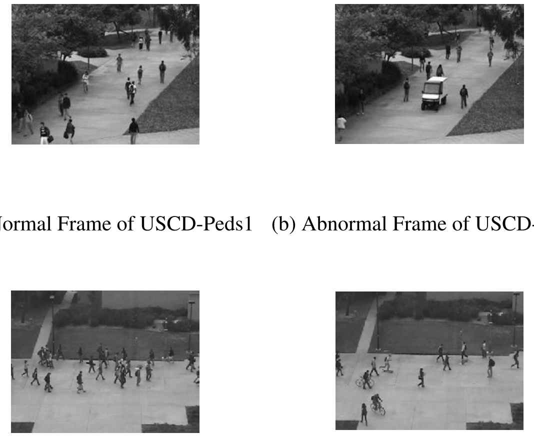
IV. PERFORMANCE EVALUATION

A. DATASETS

The performance of our proposed method is evaluated by using two different datasets: UCSD [45] and live datasets [46].

1) UCSD DATASET

This dataset is divided into two distinct scenes. First one is UCSD-Peds1 and second one is UCSD-Peds2. This dataset contains videos which capture a scene of a crowded pedestrian area. The area is restricted for vehicles, therefore, any vehicles, cyclists or skaters found moving in the area should be considered as an anomaly. Videos in the UCSD-Peds1 data



(a) Normal Frame of USCD-Peds1 (b) Abnormal Frame of USCD-Peds1

(c) Normal Frame of USCD-Peds2 (d) Abnormal Frame of USCD-Peds2

FIGURE 4. **Normal and Anomalous Frame of UCSD are shown. In (b) presence of a car makes this frame anomalous, while in (d) cyclist makes it anomalous.**

set are captured from two different angles where, camera is fixed with no lighting variations. This data set contains 98 video sequences in total, where each video sequence consists of 200 video frames. UCSD-Peds1 comprises of 70 video sequences. For the training phase we have used 34 video sequences while 36 of them are used for testing purpose. Some of the normal and abnormal frames from the dataset UCSD-Peds1 are shown in Figs. 4-a and 4-b. UCSD-Peds2 contains 28 videos sequences. For training we have used 16 of these sequence, while 12 of them are used for testing. The ground truth data for both scenes is provided to evaluate the model. Some of the normal and abnormal frames from datasets of UCSD Peds2 are shown in Figs. 4-c and 4-d.

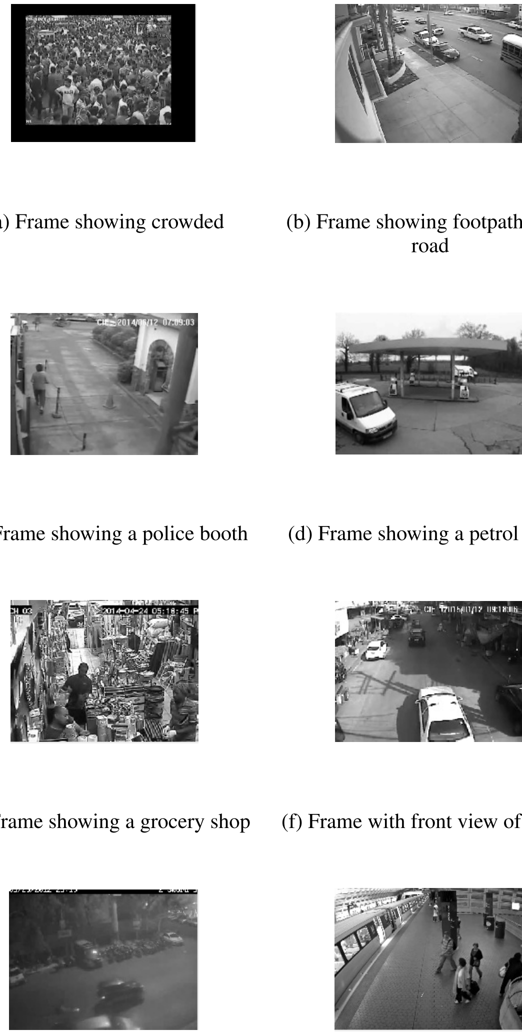
2) LIVE DATASET

It is basically a live dataset of video sequences of a real world scenario which is captured without planned movements of objects. This dataset contains 28 real videos sequences, where each video sequence consists of 300 video frames. These sequences are more challenging because of varying illumination conditions and camera movements. The sequences are captured in daylight as well as at the night time and contain different levels of crowd. The sequences also contain both normal and abnormal events and ground truths are also provided. Few of the video frames from this data set are shown in Fig 5.

B. EXPERIMENTAL SETUP

Performance of the proposed particle filtering based anomaly detection framework is evaluated by comparing its performance with several algorithms presented in [2], [5], [7], [9], [26], [31], [47], [49] which include both online and offline anomaly detections algorithms.

To compare the proposed algorithm with [31], values of different parameters are empirically chosen and kept same



(a) Frame showing crowded (b) Frame showing footpath and a road

(c) Frame showing a police booth (d) Frame showing a petrol pump

(e) Frame showing a grocery shop (f) Frame with front view of a road

(g) Frame captured in bad illumination conditions (h) Frame showing a subway station

FIGURE 5. **Normal video frames extracted from the LIVE dataset.**

while calculating evaluation results of the algorithms used for comparison. In the proposed algorithm, value of the growing rate (α) used in (1), is chosen from $\alpha \in [1.06, 1.2]$. Thresholds ϵ_o and ϵ_r used in Section III-E are chosen to be 0.6 and 0.4 respectively. All the results are calculated with the number of particles $N_p = 100$

C. PERFORMANCE METRIC

The performance metric used to evaluate the framework are Equal Error Rate (EER) and Area Under the Curve (AUC). Both are related to each other in a way that if EER approaches 1, the AUC approaches 0. ROC curve is plotted for determining the AUC value between True Positive Rate (TRP) and False Positive Rate (FPR). In order to get the EER value, a straight line is plotted between (1, 0) and (0, 1) on the ROC curve. The point at which both graphs intersect is used. The FPR value of that point is taken as the *EER* value. In our

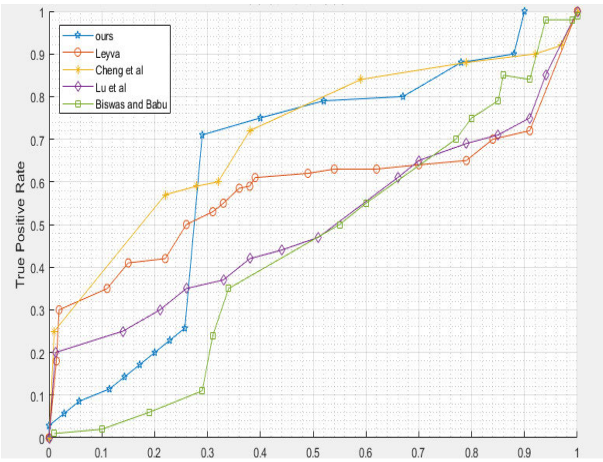


FIGURE 6. ROC Curve of Peds1.

TABLE 1. EER for the Peds1 scene of the UCSD dataset.

Author	EER value	Frame Processing Time(ms)	Performance
Javan and Levine [49]	27	190	offline
Hu et al. [5]	36	200	offline
Cheng et al. [26]	38.8	1100	offline
Cong et al. [7]	51.2	3800	offline
Zhu et al. [47]	-	4600	offline
Lu et al. [9]	59.1	6	online
Biswas and Babu [48]	50.95	14	online
Leyva et al. [31]	39.7	31	online
Y. Feng et al [12]	28.2	189.8	online
Nawaratne et al [23]	29.8	36.9	online
Proposed Algorithm	29	27	online

method, a frame is considered as abnormal if the estimated posterior probability of activities is less then the threshold ϵ_o. Video frames correctly detected as abnormal are referred to as true positive, while video frames wrongly detected as abnormal refer to as false positives.

D. RESULTS

The ROC curves showing the performance of the proposed particle filtering based algorithm and its comparison with other state-of-the-art are shown in Fig. 6. These ROCs are ploted for UCSD datasets.

Comparison of the EER value and processing time of the proposed with other state-of-the art are presented in Table 1. It can bee seen in Fig. 6 and Table 1 that the proposed technique shows better performance compared to the other state-of-the-art online anomaly detection algorithm. Proposed technique gives comparatively better detection accuracy and has smaller processing time. For offline methods, EER values are low, however, frame processing time is high. In offline case, the best method is [2] as it contains low EER value. In online case, our proposed particle filtering based algorithm gives the best EER value compared to all other online methods. It can also be seen that the processing time of the proposed algorithm is lower than all other techniques except those of [21] and [20], which have much higher EER value. Results show that the performance in terms of accuracy of the proposed anomaly detection framework is comparable to offline methods, whereas, processing speed is much higher and is capable of online detection. Deep learning based technique [12] has an EER almost similar to the proposed technique, however, its processing time is much higher.

(a) Anomalous Frame containing a car

(b) Anomalous Frame containing a cyclist

(c) Anomalous Frame containing a skater

(d) Anomalous Frame containing a motorcyclist

FIGURE 7. Abnormal Frames from the UCSD Peds1 datasets.

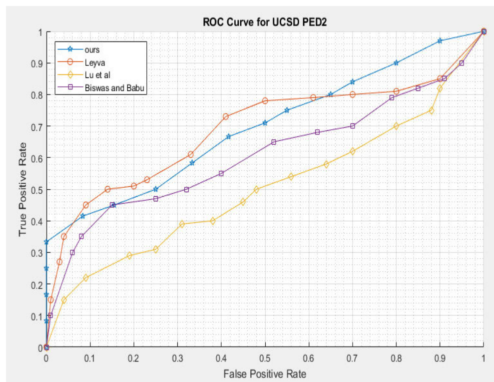


FIGURE 8. ROC Curve of Peds2.

Few of the frames of USCD-Peds1 dataset in which anomalous regions have been correctly detected by our proposed algorithm are shown in Fig. 7. It can be seen in this figure that anomalous activities, such as motion of a vehicle, a cycle, a skate board and a motorcycle in pedestrian area have been correctly detected.

Results on the UCSD-Peds2 data set are shown in Fig. 8 and Table. 2. Again, the proposed algorithm shows better performance compared to the other online and offline algorithms proposed in the literature.

Few of the frames of USCD-Peds2 dataset in which anomalous regions have been correctly detected by our proposed algorithm are shown in Fig. 9. It can be seen in this figure that

TABLE 2. EER for the PEDS scene of the UCSD dataset.

Author	EER value	Frame Processing Time ms	Performance
Javan and Levine [49]	26	220	offline
Hu et al. [5]	-	200	offline
Li et al. [21]	-	1100	offline
Lu et al. [9]	49.8	6.1	online
Biswas and Babu [48]	42.3	12.5	online
Leyva et al. [31]	36.6	31	online
Y. Feng et al [12]	27.5	191.8	online
Nawaratne et al [23]	29.8	38.1	online
Proposed Algorithm	29.5	25	online

(a) Anomalous Frame containing a cyclist

(b) Anomalous Frame containing a cyclist and a car

(c) Anomalous Frame containing a skater and two cyclists

(d) Anomalous Frame containing a cyclists

FIGURE 9. Abnormal Frames from the UCSD Peds2 dataset.

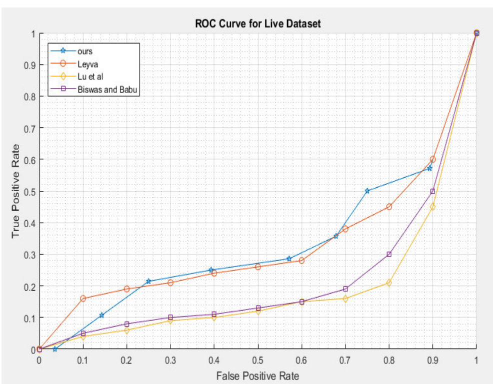


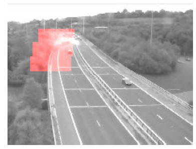
FIGURE 10. ROC Curve of LIVE dataset.

anomalous activities, have been correctly detected. The ROC curves on the LIVE dataset are shown in Fig. 10, whereas, Area under the curve (AUC) comparisons on LIVE dataset are given in the Table.3. In this case, our proposed methods outperforms all other techniques. It give larger AUC value with very low processing time. In our method we take into consideration the true positives only when the proposed technique successfully detects the Region Of Interest (ROI) otherwise it is taken as the false negative. Lowest value of AUC for [9] and [49] is due to the fact that they use motion vectors and

TABLE 3. AUC for the live dataset.

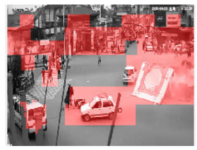
Author	AUC	Frame Processing Time ms
Lu et al. [9]	0.112	6.8
Biswas and Babu [48]	0.151	13.2
Leyva et al. [31]	0.278	32.5
Y. Feng et al [12]	0.31	185.9
Nawaratne et al [23]	0.29	39.3
Proposed Algorithm	0.32	27.3

(a) Anomalous Frame representing vandalism



(b) Anomalous Frame representing a car accident on highway

(c) Anomalous Frame representing vehicle in a pedestrian area



(d) Anomalous Frame representing earth-quack

(e) Anomalous Frame representing a street crime

FIGURE 11. Abnormal Frames from the LIVE dataset.

background subtraction in order to collect features. Although the frame processing time of both state-of-the-art methods is small, their detection accuracy is low. The AUC value of [31] is also very low as compared to our proposed method.

Few of the frames of LIVE dataset in which anomalous regions have been correctly detected by our proposed algorithm are shown in Fig. 11. It can be seen in this figure that anomalous activities, have been correctly detected.

Thus, our proposed method detects anomalous regions in anomalous video frames. The designed experiments showed the effectiveness and efficiency of the proposed method comparing it with previous methods.

V. CONCLUSION

A particle filtering based anomaly detection framework has been presented for online video surveillance. Size, motion and location features have been used to develop novel prediction and measurement models to estimate posterior probability of activities in video sequences. Based on this estimated posterior probability distribution, the proposed anomaly detection accurately detects video frames with anomalous activities. It further detects anomalous regions within anomalous video frames. The proposed framework is tested on UCSD and LIVE datasets and results are compared with the existing state-of-the-art algorithms in the literature. The comparative analysis demonstrates that the proposed anomaly detection algorithm outperforms the state-of-the art online anomaly detection algorithms in terms of EER, AUC and the processing time.

REFERENCES

[1] H. Lin, J. D. Deng, B. J. Woodford, and A. Shahi, "Online weighted clustering for real-time abnormal event detection in video surveillance," in *Proc. 24th ACM Int. Conf. Multimedia*, New York, NY, USA, Oct. 2016, pp. 536–540.

[2] Y. Zhang, H. Lu, L. Zhang, and X. Ruan, "Combining motion and appearance cues for anomaly detection," *Pattern Recognit.*, vol. 51, pp. 443–452, Mar. 2016.

[3] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis, "Learning temporal regularity in video sequences," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 733–742.

[4] Y. S. Chong and Y. H. Tay, "Abnormal event detection in videos using spatiotemporal autoencoder," in *Advances in Neural Networks*. Cham, Switzerland: Springer, 2017.

[5] W. Liu, W. Luo, D. Lian, and S. Gao, "Future frame prediction for anomaly Detection–A new baseline," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6536–6545.

[6] W. Luo, W. Liu, and S. Gao, "A revisit of sparse coding based anomaly detection in stacked RNN framework," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 341–349.

[7] Y. Cong, J. Yuan, and J. Liu, "Sparse reconstruction cost for abnormal event detection," in *Proc. CVPR*, Jun. 2011, pp. 3449–3456.

[8] B. Zhao, L. Fei-Fei, and E. P. Xing, "Online detection of unusual events in videos via dynamic sparse coding," in *Proc. CVPR*, Jun. 2011, pp. 3313–3320.

[9] C. Lu, J. Shi, and J. Jia, "Abnormal event detection at 150 FPS in MATLAB," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2720–2727.

[10] S.-H. Cho and H.-B. Kang, "Abnormal behavior detection using hybrid agents in crowded scenes," *Pattern Recognit. Lett.*, vol. 44, pp. 64–70, Jul. 2014.

[11] R. Hinami, T. Mei, and S. Satoh, "Joint detection and recounting of abnormal events by learning deep generic knowledge," 2017, *arXiv:1709.09121*. [Online]. Available: http://arxiv.org/abs/1709.09121

[12] Y. Feng, Y. Yuan, and X. Lu, "Learning deep event models for crowd anomaly detection," *Neurocomputing*, vol. 219, pp. 548–556, Jan. 2017.

[13] P. Perera, R. Nallapati, and B. Xiang, "OCGAN: One-class novelty detection using GANs with constrained latent representations," 2019, *arXiv:1903.08550*. [Online]. Available: http://arxiv.org/abs/1903.08550

[14] Y. Tang, Y. Tang, J. Xiao, R. M. Summers, and M. Han, "Deep adversarial one-class learning for normal and abnormal chest radiograph classification," *Proc. SPIE Med. Imag., Comput.-Aided Diagnosis*, vol. 10950, Mar. 2019, Art. no. 1095018.

[15] N. Tuluptceva, B. Bakker, I. Fedulova, and A. Konushin, "Perceptual image anomaly detection," in *Proc. Asian Conf. Pattern Recognit.*, 2020, pp. 164–178.

[16] K. Ouardini, H. Yang, B. Unnikrishnan, M. Romain, C. Garcin, H. Zenati, J. P. Campbell, M. F. Chiang, J. Kalpathy-Cramer, V. Chandrasekhar, P. Krishnaswamy, and C.-S. Foo, "Towards practical unsupervised anomaly detection on retinal images," in *Proc. Int. Workshop Med. Image Learn. Less Labels Imperfect Data*, 2019, pp. 225–234.

[17] D. Abati, A. Porrello, S. Calderara, and R. Cucchiara, "Latent space autoregression for novelty detection," 2018, *arXiv:1807.01653*. [Online]. Available: http://arxiv.org/abs/1807.01653

[18] R. Tibor Schirrmeister, Y. Zhou, T. Ball, and D. Zhang, "Understanding anomaly detection with deep invertible networks through hierarchies of distributions and features," 2020, *arXiv:2006.10848*. [Online]. Available: http://arxiv.org/abs/2006.10848

[19] K. P. Adhiya, S. R. Kolhe, and S. S. Patil, "Tracking and identification of suspicious and abnormal behaviors using supervised machine learning technique," in *Proc. Int. Conf. Adv. Comput., Commun. Control*, 2009, pp. 96–99.

[20] C. He, J. Shao, and J. Sun, "An anomaly introduced learning method for abnormal event detection," *Multimedia Tools Appl.*, vol. 4, pp. 29573–29588, Nov. 2018.

[21] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6479–6488.

[22] M. Mathieu, C. Couprie, and Y. LeCun, "Deep multi-scale video prediction beyond mean square error," *CoRR*, vol. abs/1511.05440, Nov. 2015.

[23] R. Nawaratne, D. Alahakoon, D. De Silva, and X. Yu, "Spatiotemporal anomaly detection using deep learning for real-time video surveillance," *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 393–402, Jan. 2020.

[24] W. Li, V. Mahadevan, and N. Vasconcelos, "Anomaly detection and localization in crowded scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 1, pp. 18–32, Jan. 2014.

[25] N. Li, X. Wu, H. Guo, D. Xu, Y. Ou, and Y.-L. Chen, "Anomaly detection in video surveillance via Gaussian process," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 29, no. 06, Sep. 2015, Art. no. 1555011.

[26] K.-W. Cheng, Y.-T. Chen, and W.-H. Fang, "Video anomaly detection and localization using hierarchical feature representation and Gaussian process regression," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2909–2917.

[27] J. Kim and K. Grauman, "Observe locally, infer globally: A space-time MRF for detecting abnormal activities with incremental updates," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 2921–2928.

[28] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 935–942.

[29] T. Hospedales, S. Gong, and T. Xiang, "A Markov clustering topic model for mining behaviour in video," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 1165–1172.

[30] L. Kratz and K. Nishino, "Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1446–1453.

[31] R. Leyva, V. Sanchez, and C.-T. Li, "Video anomaly detection with compact feature sets for online performance," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3463–3478, Jul. 2017.

[32] X. Gao, G. Xu, S. Li, Y. Wu, E. Dancigs, and J. Du, "Particle filter-based prediction for anomaly detection in automatic surveillance," *IEEE Access*, vol. 7, pp. 107550–107559, 2019.

[33] J. Varadarajan and J.-M. Odobez, "Topic models for scene analysis and abnormality detection," in *Proc. IEEE 12th Int. Conf. Comput. Vis. Workshops, ICCV Workshops*, Sep. 2009, pp. 1338–1345.

[34] V. Reddy, C. Sanderson, and B. C. Lovell, "Improved anomaly detection in crowded scenes via cell-based analysis of foreground speed, size and texture," in *Proc. CVPR Workshops*, Jun. 2011, pp. 55–61.

[35] A. B. Chan, Z.-S. John Liang, and N. Vasconcelos, "Privacy preserving crowd monitoring: Counting people without people models or tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–7.

[36] R. Leyva, V. Sanchez, and C. Li, "Video anomaly detection based on wake motion descriptors and perspective grids," in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Dec. 2014, pp. 209–214.

[37] S. Baker and I. Matthews, "Lucas-kanade 20 years on: A unifying framework," *Int. J. Comput. Vis.*, vol. 56, no. 3, pp. 221–255, Feb. 2004.

[38] N. Dalal, B. Triggs, and C. Schmid, "Human detection using oriented histograms of flow and appearance," in *Proc. ECCV*, Berlin, Germany, 2006, pp. 428–441.

[39] Y. Rui and Y. Chen, "Better proposal distributions: Object tracking using unscented particle filter," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Dec. 2001, p. 2.

[40] K. K. Ng and E. J. Delp, "New models for real-time tracking using particle filtering," *Vis. Commun. Image Process.*, vol. 7257, Jan. 2009, Art. no. 72570B.

[41] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Dec. 2002.

[42] M. T. Lin, J. L. Zhang, Q. Cheng, and R. Chen, "Independent particle filters," *J. Amer. Stat. Assoc.*, vol. 100, no. 472, pp. 1412–1421, Dec. 2005.

[43] C. M. Bishop, *Pattern Recognition Machine Learning*. Cham, Switzerland: Springer, 2006.

[44] A. Ur-Rehman, S. M. Naqvi, L. Mihaylova, and J. A. Chambers, "Multitarget tracking and occlusion handling with learned variational Bayesian clusters and a social force model," *IEEE Trans. Signal Process.*, vol. 64, no. 5, pp. 1320–1335, Mar. 2016.

[45] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 1975–1981.

[46] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, 2006, doi: 10.1109/TIP.2006.881959.

[47] X. Zhu, J. Liu, J. Wang, C. Li, and H. Lu, "Sparse representation for robust abnormality detection in crowded scenes," *Pattern Recognit.*, vol. 47, no. 5, pp. 1791–1799, May 2014.

[48] S. Biswas and R. V. Babu, "Real time anomaly detection in H.264 compressed videos," in *Proc. 4th Nat. Conf. Comput. Vis., Pattern Recognit., Image Process. Graph. (NCVPRIPG)*, Dec. 2013, pp. 1–4.

[49] M. J. Roshtkhari and M. D. Levine, "An on-line, real-time learning method for detecting anomalies in videos using spatio-temporal compositions," *Comput. Vis. Image Understand.*, vol. 117, no. 10, pp. 1436–1452, 2013.

ATA-UR-REHMAN received the B.Eng. degree in electronic engineering from Air University, Islamabad, in 2006, and the M.Sc. (Hons.) and Ph.D. degrees from Loughborough University, U.K., in 2010 and 2014, respectively. He then joined the University of Engineering and Technology, Lahore, as a Lecturer, in 2007. He moved to Loughborough University, U.K., in 2009. He worked as a Postdoctoral Research Associate at the Department of Automatic Control and Systems Engineering, University of Sheffield, from December 2013 to January 2016. Since January 2016, he has been working as an Assistant Professor with the University of Engineering and Technology Lahore. His main areas of research are computer vision, machine learning, pattern recognition, and multi-target tracking.

SAMEEMA TARIQ received the B.S. degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, in 2018. She is currently pursuing the M.Sc. degree from the Electrical Engineering Department, University of Engineering and Technology, Lahore.

HAROON FAROOQ received the Ph.D. degree in electrical engineering from Glasgow Caledonian University, U.K., in 2012. He is currently an Assistant Professor with the Electrical Engineering Department (RCET, Gujranwala), University of Engineering and Technology, Lahore, Pakistan. His research interests include power quality, renewable energy systems, electric vehicles, and demand side management.

ABDUL JALEEL received the B.S. degree in computer science and engineering, the M.S. degree in computer science, and the Ph.D. degree in computer science from UET, in 2006, 2010, and 2019, respectively. He is working as an Assistant Professor with the Rachna College, UET. His research interests include the development of self-managing software applications, machine learning, and autonomic computing.

SYED MUHAMMAD WASIF received the B.Sc. degree in electrical engineering from Riphah University, Pakistan, in 2006, and the M.Sc. degree in mechatronics and the Ph.D. degree in robotics from the King's College London, U.K., in 2008 and 2016, respectively. He is currently working as an Assistant Professor with the Department of Electrical Engineering, University of Gujrat, Pakistan. He is currently the Group Leader of the Intelligent Systems Lab, Department of Electrical Engineering, University of Gujrat. His research interests are mainly in the areas of robotics, intelligent control, active vision, machine vision, and machine learning.