Anomaly event detection and localization of video clips using global and local outliers

Sawsen Abdulhadi Mahmood[1], Azal Monshed Abid[2], Sadeq H. Lafta[3]
[1,2]Department of Computer Science, Mustansiriyah University, Baghdad, Iraq
[3]Department of Applied Sciece, University of Technology, Baghdad, Iraq

Article Info

ABSTRACT

The automatic detection of anomaly events in video sequence has become a critical issue and essential demand for the extensive deployment of computer vision systems such as video surveillance applications. An anomaly event in video can be denoted as outlier behavior within video frames which formulated by a deviation from the stable scene. In this paper, an anomaly event detection and localization method in video sequence is presented including multilevel strategy as temporal frames differences estimation, modelling of normal and abnormal behavior using regression model and finally density–based clustering to detect the outliers (abnormal event) at clips level. Hence, outlier score is obtained at the segment or clip level along video frames sequences. The proposed method seplits video frames into non-overlapped clips using global outlier detection process. Afterward, at each clip, the local outliers are determined based on density of each clip. Extensive experiments were conducted upon two public video datasets which include dense and scattered outliers along video sequence. The experiments were performed on two common public datasets (Avenue) and University of California, San Diego (UCSD). The experimental results exhibited that the proposed method detects well outlier frames at clip level with lower computational complexity comparing to the state-of-the-art methods.

Corresponding Author:

Sawsen Abdulhadi Mahmood
Department of Computer Science
Mustansiriyah University
Baghdad, Iraq
Email: sawsenhadi@uomustansiriyah.edu.iq

1. INTRODUCTION

Outliers, also known as anomalies, abnormalities or rare events, are data samples or objects in raw data that do not adapt to a concept of normal behaviour [1]. Owing to the growing requirement and applications in wide domains, such as video surveillance, security, healthcare and medical risk [2] as well as intrusion detection [3], anomaly and outlier detection task plays significant role in variant fields including computer vision, statistical analysis and machine learning. Outlier detection is a task of determining a sub-region in a given data samples which are considered as abnormal subset and have unusual behaviour from other samples. It is one of the essential data mining objectives and basic field in many applications, such as video surveillance systems which are used to recognize the potential threats [2]-[4].

Outliers or anomaly events can deviate or distort the statistical measurements and data distribution due to its rarely appearance in the real life events, allowing a confused representation of the implicit data and relationships [5]. Practically, the variant distribution of normal and anomaly events within video frames is unannounced with rare presentation of outliers. In surveillance videos, the predominant events occurring

frequently are denoted as normal behaviour, while abnormal events are referred to the events occurred with low probability [6]. The aim of outlier detection and localization is to determine the spatial and temporal sub-regions involved the anomalous events with automatic manner [7]. Outlier detection methods can be classified in terms of the data type required to train the model such as; point, contextual and collective anomalies. Point outlier is referred to a singular data instance deviated from the whole data samples. A contextual outlier represents an individual data sample deviated within a context which requires a notion of context as well as conditional anomalies. Collective anomalies are denoted by a set of correlated data instances and needs a relationship among data instances represented by sequential and spatial data. The singular patterns within a collective anomaly are considered not anomalous by themselves. Based on the anomaly definition provided in [8], "Video anomalies can be thought of as the occurrence of unusual appearance or motion patterns or the occurrence of usual appearance or motion attributes in unusual locations or times" we formulate the problem of anomaly event detection as regression model fitting as well as clustering based-local outlier detection task.

In this section, a discussion of recent researches relevant outlier and anomaly event detection in video sequence will be presented. Recently, video anomaly detections based on deep learning methods are introduced including different scenarios such as; saliency video detection method based on spatial-temporal features and 3D convolution network [9], temporal segment network [10], autoencoder network-based motion pattern learning [11], end to end anomaly detection of video based on pre-trained deep networks [12], deep anomaly detection with deviation networks [13]. However, deep learning methods are required a huge dataset sample to obtain an accurate performance measure for detecting anomaly events within video frames. Other researchers formulated the problem of video anomaly detection as a regression model to provide an anomaly score by frame or clip level [14], [15]. Simultaneously, other researchers have interested by trajectory-based video anomaly detection methods [16], [17].

Trajectory based methods include learning the normal trajectories of pedestrian with assistance of tracking algorithms to decide whether the video normal or anomaly. The main obstacle of trajectory-based methods is the influence of occlusion especially in crowded scenes as well as the likelihood of existing different trajectories for single scene [18]. Local and global anomaly detection method using hierarchical feature representation and Gaussian process regression was introduced in [19], where global anomalies refer to the anomalies among sequence frames and local anomalies denote the anomalous regions within a frame. Local anomalies are commonly detected in videos based on local spatio- temporal features, where motion is occurring and generating due to multiple objects moving within single scene [20]. Video anomaly detection methods are interested in determining whether the current frame of a given video demonstrates an anomaly or not [20], [21].

Another type of anomalies named contextual anomaly which corresponds to the samples having significant variation causing anomalies related to certain context [20]. Contextual anomalies can be categorized based on spatial and temporal features of video frames [22], [23]. Outliers detection methods based on clustering algorithms are inspired by many early researchers to efficiently detect outliers based on the entire dataset [24]. The proposed work in [25] handled the anomaly detection problem as low likelihood pattern detection due to limited availability of adequate instances of anomaly events with assistance of neural network for normal events learning. Lu *et al.* [26], suggested a dictionary-based approach to learn normal behaviors and detect anomaly events using reconstructed error. However, this method is not robust adequately to distinguish between normal and abnormal events based on the reconstruction error. The proposed work in [27] presented a learning model of normal scenes using two-stream recurrent autoencoder in a semi supervised learning manner and trajectory-based spatio-temporal features. Another work [28] based on autoencoder network and spatial features for learning normal events. The main limitation of this method represented by its dependency on succeeding the trajectory-based features extraction phase which is affected in crowded scenes.

The main objective of this research is to design and implement a computer vision system to address the problem of anomaly event detection in video. The Anomaly events (outiers) are detected in video sequence based on clip level. We can summarize the essential contributions of the proposed method by;

− Determining automatically the global and local outliers in video based on temporal redundancy of frames and dense clips clustering.
− Further, training a regression model can estimate the global anomaly score for the entire video frames, hence detects the anomalous potential frames within each clip in term of local outlier detection.
− In a continuous manner, the trained regression model is updated with normal segments which detected in the testing mode in order to take advantage of normal behaviour variation.

The rest of this paper is organized as follows: sections 2 and 3 present the detail description of anomaly event detection and localization method of video clips using global and local outliers. The results and discussion are illustrated in section 4. Section 5 concludes this paper.

2.   THE PROPOSED METHOD

In this section, we will demonstrate the proposed method of anomaly event detection and localization along video frames based on global and local outlier prediction as shown in Figure 1. In this paper, a multilevel framework is adopted to determine the anomaly behaviour (outliers) in a given video sample. First level is represented by estimation the temporal differences between sequenced frames based on visual features such as motion and histogram similarity features.

Second level mimics the detection of global outliers using temporal difference score and polynomial fitting model. To this end, an assumption of the regression model is presented and trained to specify the polynomial fitting with the temporal difference raw data of the entire video sequence. Consequently, the samples that deviate from the regression model are candidate to be global outliers. Accordingly, the suggested global outlier detection model is trained using only normal videos with normal behaviours to predict the optimal polynomial coefficients. We assume that an instant segment has a normal /abnormal behaviour based on the prediction score of the adopted regression model instead of using predefined threshold. In this level, the video frames are segmented into set of primitive clips or shots isolated by the corresponding indexes of temporal frames differences. The global outliers are exploited to segment the video frames into clips.

Finally, the local outliers' detection level is implemented where the contextual outliers within each video clip was taken in our consideration. Practically, the framework of an efficient anomaly event detector should be eligible to update its determination of normal events in a continuous manner. Thus, the recent approach is able to update the regression model through feeding the detected normal segments to the training samples from any acquired video. Figure 1 illustrates the workflow of the proposed anomaly event detection method.
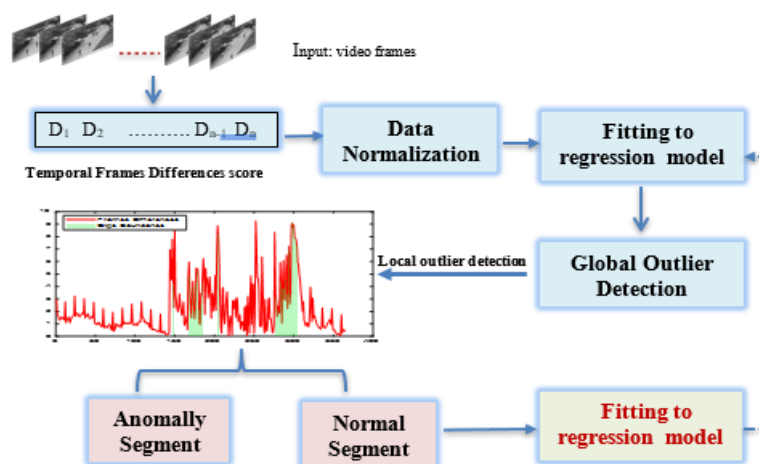

Figure 1. Workflow of the proposed anomaly event detection method

2.1.   Temporal frames differences estimation

As a pre-processing of video analysis, we segment the video into individual frames and convert each frame to YCbCr format in order to extract the luminance layer. In this paper, an efficient understanding of the video clip behaviours has been provided by formulating the temporal relevance among sequenced frames. In order to detect and extract the foreground objects (moving objects) over the video sequence, the temporal differencing between each two consecutive frames is estimated. The temporal frames differences are obtained based on two main integrated parameters; motion estimation and colour histogram difference. The motion parameter between frames F_i, F_{i-1} is obtained through binarization the subtracted image (F_i - F_{i-1}) with a *Threshold* value that is specified according to the intensity range of the subtracted image (F_i - F_{i-1}) and Otsu's method [29] as illustrated in (1):

$$MO_i = |F_i - F_{i-1}| > Threshold \tag{1}$$

Meanwhile, the foreground objects are extracted and labelled using (8–connected) neighbouring pixels of MO_i image which represents the foreground objects in the current frame F_i, where $i=2,3,....N$, and N is the total number of frames in the acquired video sample. Afterward, the area property AR_i is extracted from MO_i data points to highlight the interested regions, minimizing the computation time as well as to make a

decision about occurring motion between the sequence frames. We assumed that the higher values of AR_i measured in pixels (white pixels) referred to motion capturing between the sequence frames.

The second parameter used in our framework for temporal frames differences estimation is represented by the similarity benchmark commonly used in machine learning and data analysis. The Hausdorff distance (HD) distance measurement [30] is a lightweight yet effective similarity measure for quantifying the proximity of two graphs. The Hausdorff (HD) distance measure quantifies the difference between two sequenced frames based on block-based colour histogram differences. The temporal difference between each sequence frames F_i, F_{i-1} using Hausdorff (HD) distance measure is computed according to (2):

$$Diff_i = \sum_k HD(F_i^k, F_{i-1}^k)$$

(2)

where $Diff_i$ parameter denotes the weighted sum of blocks differences between sequence frames F_i and F_{i-1} and k is the corresponding block index in the two frames F_i, F_{i-1}.

A linear integration of the two parameters (AR_i and $Diff_i$) is specified to formulate the score of temporal differences between each two sequenced frames as given in (3):

$$TempDiff_i = \alpha AR_i + (1 - \alpha) Diff_i$$

(3)

where α is a scalar stand at 0.6 and it is determined through the experiments. Obviously, area feature AR is augmented by scalar α due to its effectiveness of reflecting the dissimilarity between two sequence frames. The temporal differences estimation strategy and its visualization result of sample video are illustrated in Figure 2 and Figure 3 respectively.
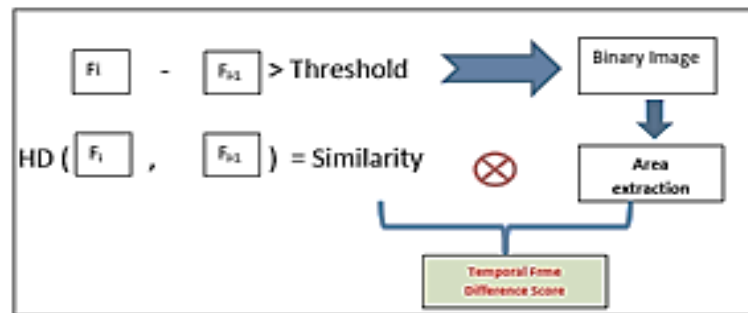

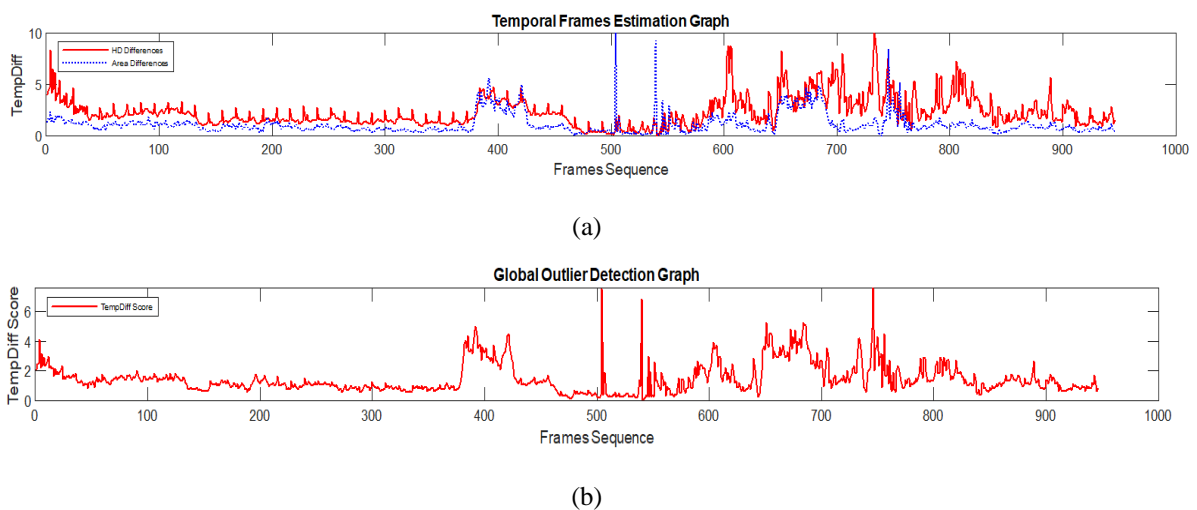Figure 2. Block diagram of temporal frame differences estimation strategy

(a)

(b)

Figure 3. Visualization result of temporal frame differences estimation based on; (a) area (blue line) and HD distance measure (red line); (b) TempDiff score of video sample (04) from AVENU dataset

3. MODELLING NORMAL BEHAVIOUR AND ANOMALY SCORE
3.1. Global outlier detection model

Based on the assumption; a data sample is considered as global outlier if it deviates sufficiently from the whole data set, the first level of detecting outliers is realized based on extracting the global outlier using the regression model. Regression analysis is a fundamental and common concept adopted in machine learning models. Specifically, it falls under supervised learning in which the training model is trained using input properties and output labels. It assists in investigating and establishing the relevance of the model variables through guessing how one variable influences the other.

In the case of occurring abrupt illumination change or sudden object motion along video sequence, we need to label the corresponding video frames with appropriate outlier score. In this context, a non-linear regression model is employed to achieve a polynomial fitting process with the raw data of temporal frames differences TempDiff. Consequently, the normal behaviour and the global outliers of the entire video frames are estimated and detected respectively. As a consequence, global outliers are detected by leveraging from the residual errors which sufficiently described data deviation from the fitted polynomial curve. Hence, predicting the set of potential values of polynomial coefficients will inform us of how appropriately our current model is able to describe the raw data behaviour. The main idea of adopting a regression model for global outlier detection in the whole video frames is based on the assumption that the temporal frames differences can be represented by non-linear polynomial model. The outliers are those temporal difference instants that are not well represented by the regression model. In this level, the sudden or abrupt changes along video frames are detected which corresponded to the differences samples that far sufficiently from the interval $[p+2\Delta, p-2\Delta]$, where p represents the polynomial coefficients and Δ refers to the residual error. In this case, the frames lying out of this range are labelled as global outlier frames as shown in Figure 4.

Henceforth, the proposed framework of global outlier detection task is designed as a regression model fitting, which considers a certain segment of the video has abnormal behaviour based on regression prediction score instead of adopting a predefined threshold for detecting anomaly. When there are anomaly changes, the anomaly score ascents significantly. It worth noting that the detected global outliers are exploited to automatically segment video frames into primitive shots or clips, yet to find and localize the local outliers within each clip. Neighbour frames of the detected global outlier frames are extracted and formulated into distinguishing regions and yet added to the memory buffer to collect the outlier regions along video sample. Figure 5 shows the detected global outliers constricted by the primitive shots boundaries of video samples 04, 06 taken from AVENU dataset.
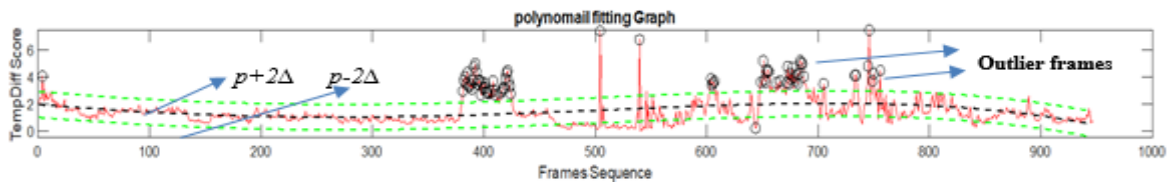

Figure 4. Global outlier detection based on polynomial fitting strategy of video sample (04) from AVENU
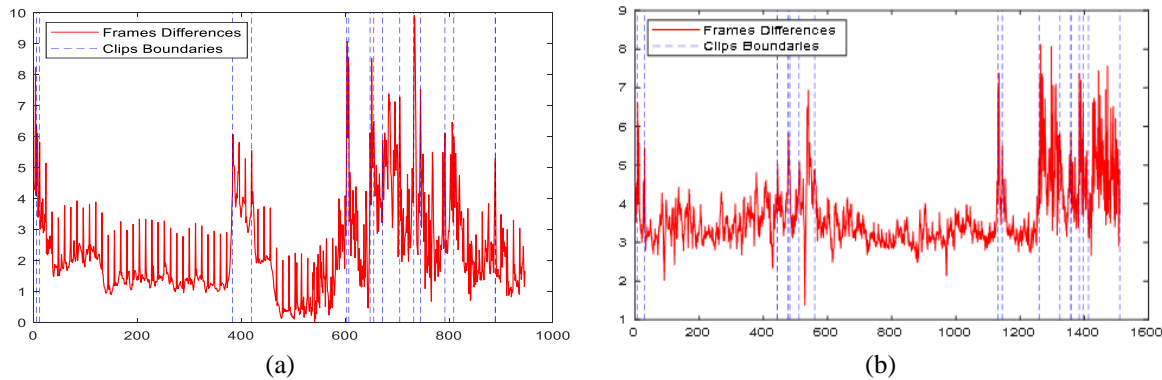

(a)   (b)

Figure 5. Detection of clips boundaries of: (a) video sample 04, (b) video sample 06 from AVENU dataset, X-axis represents frames indexes; Y- axis denotes the temporal differences scores

3.2. Local outliers detection

Local outliers can be defined as the patterns that prominent from their neighbourhood in a specific data region [24]. In this section, we will explain how to detect the local outliers in the acquired video sample based on clips density. The video frames are segmented automatically into non-overlapped K clips or shots based on the global outlier detection process illustrated in the previous section. We cope with each segmented clip as cluster with variable size S and assumed the average score of temporal frame difference of each cluster is the centre Co_m of cluster C_m. Afterward, the density of each cluster is determined based on the mean distance Md benchmark between each sample x_j belongs to m^{th} cluster and its centroid Co_m according to (4):

$$Md(C_m) = \frac{1}{S_m}\sum_j \|x_j - Co_m\| \tag{4}$$

where $m=1, .K$, and K is the number of the detected clips or clusters. Based on the mean distance Md mentioned in (4), a lower value of Md refers to present a dense cluster while a higher value of Md indicates that the cluster is scattered or sparse. Since the Md value is inversely related with the cluster density, the density DS of the cluster C_m is identified to be the inverse of Md (C_m) as illustrated in (5):

$$DS(C_m) = \frac{1}{Md(C_m)} \tag{5}$$

According to cluster density definition, the cluster with higher density leads to be its data elements are closer to its centroid. In contrast, low-density clusters tend to have data elements on average far from cluster centre. Regards to local outlier's detection task, we have selected the higher density clusters which included dense outliers and represented a candidate anomaly event with respect to the whole video sequence as shown in Figure 6. The corresponding frames of this type of clusters are labelled as anomaly event frames. The selection benchmark is based mainly on a predefined threshold to take a decision about lower and higher density clusters taking in our consideration its mean value. Further, the centroid Co of lower density cluster; i.e ($DS\approx 0$) is considered to be singular outlier within cluster or clip and the corresponding frame is labelled as anomaly frame as illustrated in Figure 6.
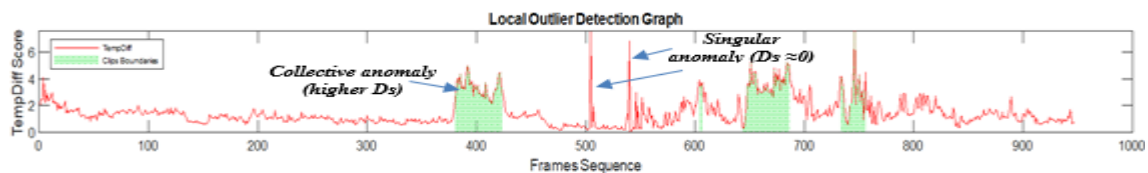

Figure 6. Demonstration results of singular and collective anomalies frames captured from video sample (04) from AVENU

4. RESULTS AND DISCUSSION

In this section, detailed description of common event detection datasets adopted in our experiments is presented as well performance evaluation metrics of anomaly detection score are computed and simulated graphically. Two main public dataset are employed in our experiments to evaluate the performance of the proposed anomaly detection task in video sequence. In the training mode, video samples with normal behaviour are deployed in our experiments in order to estimate the regression model coefficients. The testing mode is standing to validate the occurring of outlier's scores in the acquired video sample. Figure 7 presents some frames examples of normal and anomaly behaviour taken from UCSD [31] and avenu [26] datasets respectively. Avenue dataset [26] consists of 16 training videos and 21 testing video. Each video sample takes a short interval, around (1-2) minutes with frame resolution 640×360 pixels.

4.1. UCSD datasets (Ped1 and Ped2)

The UCSD dataset [31] composes of video samples of pedestrian walkways. The crowd density in the walkways was variable, ranging from sparse to very crowded. In the normal setting, the video contains only pedestrians. Abnormal events are due to either: the circulation of non-pedestrian entities in the walkways and anomalous pedestrian motion patterns. Peds1 dataset contains 34 training video samples and 36 testing video samples with a resolution of 238×158 pixels. Peds2 dataset contains 16 training video samples and 12 testing video samples with a resolution of 360×240 pixels. All testing samples are associated with a manually-collected frame-level ground truth relevant to abnormal events annotation. Figure 7 shows frames examples of normal and anomaly events taken from Avenu and UCSD datasets.
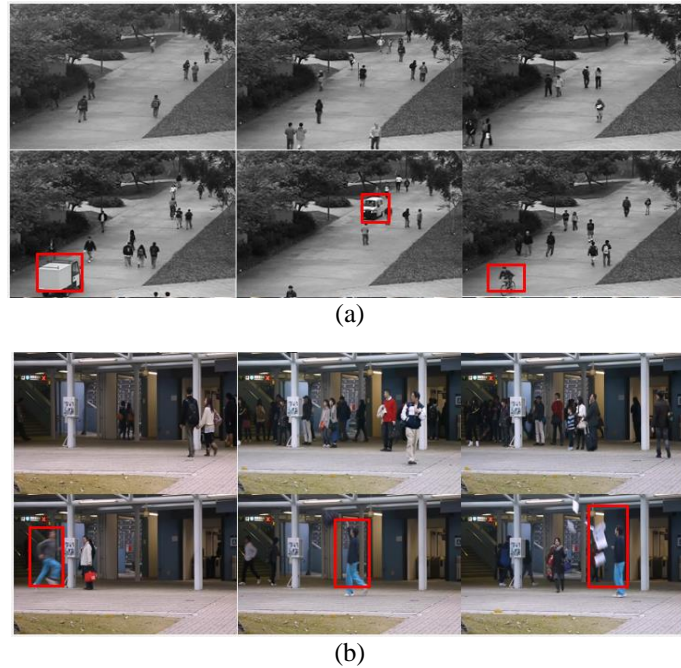
(a)

(b)

Figure 7. Examples of normal and anomaly frames in; (a) UCSD -Ped2 dataset, (b) Avenue dataset, the anomaly regions are demonstrated by red rectangle

4.2. Implementation details

The video frames are extracted and resized into resolution (320×200) pixels, then convert each frame into YCbCr color model to exploit the luminance component Y in the implementation framework. In order to estimate the anomaly score for current frame F_i, we adopt a Hausdorff distance measure as well as motion features between the two sequence frames (F_i, F_{i-1}). Then, the anomaly score is estimated based on integrating these two features by using linear (3). The estimated anomaly scores values are normalized into [0, 10] interval to facility the computation process. The normalization technique adopted in this work has been performed according to the following equation: $x_n = (x - x_m)/std$, where x_n represents the normalized data point, x_m and std are the mean and standard deviation of all data points respectively. The validation workflow of the proposed anomaly event detection in video is composed of two main phases; training and testing models as illustrated in the following subsections.

4.3. Training model

We pursue to determine anomalies in a given video sequence with lower computation time as well as specifying the required false alarm rate. Significantly, the proposed anomaly detection approach assumes that any anomalous pattern would hold for an obscure period of time. Meanwhile, we have no pre-knowledge about the anomaly event in the video sequence. Thus, the training mode is desired to address and describe the normal behaviour of the frame sequences. The training mode is configured by selecting the videos samples $V = \{vi: i=1, 2, ..., M\}$ with normal behaviour to perform the fitting process with a predefined regression model with 4th polynomial equation having four coefficients (b0, b1, b2, b3) and one temporal difference score (independent variable). The temporal frames differences *TempDiff* for each video sample have been computed. Consequently, we applied the polynomial fitting model upon the differences values to extract the suitable coefficients and residual errors of the regression model. In this context, the global outliers denote temporal differences instants that are not well represented by the estimated regression model.

4.4. Testing model

Any new raw data of different values obtained from the temporal differential phase are fitted with the trained regression model to detect the outlier in the raw data. The polynomial regression model fits a curve line to the temporal differences data as illustrated in Figure 8. Global outliers are detected based on the assumption that outliers are far away from the fitted model and have a higher residual error than the other data samples. The corresponding frames of this type of anomalies are labelled as anomaly frame. Further, the local outliers of the video sample have been detected at clip level based on density clustering strategy. The corresponding frames of this type of clusters are labelled as anomaly frame event. The selection benchmark

is based mainly on a predefined threshold to take a decision about lower and higher density clusters taking in our consideration its mean as the centre of each cluster.
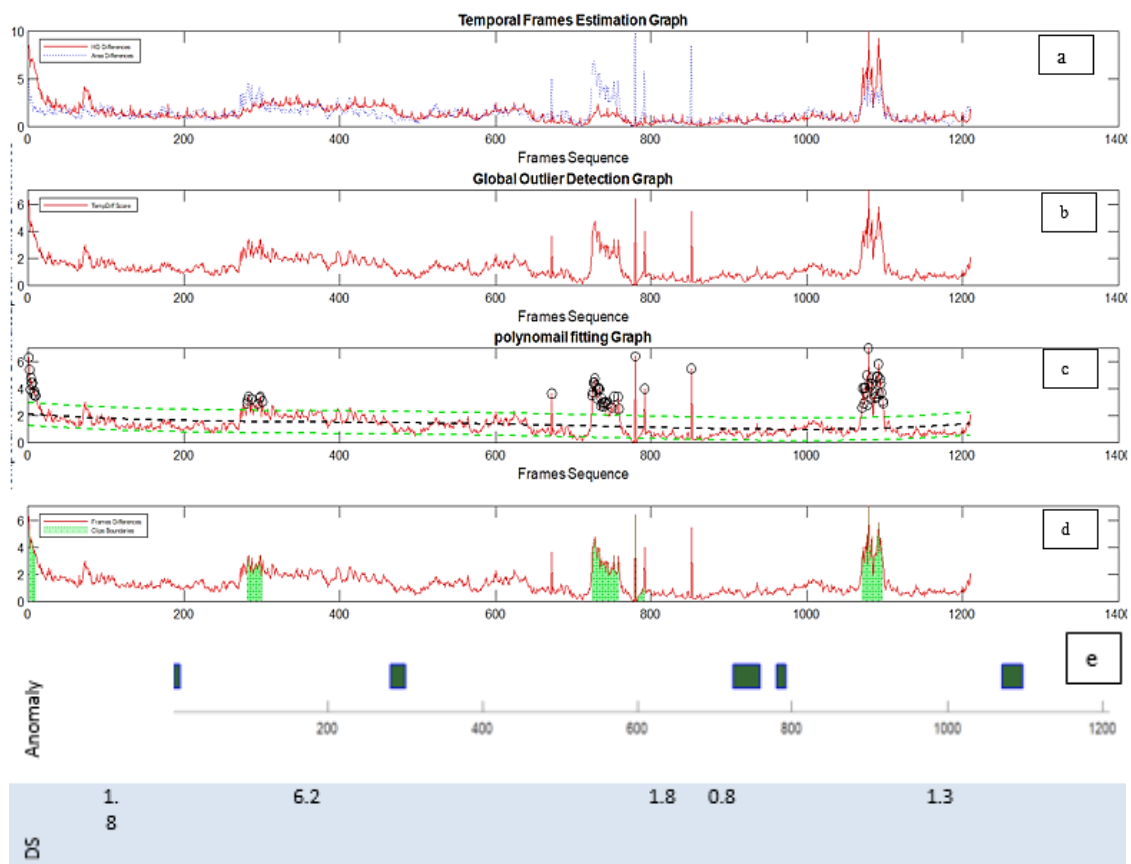

Figure 8. Visualization results of the proposed anomaly event detection method. (a) raw represents temporal frames differences, (b) raw represents the global outlier detection, (c) raw demonstrates the fitting process with regression model, (d) raw described the local outlier detection, (e) raw represents the extracted anomaly regions or clusters, last raw states the density of each cluster DS

4.5. Performance metrics

In our experiments, the quantitative results-based performance evaluation has been conducted using a receiver operating characteristic (ROC) curve and corresponding area under the curve (AUC) metrics. In this work, the anomaly event is detected based on a clip (event) level and frame level. Consequently, we record the begin and end of each anomaly event with assistance of ground truth results provided for each testing video in order to construct a suitable ground truth-clip level results for validation and comparison purposes. To obtain ROC curve demonstration for the tested video samples, true positive rate (TPR) and false positive rate (FPR) metrics are calculated at clip level. AUC area measures are computed for each video in both datasets, then we average the obtained results to show the AUC for each dataset compared with recent methods that based on neural networks such as [25] and [27]. Table 1 exhibits the correct detection score and false alarm benchmark for Avenu, Ped1 and Ped2 datasets respectively. Figure 9 illustrates the ROC curves quantitative results for Avenu, Ped2 and Ped1 datasets respectively.

Table 1. A comparison results of event level detection in term of correct detection and false alarm rate

Method	Correct detection/ false alarm		
	Avenu	Ped1	Ped2
Anomaly events	47	40	12
M. Hasan *et al.* [25]	45/4	38/6	12/1
S. Yan *et al.* [27]	34/6	38/5	12/0
Y. S. Chong and Tay [28]	43/8	36/11	12/3
The proposed work	45/3	38/4	12/1

A commonly, the higher values of the AUC area value referred to higher performance of the detector or classifier method. As shown in Figure 8, the proposed method presents a higher AUC value when applied on avenue dataset than Ped1 and Ped2. The crowded scenes included in video samples are effects mainly of increasing false alarm rates as shown in Figure 10. From other hand, we get competitive results and higher performance for avenu and Ped2 datasets when compared with the proposed approach to state of the art methods as illustrated in Table 2.
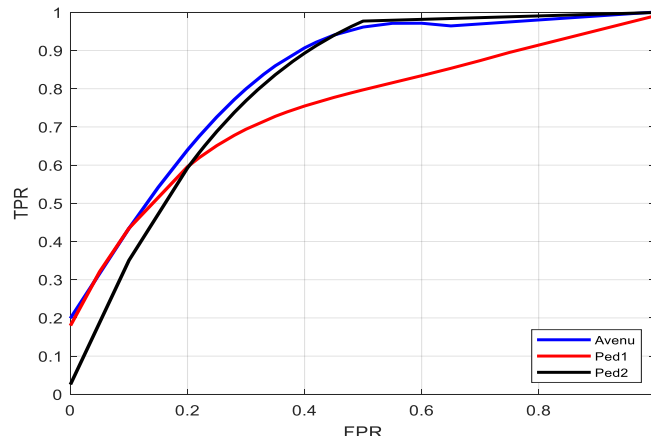

Figure 9. Roc curves visualization results of the proposed work at clip level for avenue, Ped2 and Ped1 datasets
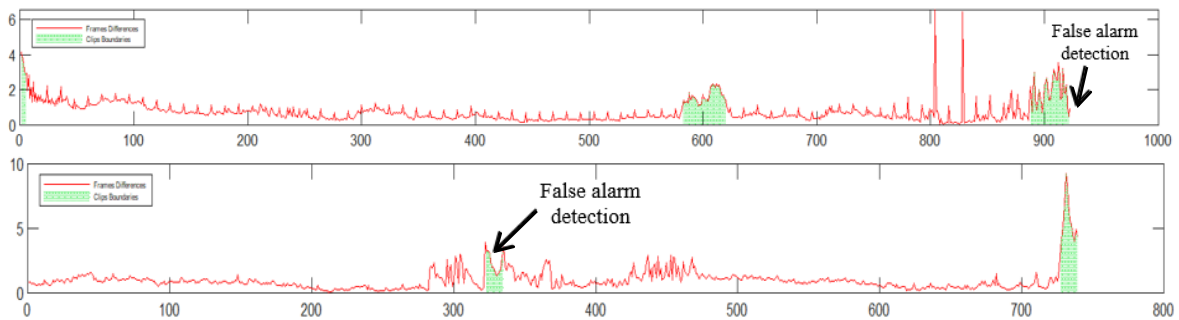

Figure 10. Two examples of failure cases of the proposed anomaly detector

Table 2. A comparison results of frame level detection in term of AUC metric

Method	AUC (%)		
	Avenu	Ped1	Ped2
G. Pang *et al.* [12]	-	83.2	71.7
M. Hasan *et al.* [25]	75.2	90	81.5
Lu *et al.* [26]	65.5	-	63.8
S. Yan *et al.* [27]	79.6	71.9	75.0
The proposed work	83	76	82

From the experiments, we found that the cluster with density close to zero value is considered as a singular anomaly frame (sudden change), while the clusters with higher density represent an anomaly collective frame in term of anomaly event detection in the acquired video sample. Further, the area parameter used for foreground objects extraction is more accurate and reflects an efficient criterion of the sequence frames behaviour compared to the similarity distance measure between each two sequenced frames. Furthermore, the proposed anomaly event detection scheme is validated based on the available dataset without needing to perform an augmentation of the datasets in contrast to deep learning-anomaly event detection-based methods which required a large dataset to learn the anomaly behaviours.

4 CONCLUSION

We have suggested and presented an anomaly event detection and localization scheme based on determining the global and local outliers in video sequence. The problem of global outlier detection was assumed as regression model fitting in order to segment the video frames into non overlapping clips. The proposed model is able to update the trained regression model using newly detected anomaly clips. The global outliers are exploited to segment the video frames into clips. Furthermore, the local outliers at each clip are detected using density of each clip taking in our consideration the contextual outliers within each video clip. Based on the experiments that implemented on public anomaly event detection dataset, we exhibited that the proposed anomaly detector significantly has a competitive performance compared to the state-of-the-art methods in terms of detection accuracy and false alarm rates at clip level.

ACKNOWLEDGEMENTS

REFERENCES

[1] G. Pang, C. Shen, L. Cao, and A. Van Den Hengel, "Deep learning for anomaly detection: a review," *ACM Computing Surveys,* vol. 54, no. 2, pp. 1-38, Mar. 2021, doi: 10.1145/3439950.
[2] K. Yan, X. You, X. Ji, G. Yin, and F. Yang, "A hybrid outlier detection method for health care big data," *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom),* Oct. 2016, pp. 157-162, doi: 10.1109/BDCloud-SocialCom-SustainCom.2016.34.
[3] Y. Wang, Z. Wu, Y. Zhu, and P. Zhang, "Research on anomaly detection algorithm based on generalization latency of telecommunication network," *Future Generation Computer Systems*, vol. 85, pp. 9-18, Aug. 2018, doi: 10.1016/j.future.2018.02.022.
[4] W. Li, V. Mahadevan, and N. Vasconcelos, "Anomaly Detection and Localization in Crowded Scenes," in *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 36, no. 1, pp. 18-32, Jan. 2014, doi: 10.1109/TPAMI.2013.111.
[5] Y. Cong, J. Yuan, and J. Liu, "Sparse reconstruction cost for abnormal event detection," *CVPR 2011*, 2011, pp. 3449-3456, doi: 10.1109/CVPR.2011.5995434.
[6] T. Xiao, C. Zhang, and H. Zha, "Learning to detect anomalies in surveillance video," in *IEEE Signal Processing Letters*, vol. 22, no. 9, pp. 1477-1481, Sept. 2015, doi: 10.1109/LSP.2015.2410031.
[7] L. Kratz and K. Nishino, "Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models," *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1446-1453, doi: 10.1109/CVPR.2009.5206771.
[8] V. Saligrama, J. Konrad, and P. Jodoin, "Video Anomaly Identification," in *IEEE Signal Processing Magazine*, vol. 27, no. 5, pp. 18-33, Sept. 2010, doi: 10.1109/MSP.2010.937393.
[9] K. Min and J. Corso, "TASED-Net: Temporally-aggregating spatial rncoder-decoder network for video saliency detection," *2019 IEEE/CVF International Conference on Computer Vision (ICCV),* 2019, pp. 2394-2403, doi: 10.1109/ICCV.2019.00248.
[10] L. Wang *et al.*, "Temporal segment networks: towards good practices for deep action recognition," *2016 European Conference on Computer Vision (ECCV),* vol. 9912, 2016, pp. 20-36, doi: 10.1007/978-3-319-46484-8_2.
[11] Y. Zhu and S. Newsam, "Motion-aware feature for improved video anomaly detection," *arXiv:1907.10211,* pp. 1-12, 2019.
[12] G. Pang, C. Yan, C. Shen, A. Van, Den Hengel, and X. Bai, "Self-trained deep ordinal regression for end-to-end video anomaly detection," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* 2020, pp. 12170-12179, doi: 10.1109/CVPR42600.2020.01219.
[13] G. Pang, C. Shen and A. Van Den Hengel, "Deep anomaly detection with deviation networks," *2020 Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 353-362, doi: 10.1145/3292500.3330871.
[14] F. Landi, C. G. Snoek, and R. Cucchiara, "Anomaly locality in video surveillance," *arXiv:1901.10364*, 2019.
[15] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition,* 2018, pp. 6479-6488, doi: 10.1109/CVPR.2018.00678.
[16] F. Jiang, J. Yuan, F. Jiang, J. Yuan, S. A. Tsaftaris, and A. K. Katsaggelos, "Anomalous video event detection using spatiotemporal context," *Computer Vision and Image Understanding*, vol. 115, pp. 323-333, 2011, doi: 10.1016/j.cviu.2010.10.008.
[17] Ce Li, Z. Han, Qixiang Ye, and J. Jiao, "Visual abnormal behavior detection based on trajectory sparse reconstruction analysis," *Neurocomputing*, vol. 119, pp. 94-100, 2013, doi: 10.1016/j.neucom.2012.03.040.
[18] M. G. Narasimhan and S. Kamath, "Dynamic video anomaly detection and localization using sparse denoising autoencoders," *Multimedia Tools Appl.,* vol. 77, no. 11, pp. 13173-13195, Jun. 2018, doi: 10.1007/s11042-017-4940-2.

[19] K. Cheng, Y. Chen, and W. Fang, "Video anomaly detection and localization using hierarchical feature representation and Gaussian process regression," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2015, pp. 2909-2917, doi: 10.1109/CVPR.2015.7298909.

[20] A. Mohan, M. Choksi, and M. A. Zaveri, "Anomaly and activity recognition using machine learning approach for video based surveillance," *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT),* 2019, pp. 1-6, doi: 10.1109/ICCCNT45670.2019.8944396.

[21] R. Nayak, U. Chandra Pati, and S. Kumar Das, "A Comprehensive review on deep learning-based methods for video anomaly detection," *Image and Vision Computing,* vol. 106, pp. 1-69, Feb. 2021, doi: 10.1016/j.imavis.2020.104078.

[22] K. Pawar, and V. Attar, "Deep learning approaches for video-based anomalous activity detection," *World Wide Web*, vol. 22, pp. 571-601, 2019, doi: 10.1007/s11280-018-0582-1.

[23] M. J. Leach, E. P. Sparks, and N. M. Robertson, "Contextual anomaly detection in crowded surveillance scenes," *Pattern Recognition Letters*, vol. 44, pp. 71-79, 2014, doi: 10.1016/j.patrec.2013.11.018.

[24] S. Su *et al.,* "An efficient density-based local outlier detection approach for scattered data," in *IEEE Access*, vol. 7, pp. 1006-1020, 2019, doi: 10.1109/ACCESS.2018.2886197.

[25] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury and L. S. Davis, "Learning temporal regularity in video sequences," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2016, pp. 733-742, doi: 10.1109/CVPR.2016.86.

[26] C. Lu, J. Shi, and J. Jia, "Abnormal event detection at 150 FPS in MATLAB," *2013 IEEE International Conference on Computer Vision*, 2013, pp. 2720-2727, doi: 10.1109/ICCV.2013.338.

[27] S. Yan, J. S. Smith, W. Lu and B. Zhang, "Abnormal event detection from videos using a two-stream recurrent variational autoencoder," in *IEEE Transactions on Cognitive and Developmental Systems*, vol. 12, no. 1, pp. 30-42, Mar. 2020, doi: 10.1109/TCDS.2018.2883368.

[28] Y. S. Chong and Y. H. Tay, "Abnormal event detection in videos using spatiotemporal autoencoder," *International Symposium on Neural Networks*, pp. 189-196, 2017, doi: 10.1007/978-3-319-59081-3_23.

[29] M. Huang, W. Yu, and D. Zhu, "An improved image segmentation algorithm based on the otsu method," *2012 13th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing,* 2012, pp. 135-139, doi: 10.1109/SNPD.2012.26.

[30] I. S. Kim and W. McLean, "Computing the hausdorff distance between two sets of parametric curves," *Communications of the Korean Mathematical Society*, vol. 28, no. 4, pp. 833-850, Oct. 2013, doi: 10.4134/CKMS.2013.28.4.833.

[31] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1975-1981, 2010, doi: 10.1109/CVPR.2010.5539872.