Abnormal Event Detection and tracking in Surveillance Videos using Neural Networks

LRTDM Bandara[1#], G. Gayamini [2], and Dr. Thushara Weerawardane [3]

1General Sir John Kotelawala Defence University1, Sri Lanka
2 General Sir John Kotelawala Defence University1 2, Sri Lanka
3 General Sir John Kotelawala Defence University13, Sri Lanka

DOI: 10.29322/IJSRP.12.09.2022.p12944
http://dx.doi.org/10.29322/IJSRP.12.09.2022.p12944

Paper Received Date: 13th August 2022
Paper Acceptance Date: 15th September 2022
Paper Publication Date: 26th September 2022

Abstract— *In recent years abnormal event detection has become an important research area because of the following reasons : the increase of the applications of it, the increase of the number surveillance cameras in public locations such as museums, markets, airports, stadiums, train stations, etc., and the advancements of the peoples' public safety awareness. AED is a subtopic in event detection, In the study of video content, event detection is critical. Abnormal event detection can be applied in 'smart surveillance monitoring' systems, which can recognize abnormal incidents in video footage automatically and decide what to do based on the type of abnormal event. The primary premise behind abnormal event detection is to use computer vision technology to extract the characteristics that represent unusual activity from a video and then train a model that can recognize abnormal events. Surveillance footage is typically viewed by human operators who must spot unusual activities in real time and react accordingly. But recently this traditional method has failed due to the increased number of surveillance cameras. When considering about the object or people tracking it is hard to track people using the faces because the low visibility if the video footage in surveillance cameras. Therefore, several other alternate methods are used. By combining the above two technologies a system to detect abnormal events then to identify and track the related persons through multiple videos from multiple cameras is to be developed.*

Keywords— **Artificial Intelligence, Neural Networks, Abnormal Event Detection (AED), Transfer Learning, Object identification**

I. INTRODUCTION

Abnormal event detection also known as AED, a subtopic of event detection, is a difficult study area that has drawn a large number of researchers in recent years. Human activity identification, scene identification, tracking people and objects, and human behavior identification are all examples of event detection. The video surveillance application is the major focus here. Furthermore, this can be used to monitor public locations including railway stations, shopping complexes, packed sports venues, airports, military sites, and so on, as well as smart healthcare institutions for motion and activity tracking of patients and for detecting accidents in elderly homes [1].

Recently the use of CCTV (Closed-circuit television) to record human activity has become well-known. However, there are insufficient human resources to analyze the data. It could also be more expensive, wasteful, and less accurate to use human labor. Human variables such as exhaustion and boredom might also have an impact on the ultimate result. The major objectives of AED are to identify abnormal occurrences in surveillance videos and to increase the accuracy and the efficiency of abnormal event recognition process [2]. AED techniques are primarily divided into two groups: Object-centric approach where the attention is on the abnormal occurrence and the object that causes it. And holistic approach, which is concerned with the direction and flow of objects and people [3].

There are several methods for detecting anomalous events, but the general processes are as follows: (i) raw data is pre-processed to remove noise, (ii) video segmentation, (iii) feature extraction, (iv) feature reduction for accuracy, and (v) event classification using specific algorithms. When it comes to detecting unusual events, there are number of problems and obstacles to overcome such as : Low quality footages can cause poor feature extraction, as a result of the incorrect interpretation of the features, the categorization of events was poor, some similarities in the detected motion pattern may cause separate activities to be identified as the same [4]. Many researchers presented many approaches to improve the detection process' efficiency, accuracy, and performance. Different methods and

algorithms for abnormal event detection videos utilizing computer vision are examined and reviewed in this paper. [5].

II. LITERATURE REVIEW

There have been a large number of studies on abnormal event detection. Some are focused on Spatio Temporal data [6], CNN (convolutional neural networks), SVM (Support Vector Machine) & LSTM (Long Short-Term Memory) [7], Graph Formulation [7], Oriented GMM (Gaussian mixture model) model [8], etc. have been proven to be successful.

Yilian Wang [6]. Normally, supervised methods are used to detect abnormalities; however, in this research paper, an unsupervised method is proposed, which has the advantage of being able to train and test on unlabeled data sets. Spatial Temporal Saliency is utilized for two key objectives in this approach. (i) Saliency Prediction and (ii) Abnormality Detection. FFT (Fast Fourier Transformation) and Signals are converted into frequency domain with amplitude and phase using IFFT (Inverse Fast Fourier Transformation). The original video signal is then compressed using a High Pass Filter (HPF) in the frequency domain in the temporal dimension, and the analysis is performed using Parseval's Theorem. The sum of all dimensions in the regularity domain equals the sum of all regularities, according to this theorem.

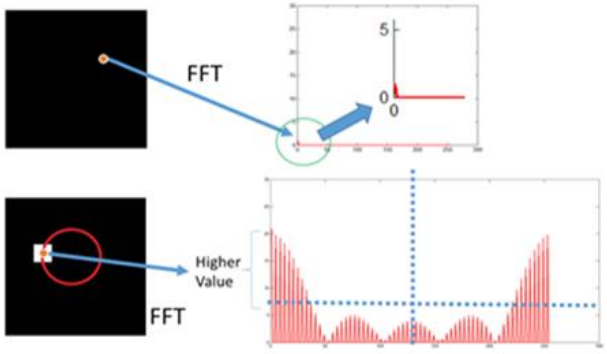
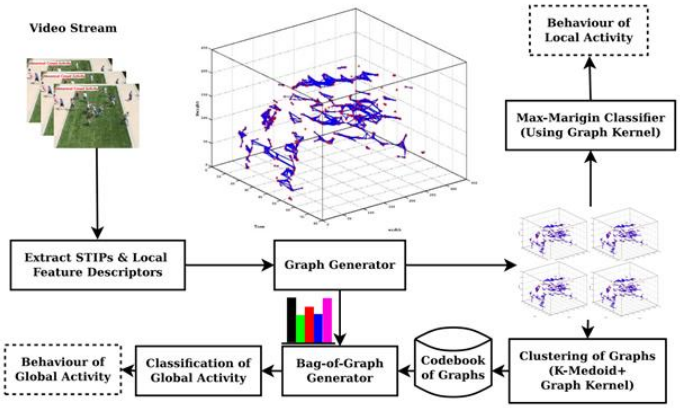


Figure 1. The spectrum analysis for normal videos. The first column is sample frame, the second column shows sampled video points in frequency domain.

Dinesh Singh, 2017 [7] , For abnormal video detection, a graph-based technique is used. A graph design technique for video activities was utilized to discover abnormal behavior in video surveillance using graph kernel SVM (Support Vector Machine). The spatial-temporal interest point is represented by the vertices in this graph and the link between the arrival and dynamic forces surrounding the interest point is represented by the edges. Fuzzy membership functions based on their closeness are utilized to find the graph's edge. Kernel graphs are also used to detect similarities. Finally, the graph is drawn from the video, and abnormalities are detected from the graph using SVM.

Figure 2. Block diagram of the proposed framework for abnormal event detection or recognition in surveillance videos



Feiping Li [8] This paper discusses a method called as Oriented GMM (Gaussian mixture model). Optical flow is utilized as a low-level characteristic in this technique. The GMM model is used at each stage and position during the training phase. In the testing stage, use the likelihood approach to determine probability and discover anomalous activity. It is a non-tracking method in which a likelihood map is built for each frame to discover abnormalities. These likelihood maps help in determining whether or not the behavior is normal.

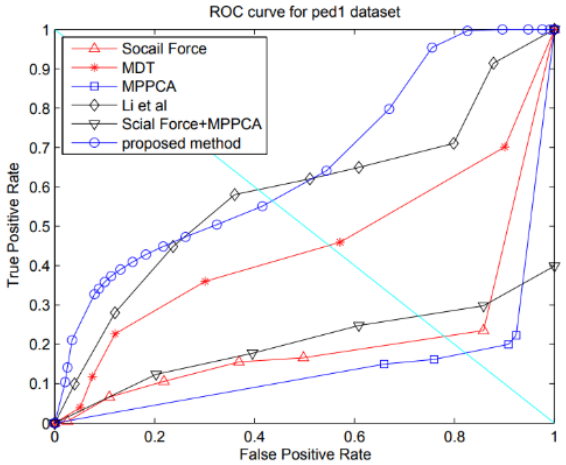


Figure 3. Curves of approach and comparison approach on Peds 1 dataset with pixel level ground truth

Zhaohui Luo [9] Sensitive movement point is a new feature (SMP) method is discussed in this paper. First, the video is analyzed with the help of GMM, and then the SMP (Support Vector Machine) is obtained. Afterwards, the GMM's threshold value is established. SMP is investigated using Spatio-temporal modeling. Anomalies in spatial temporal modeling are discovered during the study. Normal behaviors are then assigned to the normal state, while abnormal behaviors are assigned to the variable and unstable state.

Yuan Yuan [10], This paper discusses a statistical detector-based strategy for detecting anomalies. Each data sample is analyzed as they are a collection of patterns. Initially, abnormal events are detected when abnormality scores are high. Simple

distribution is used due to noise and complexity of the video data. A Gaussian mixture is used for complicated noise complexity.

Dan Xu [11], For the appearance and motion pattern, a double fusion framework is used. To differentiate the motion and appearance patterns, and also joint representation, de noising auto-encoders are used. Finally, SVM is utilized to classify each input and estimate its score. After that, the final fusion is applied to these scores in order to detect abnormal events.

Yuan Gao [12], Oriented Violent Flow (OViF) is a method for changing the magnitude of motion in statistical motion orientations in video data. The goal of OViF planning is to fully utilize the alignment data from the videos. Linear SVM classifier is used for multiple classifiers, which is done on feature specific by 'AdaBoost', therefore it improves the rate of violence activity identification.

J.F.P.Kooija [13], The purpose of this paper is to detect hostile behavior in a public location. CASSANDRA is the name of a smart surveillance system discussed here. Audio and video streams are merged into a composite unit here. DBN is used to predict the level of hostility in this method. Then, after removing any further views, the remaining volume EM technique locates the potential object.

Xinlu Zong [14]. The method is focused upon sparse representation. The basic idea behind this technique is to train this model using both normal video and abnormal video data, resulting in two sparse representation models, and then determining whether or not there are abnormal activitoes based on the outputs of the two models. The identification of abnormal occurrences in video is essentially a different problem, in which the model divides a video sequence into two categories: normal and abnormal. Now, abnormal event detection research is mostly focused on two aspects : extraction of different features and classification model design. The construction of a classification model is a model for classifying video frames (normal and abnormal). It is produced through the extraction of features prior to training. The detection model is classified into supervised, semi-supervised, and unsupervised models based on the varying needs of the training set. The semi-supervised model is the most widely used detection model among them. The Histogram Oriented Gradient (HOG) feature is employed in the feature extraction stage.

Feature extraction

Feature extraction is the process of obtaining features from a video stream that signify abnormal events. The model's detection performance is directly affected by the results of feature extraction process. The direction gradient histogram collects features from gradient direction information in the graphic picture's local area to produce a histogram and to communicate the image information. The HOG approach extracts the target contour information in the video with efficiency. Before extracting HOG features, the motion foreground area is split in this paper. For HOG feature extraction, the obtained binarization foreground region is used as the input image. Because most pixels in the binarization image have a pixel value of 0, therefore the time spent extracting HOG feature can be significantly decreased [14].

B. VGGNet (Visual Geometry Group Network)

VGGNet, developed by Karen Simonyan and Andrew Zisserman from Oxford University's Visual Geometry Group (VGG) research lab, came in second place in the ILSVRC 2014 challenge. It consist of a very simple and classic architecture, with two or three convolutional layers and a pooling layer, then two or three convolutional layers and a pooling layer, and so on (for a total of 16 or 19 convolutional layers, depending on the VGG type : VGG16 and VGG19), plus a final dense network with two hidden layers and the output layer [15].

C. Transfer learning

If there is a trained to recognize faces in photos and if you want to train a second neural network to recognize hairstyles, you can jumpstart the process by reusing the first network's lower layers. Rather than randomly implementing the weights and biases of the first few layers of the new neural network, you can set them to the values of the lower levels of the original network's weights and biases. The network will no longer have to learn all of the low-level structures that appear in most photos from scratch; instead, it will only have to learn the higher-level structures. This is referred to as transfer learning [15].

ISSN 2250-3153

Reference no	solution	Detection types
[6]	Spatio-Temporal Saliency Detection	Crowd activity detection
[7]	Using CNN, SVM & LSTM	Robust framework for recognizing motion of activities
[8]	Oriented GMM model	Anomaly Detection
[9]	GMM, Spatial Temporal modelling	Activity recognition and behavior analysis.
[10]	Statistical Hypothesis Detector	Crowd activity detection
[13]	Multi-model human aggression detection using DBN	Violence Detection
[14]	Using Compact Feature Sets & GMM, Markov model	Anomaly Detection
[16]	Using Generative Adversarial Network	Anomalous event detection
[12]	Violence detection using OViF, SVM & AdaBoost	Violence Detection

Table I. comparison of methods present in above discussed models [3]

III. PROPOSED SOLUTION

The proposed solution is a system that consist of two main parts.

1. an event detection module (identification)
2. an object re-identification module (person tracking)

The inputs of the first module are the surveillance videos which may or may not contain any abnormal event. Once the videos are fed into the first module of the system it identifies whether an abnormal or an unusual event occurs. If there is any event which is out of the ordinary, it will produce an alert.

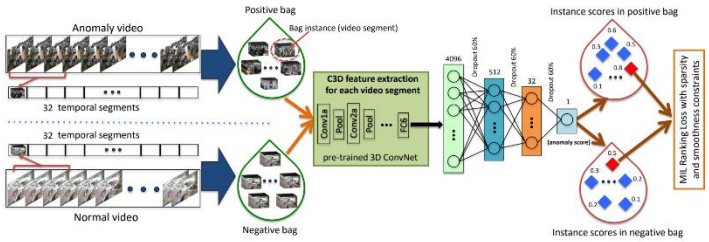


Figure 4. Shows the abnormal event identification process which occurs in the first module

The inputs of the second module are the surveillance videos which may contain the related individual / individuals. Then the second module can re-identify and track by selecting the involved individual / individuals. Here whether an abnormal event occurs or not does not affect the result of the second module. It tracks the identified individual / individuals across number of videos. Here face recognition is not the most suitable technology since the surveillance footages are not the best videos to identify a face due to the low resolution and the various viewing angles. Therefore, the appearance of the whole person is used to re-identify the individual / individuals. Therefore even if the subject / subjects cover their faces it does not affect the result.

IV. RESULTS AND DISCUSSION

When a surveillance video is uploaded to the first module it produces the following graph with a y-axis which goes from 0 to 1. Where 0 represents the normal frames and 1 represent the abnormal frames. Here the video used is from the dataset UCF-Crimes and the content is a fight which goes on between few people which starts as a normal incident and gradually turns into a fight. When the graph exeeds a curtain threshold the alert is produced.

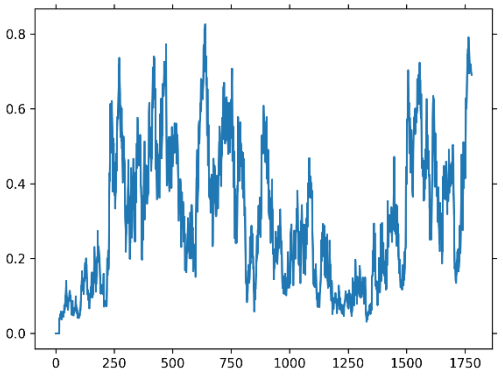


Figure 5. shows the graph output of the first module after processing a video containing an abnormal event (fighting)

The results from the second module is show below. Once the source image / images (source image : the cropped image from

the video from the first module which contains the person / people) is uploaded to the second module along with the surveillance videos it identifies the person / people and produce an output

Figure 6. this is the source image which is fed into the second module along with the videos

Figure 7. this is the images which the second module identifiied from videos using the source image

IV. DISCUSSION

Object recognition and event detection are two of the most necessary study topics among image processing experts. In the context of security, identifying abnormal events or behaviors in video surveillance becomes more important. Researchers are becoming increasingly interested in learning about the procedures involved and the algorithms used in this recognition process. In the subject of abnormal activity detection, there are numerous methodologies and strategies available. The most prominent and promising technologies and methodologies are covered in this review paper. In addition, advantages of different technologies and algorithms are compared and reviewed. Many techniques such as Spatio-temporal Saliency Detection, Graph Formulation, Oriented GMM, Based on Sparse reconstruction, and utilizing Statistical Hypothesis Detector have been used in related work. The following are some of the most common technologies used in almost all of the methods listed above:

IV. CONCLUSION

With the fast development of surveillance cameras in various fields to record human behavior, so does the demand for systems that automatically recognize abnormal situations. Abnormal event detection has become a popular topic in computer vision, attracting new researchers. Many scholars have proposed various strategies for detecting such events in video. The primary purpose of this review is to examine the most recent studies in the field of abnormal event detection. The review covers approaches for detecting abnormal events using SVM, CNN, and machine learning classification. These

IV. FUTURE WORK

Currently the 2 modules are running independently with a much more inteventoin from the user. As the further development, to combine the two modules and to make it more autonomous. To create an interface which is easy to use. Furthermore to increase the efficiency and accuracy of both the abnormal event detection and the person re-identification processes.

IV. REFERENCES

[1] Oluwatoyin Popoola and Kejun Wang, “Video-Based Abnormal Human Behavior Recognition—A Review,” presented at the IEEE Transactions on Systems Man and Cybernetics Part C (Applications and Reviews), 2012.

[2] Mona Izadi, Zohreh Azimifar, and Gholam-Hossein Jowkar, “Abnormal Event Detection in Indoor Video using Feature Coding,” presented at the 2017 Artificial Intelligence and Signal Processing (AISP), 2017.

[3] Alisha Ahir, Pushpendra Kumar Pateriya, Devinder Kaur, and Mritunjay Kumar Rai, “A Review on Abnormal Activity Detection Methods,” presented at the 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), 2018.

[4] N. Kumaran and U. Srinivasulu Reddy, “Object detection and tracking in crowd environment — A review,” presented at the 2017 International Conference on Inventive Computing and Informatics (ICICI), 2017.

[5] Muhammad Ramzan et al., “A Review on State-of-the-Art Violence Detection Techniques,” 2019, vol. 7.

[6] Wang Y, Zhang Q, Li B.Efficient unsupervised abnormal crowd activity detection supported a spatiotemporal strikingness detector.InApplications of Computer Vision (WACV), 2016 IEEE Winter Conference on 2016 Mar 7 (pp. 1-9). IEEE.

[7] Singh D, Mohan CK.Graph formulation of video activities for abnormal activity recognition.Pattern Recognition. 2017

[8] Li F, Yang W, Liao Q. An economical anomaly detection approach in police work video supported oriented GMM. InAcoustics, Language and Gesture process , twenty16 IEEE Intercontinental Session on 2016 (pp. 1981-1985). IEEE.

[9] Luo Z, He W, Liwang M, Huang L, Zhao Y, Geng J. Actual finding algorithmic program of irregular actions in gatherings supported mathematician mixture model. InComputer Science

and Education (ICCSE), 2017 12th International Conference on 2017 (pp. 183-187). IEEE.

[10] Yuan Y, Feng Y, Lu X. Geometric assumption indicator for anomalous incident recognition in jammed scenes. IEEE dealings on robotics. 2017

[11] Xu D, Yan Y, Ricci E, Sebe N. Identifying irregular actions in cartridges by knowledge profound symbols of look and signal. Processor Visualization and Appearance Considerate. 2017

[12] Gao Y, Liu H, Sun X, Wang C, Liu Y. Viciousness finding exhausting concerned with intense movements. Appearance and image figuring. 2016

[13] Kooij JF, Liem MC, Krijnders JD, Andringa TC, Gavrila DM. Multimodal social anger exposure. Processor Visualisation and Appearance Appreciative. 2016

[14] Xinlu Zong et al., "Abnormal Event Detection in Video Based on Sparse Representation," presented at the 2020 15th International Conference on Computer Science & Education (ICCSE), 2020.

[15] Aurélien Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition. O'Reilly Media, Inc., 2019.

[16] Ravanbakhsh M, Nabi M, Sangineto E, Marcenaro L, Regazzoni C, Sebe N. Abnormal event detection in videos mistreatment generative adversarial nets. InImage process (ICIP), 2017 IEEE International Conference on 2017 Sep seventeen (pp. 1577-1581). IEEE.

IV. ACKNOWLEDGEMENT

I would like to take this opportunity to humbly extend my sincere gratitude and thanks to a few individuals who inspired and motivated me throughout the course of this project, to venture into new developments in the field of computer science. My sincere gratitude also goes to Mrs. G. Gayamini who helped me this ideas and advices on event detection and object tracking and developing such a software. With the help of Dr. Thushara Weerawardane' s inspirations given to conduct the project. To all the past researchers who have conducted research in the fields explored in this study, for all the days of challenging work they put in to achieve their goals. Of special note is everyone who generously made available the research they conducted, online and otherwise, in relevant fields. I am indebted to them for all the material I used for this project. It was in fact those which helped me immensely whenever I got stuck in the course if research.