Real Time Video Surveillance for Automated Weapon Detection

Bhagyalakshmi. P[1], Indhumathi. P[1], Lakshmi. R[1], Dr. Bhavadharini[2]

[1]UG Student, [2]Associate Professor

[1,2]Computer Science and Engineering, Easwari Engineering College, Chennai, Tamil Nadu, India

ABSTRACT

Closed circuit television systems (CCTV) play a vital role in evidence collection against crimes and criminals. The existing systems does not classify normal and abnormal events leading the police to become more reluctant to attend the crime scenes unless there was a visual verification, either by manned patrols or by electronic images from the surveillance cameras. The Proposed work is being used for surveillance, monitoring and classifications of weapons, live tracking and many more purposes. In this work, live surveillance videos is taken for monitoring and detecting the abnormal events based on real time image processing techniques. Operations of proposed project has three processing modules, first processing module is for object detection using Convolutional Neural Networks(CNN) and second processing module will handle the classification of weapons, monitoring and alarm operations will be carried out by the third processing module. CCTV will monitor circular area and it will automatically perform all operations and be controlled. Shape detection algorithms and object detection algorithms have been tested to find accuracy in detection and analysis the processing time before implementing in such environment and results provide optimal accuracy in matching weapons and objects type with name and shape in predefined database like ALEXNET. The proposed work drastically reduces the crime rate and it also provide a higher level security in certain areas and it will reduce the time required to catch the criminal.

Keywords: Video surveillance, abnormal events, object detection, CNN, ALEXNET

I. INTRODUCTION

Closed circuit television systems (CCTV) are becoming more and more popular and are being deployed in many offices, housing estates and in most public spaces. There are a million of CCTV cameras that are currently in operation in India. This makes for an enormous load for the CCTV operators, as the number of camera views a single operator can monitor is limited by human factors. The task of the CCTV operator is to monitor and control, detect, observe, recognize and identify individuals and situations that are potentially harmful to other people and property but it becomes harder to monitor when there are a lot of CCTV cameras.

A solution to the problem of overloading the human operator is to apply automated image-understanding algorithms, which, rather than substituting the human operator, alert them if a potentially dangerous situation is at hand.

When an individual carries a weapon (firearm or a knife) out in the open, it is a strong indicator of a potentially dangerous situation. While some countries allow for open carry firearms, in such an event, it is still advisable to grab the CCTV operators' attention in order to assess the situation at hand.During recent years, an increase in the number of incidents with the use of dangerous tools in public spaces can be observed.

Fig 1.1 CCTV images of crime scenes

Automated methods for video surveillance have started to emerge in recent years, mainly for the purpose of intelligent transportation systems (ITS). They include traffic surveillance and recognition of cars.. In this study, we have focused on the specific task of automated detection and recognition of dangerous situations applicable in general for any CCTV system. The problem we are tackling is the automated detection of dangerous weapons—knives and firearms, the most frequently used and deadly weapons. The appearance of such objects held in a hand is an example of a sign of danger to which the human operator must be alerted.

We propose an initial approach to systems designed for knife and firearm detection in images, respectively. In this work, we summarize this effort and present the current versions of the algorithm. Even if different methods are also used, the algorithms presented in this paper aim towards a similar goal; our motivation is to solve the problem of knife or

firearm recognition in frames from camera video sequences. The aim of these approaches is to provide the capability of detecting dangerous situations in real life environments, e.g., if a person equipped with a knife or firearm starts to threaten other people. The algorithms are designed to alert the human operator when an individual carrying a dangerous object is visible in an image. We present the complex problem of fully-automated CCTV image analysis and situation recognition. We define the requirements for a fully-automated detection and recognition solution, and we propose a complex, multi-stage algorithm and evaluate its effectiveness and limitations in given conditions. Finally, we discuss the results and point to further development paths for our solution and similar techniques.

II. LITERATURE SURVEY

Qichang Hu et al [1] proposed a detection framework which involves three phases detection of objects of interest, recognition of detected objects and tracking of motion. Single learning based detection framework is used because of which high processing speed is achieved. Because dense features need only to be evaluated once rather than individually for each detector. And this framework also introduces spatially pooled features as a part of aggregated channel features to enhance the feature robustness. For object detection a framework using a linear support vector machine classifier with histogram of oriented gradients features.

For large inter-class variation, cannot be tackled by conventional VJ framework(Viola and Jones) instead it combines object sub categorization to cluster the object classes. So this becomes an disadvantage in using VJ framework. Using a combination of ACF(Aggregated Channel Features) features and sp-LBP(Local binary pattern)features can provide a better trade-off between detection performance and system runtime. The KITTI dataset provides a wide range of images from various traffic scenes with fully annotated objects. information. To improve detection performance, some techniques are used to post-process raw detection results. And they are Calibration of Confidence Scores, Non-Maximum Suppression (NMS) and Fusion of Detection Results.

The issues with above approaches are that they are not adaptive under severe weather and lighting conditions. And another challenging problem is that detection of car is difficult with large intra class variation at different viewpoints. Uses shrinkage version of AdaBoost as the strong classifier and use decision trees as weak learners. To train the classifier, the procedure known as bootstrapping is applied.

Shifu Zhou et al [2] suggested a method for detecting and locating anomalous activities in video sequences of crowded scenes. The key for method is the coupling of anomdescribon with a spatial-temporal Convolutional Neural Networks. This architecture allows us to capture features from both spatial and temporal dimensions by perform ing spatial-temporal convolutions, thereby, both the appearance and motion information encoded in continuous frames are extracted. The spatial-temporal convolutions are only performed within spatial temporal volumes of moving pixels to ensure robustness to local noise, and increase detection accuracy.

The existing approaches for detecting anomalies can be classified into two categories they are object-centric

approaches and holistic methods. The spatial-temporal CNN model is applied only on spatial-temporal volumes of interest (SVOI) which reduces the computational cost. SVOI contains only pixels carry rich information relevant to the event taking place not the entire video.

This method makes use of four benchmark datasets, i.e. UCSD, UMN, Subway, and U-turn. Two criterions are used for evaluating anomaly detection accuracy namely a frame level criterion and a pixel level criterion. Motion pattern and FRP (False positive rates) are calculated for evaluating performance. And DR(Detection Rate) corresponds to the successful detection rate of the anomalies happening at EER(Equal Error Rate).This issues with this method are that there is no predefined set of anomaly patterns. It depends on the current scenario. One of the main challenges is to detect anomalies both in time and space domains. This implies to find out which frames that anomalies occur and to localize regions that generate the anomalies within these frames.

Hossein Mousa et al [3] presents a novel video descriptor, referred to as Histogram of Oriented Tracklets, for recognizing abnormal situation in crowded scenes. Unlike standard approaches that use optical flow, which estimates motion vectors only from two successive frames, have built descriptor over long-range motion trajectories which is called tracklet. video sequences in spatio-temporal cuboids within which we collected statistics on the tracklets passing through them. Fames are classified as normal and abnormal by using Latent Dirichlet Allocation and Support Vector Machines.

demonstrated (i) very promising results in abnormality detection, (ii) setting new state-of-the-art on two of them, and (iii) outperforming former descriptors based on the optical flow, dense trajectories and the social force model. Three different detection strategies are BW(Fully bag of words),FS(Per-frame, Per-sector) and FiS (Per-frame, Per-independent-sector).

One of the main challenges is to detect abnormalities in densely crowded environments. This implies to isolate the frames where abnormalities occur and to localize within these frames, the area that generated the abnormalities. The other major challenges are that there is no clear definition of abnormalities as they are basically context dependent and can be defined as outliers of normal distributions.

Shuiwang Ji et al [4] put forward a method for the automated recognition of human actions in surveillance videos. Developed a novel 3D CNN model for action recognition Convolutional neural networks (CNNs) are a type of deep model that can act directly on the raw inputs. To boost the performance, it includes regularizing the outputs with high-level features and combining the predictions of a variety of different models.

Limitations of the previous models are that they are limited to handling 2D inputs alone. This model extracts features from both the spatial and the temporal dimensions by performing. 3D convolutions, is achieved by convolving a 3D kernel to the cube formed by stacking multiple contiguous frames together. The developed model generates multiple channels of information from the input frames, and the final feature representation combines information from all channels.

The developed includes model regularization and combination schemes to further boost the model performance. The issues are that , accurate recognition of actions is a highly challenging task due to cluttered backgrounds, occlusions, and viewpoint variations perform 3D convolution in the convolutional layers of CNNs so that discriminative features along both the spatial and the temporal dimensions are captured.3D convolution is achieved by stacking multiple contiguous frames together. The developed 3D CNN model was trained using a supervised algorithm , and it requires a large number of labeled samples.

Chengkun et al[5] proposed an anomaly-introduced learning (AL) method to detect abnormal events. A graph-based multi-instance learning (MIL) model is formed with both normal and abnormal video data. A set of potentially abnormal instances and a coarse classifier are generated by the MIL model. These instances are adopted for an improved dictionary learning, which we call anchor dictionary learning (ADL). The sparse reconstruction cost (SRC) is selected to measure the abnormality. Compared with other methods, this (i) make use of abnormal information and (ii) prune testing instances with a coarse filter and reduce time cost of computing SRC.

This work uses the concept of multi-instance learning (MIL) to solve the task, which utilizes the abnormal event videos as training samples. Moreover, compared with some supervised methods for abnormal event detection, MIL is a kind of weakly-supervised method, which only needs to provide video-level labels, but does not need the finer labels. MIL has been widely used on some other tasks of video, such as object tracking [3, 6], action recognition [2], and video retrieval [24]. Nonetheless, MIL is rarely applied to abnormal event detection. Compared with other tasks, the key point of abnormal event detection is not only to detect when an exception occurs, but specifically to locate where an exception occurs.

The main contributions of the work are as follows:
➤ proposal of a novel approach based on MIL and dictionary learning for abnormal event detection.
➤ utilizing the abnormal videos to improve the performance of abnormal event detection.
➤ employing dictionary learning to further classify the result derived from MIL classifier, which improves classification efficiency.

The experimental results show consistent improvement over the state-of- the-art abnormal event detection methods which only use normal videos. In the future, changes would be made to study the information contained by the abnormal video data.

Jiayu Sunet al [6].Abnormal event detection and localization is a challenging research problem in intelligent video surveillance. It is designed to automatically identify abnormal events from monitoring videos. The main difficulty of this task lies in that there is only one class called "normal event" in training video sequences.

we propose a novel end-to-end model which integrates the one-class Support Vector Machine (SVM) into Convolutional Neural Network (CNN), named Deep One-Class (DOC) model. Specifically, the robust loss function derived from the one-

class SVM is proposed to optimize the parameters of this model. Compared with the hierarchical models, our model not only simplifies the complexity of the process, but also obtains the global optimal solution of the whole process. In the experiments, we validate our DOC model with a publicly available dataset and compare it with some state-of-art methods.

In this paper, deep learning is applied to the challenging video anomaly detection problem. We proposed a deep one-class learning model for abnormal event detection from video sequences by combining CNN and one-class SVM. CNN is utilized to learn the underlying high-dimensional normal representations to effectively capture normal features. One-class SVM layer not only distinguishes normal/abnormal cases as a discriminator, but also optimizes parameters of the whole model as an optimization objective. Moreover, the enhanced objective function based on original one-class SVM makes the robust optimal solution. This method greatly reduces the cumbersome intermediate operation compared with other methods. For future work, we will investigate how to improve the result of video anomaly detection with two-stream deep one-class learning model, exploiting the fusion of spatial and temporal features to generate integrated and comprehensive representations.

III. PROPOSED SYSTEM

The proposed work consists of three modules. The first module is object detection module, the second module is behaviour analysis, and the third module is alert module.
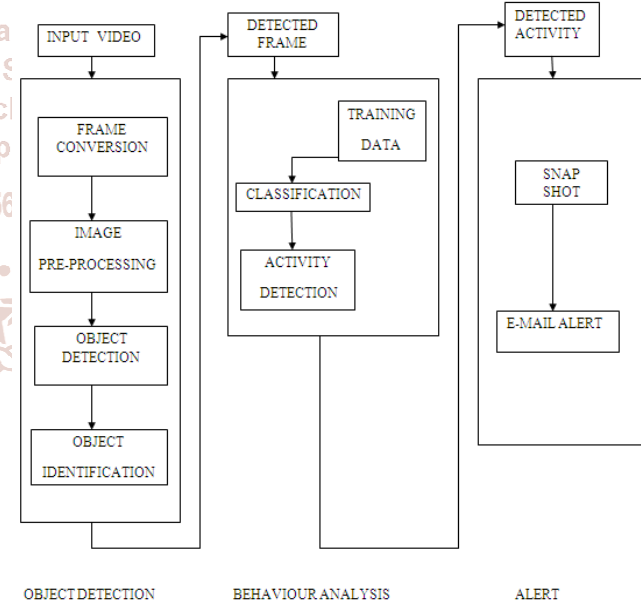

Fig 3.1 Functional module

The first module takes the CCTV live video as the input. The video is converted into frames in the frame conversion block which uses Sum Of Differences algorithm(SAD).Each frame from the frame conversion block is sent to the image processing module, where the edge distortions and high quality frames are produced. These frames are then processed using the Convolution Neural Networks (CNN). After the detection of the object is alone taken and sent to the second module to classify if the object is knife or an iron rod.

The behaviour analysis module takes the detected frame as input. The input is given to classification sub stage where it

uses support Convolutional neural networks (CNN).Classification Sub stage gets the supporting reference frames from training dataset. The Alexnet dataset consists of various annotated objects. Based on this it classifies the given frame as normal or abnormal events. The detected activity is sent to the third module.

The detected activity is sent to the alert system module for classification. In the activity classification sub stage the activity is classified as abnormal activity based on the training data set. When ever an abnormal activity is found it is sent to the alert system for providing email alerts to the operator along with the snapshot of the criminal activity caught in the CCTV.

IV. SYSTEM IMPLEMENTATION AND PERFORMANCE ANALYSIS

A. IMPLEMENTATION OF MODULES

We carried out the implementation part in MATLAB which has produced the results. On execution, we have got the following results for each module. The time efficiency is shown in the performance analysis part.

Module 1 is provided with a video of duration 12 seconds. The given video is converted into 180 frames and then it is processed. The processing of each frame includes various image pre-processing techniques such as pixel subtraction and gray scale conversion. And the border outline values for the detected object is calculated and used for highlighting. The output of this module is the frame image that contains the weapon along with highlighting the boundaries of it. And also it gives the additional information in a form of dialog box on what type of weapon is detected.
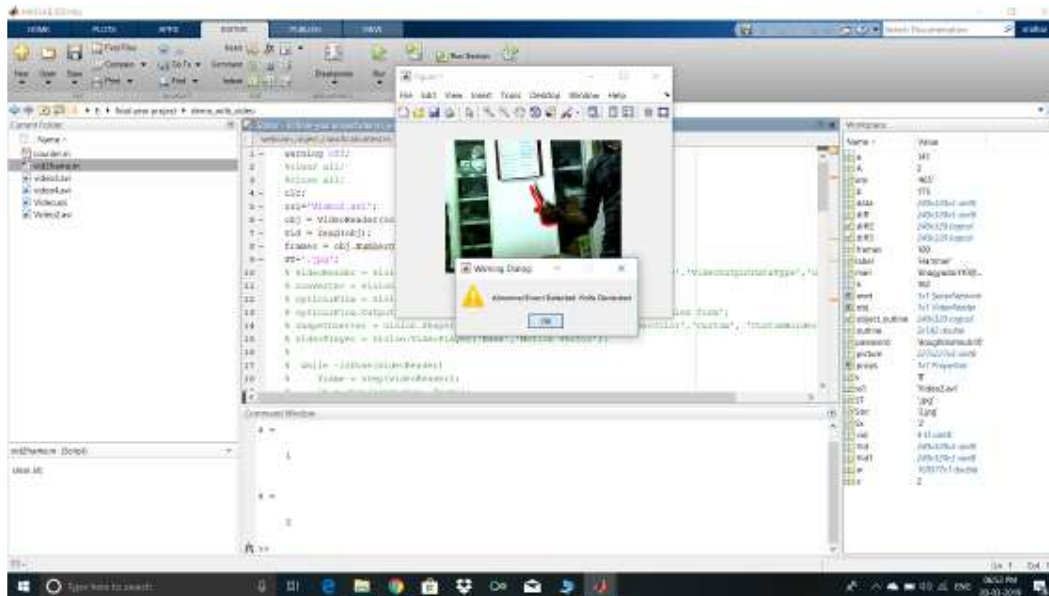

Fig 4.1.1 Module one result screenshot

Module 2 is given with the input of live video. And also it takes the ALEXNET neural network as the input training dataset.Each frame is converted into ycbrc color space and resizes the frame into 227:227 ratio.The frame which is detected along the dataset object in ALEXNET is fed into the classification phase which uses CNN. When the label matches with the given input frame,it gives a warning alert on the operator screen along the weapons name.
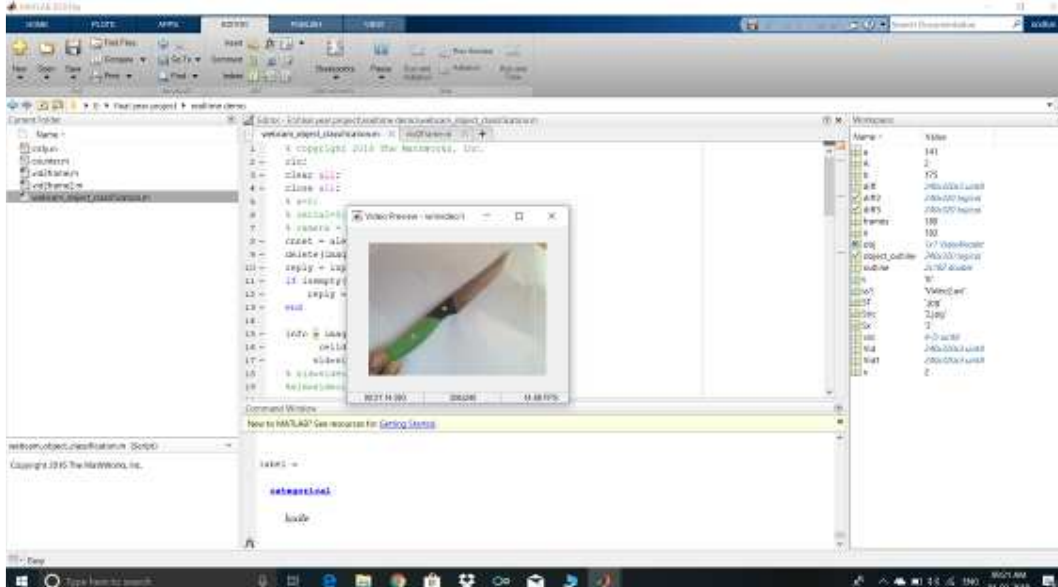

Fig 4.1.2 Module two result screenshot

Module 3 takes the detected frame as input from module 2. The received frame is further fed into the classification phase with is based CNN. This phase classifies the input frame as either normal or abnormal event based on the input training dataset. If it is classified as normal event no action is performed. Orelse for abnormal event, snapshot of the crime scene is taken and sent as email to the operator. Additionally voice alert is also given in the control station.
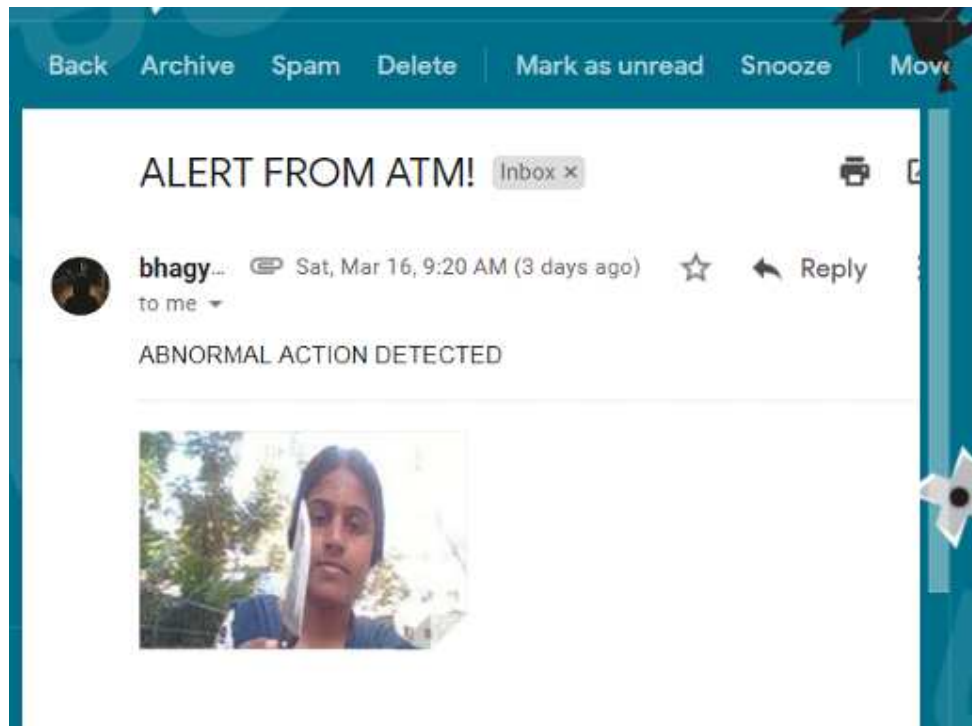

Fig 4.1.3 Module three result screenshot

B. PERFORMANCE ANALYSIS

Our work is done in Matlab. The parameter that we have taken for calculating the performance of the proposed system is processing time. We conduct an experiment on the evaluation of the overall runtime of the proposed system. All experiments are carried out on a computer with an octa-core Intel i5 2.50 GHz processor and the following results are achieved. The achieved results in [1] is as follows.

METHOD	PLATFORM	GPU	MEMORY	RUNNING TIME(IN SECONDS)
SPARSE RECONSTRUCTION	MATLAB	-	2.0GB	3.8-4.9

Table 4.2.1 Results of the work [1]

Our work produced the following results

METHOD	PLATFORM	GPU	MEMORY	RUNNING TIME(IN SECONDS)
CNN	MATLAB	-	8.0GB	2.5-3

Table 4.2.2 Results of our proposed work

The time efficiency of our work is displayed in the graph below.
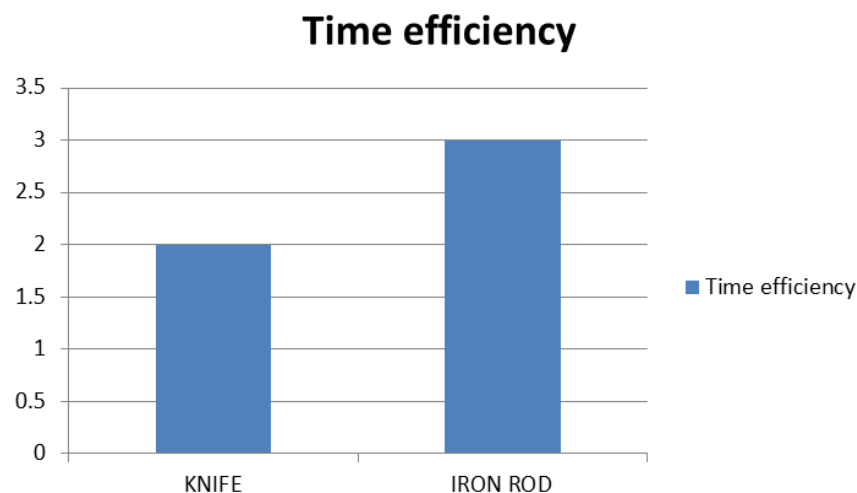

Fig 4.2.3 Processing time for knife and iron rod in our proposed work

V. CONCLUSION AND FUTURE WORK

In this paper,we propose a common detection framework for detecting abnormal events.In our method, real time surveillance video that carry rich motion information are fed to train the CNN model for anomaly detection.For extracting features we use HOG(histogram of gradients).Since our work is applicable only in confined areas, in future the challenges of crime detection in roads can be addressed.imaging techniques based on a combination of sensor technologies and processing will potentially play a key role in addressing the weapon detection problem.

VI. REFERENCES

[1] Qichang Hu , Sakrapee Paisitkriangkrai , Chunhua Shen, April 2016,"Fast Detection of Multiple Objects in Traffic Scenes With a Common Detection Framework", IEEE Transactions on Intelligent Transportation Systems, Volume: 17, Issue: 4 , 1002 - 1014

[2] C. He, J. Shao, and J. Sun , "An anomaly-introduced learning method for abnormal event detection," Multimedia Tools and Applications, pp. 1–16, 2017.

[3] Hossein Mousavi Sadegh Mohammadi Alessandro Perina Ryad Chellali Vittorio Murino,"Analyzing Tracklets for the Detection of Abnormal Crowd Behavior"2015 IEEE Winter Conference

[4] Shuiwang Ji, Wei Xu, Ming Yang, Member, IEEE, and Kai Yu,"3D Convolutional Neural Networks for Human Action Recognition", Volume: 35 , Issue: 1 , Jan. 2013

[5] Chengkun He, Jie Shao, Jiayu Sun, "An anomaly-introduced learning method for abnormal event detection", Multimedia Tools and Applications November 2018, Volume 77, Issue 22, pp 29573–29588

[6] Jiayu Sun,Jie Shao, Chengkun He, "Abnormal event detection for video surveillance using deep one-class learning", Multimedia Tools and Applications February 2019, Volume 78, Issue 3, pp 3633–3647

[7] F. García-Rial, L. Úbeda-Medina, and J. Grajal, "Real-time GPU-based image processing for a 3-D THz radar," IEEE Trans. Parallel Distrib.Syst., vol. 28, no. 10, pp. 2953–2964, Oct. 2017.

[8] J. Grajal et al., "Compact radar front-end for an imaging radar at 300 GHz," IEEE Trans. THz Sci. Technol., vol. 7, no. 3, pp. 268–273,May 2017.

[9] V. Krozer, "Terahertz imaging systems with aperture synthesis techniques," IEEE Trans. Microw. Theory Techn., vol. 58, no. 7,pp. 2027–2039, Jul. 2017.

[10] D. A. Robertson et al., "High resolution, wide field of view, realtime 340 GHz 3D imaging radar for security screening," Proc. SPIE, vol. 10189, p.101890C, May 2017.

[11] Cheng, K.W., Chen, Y.T., Fang, W.H.: Video anomaly detection and localization using hierarchical feature representation and gaussian process regression. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2909–2917(2015)

[12] P. Viola and M. J. Jones, "Robust real-time face detection," Int. J. Comput.Vis., vol. 57, no. 2, pp. 137–154, May 2016.

[13] Adam, A., Rivlin, E., Shimshoni, I., Reinitz, D.: Robust real-time unusual event detection using multiple fixed-location monitors. Pattern Analysis and Machine Intelligence, IEEE Transactions on 30(3), 555–560 (2016)

[14] Ali, S., Shah, M.: A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In: Computer Vision and Pattern Recognition, 2007. CVPR' 07. IEEE Conference on, pp. 1–6. IEEE (2015)

[15] Benezeth, Y., Jodoin, P.M., Saligrama, V., Rosenberger, C.: Abnormal events detection based on spatio-temporal co -occurences. In: Computer Vision and Pattern Recognition,2009. CVPR 2009. IEEE Conference on, pp. 2458–2465. IEEE (2017)

[16] Bertini, M., Del Bimbo, A., Seidenari, L.: Multi-scale and real-time non-parametric approach for anomaly detection and localization. Computer Vision and Image Understanding 116(3), 320–329 (2014)

[17] Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. the Journal of machine Learning research 3, 993–1022 (2016)

[18] Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: A survey. ACM computing surveys (CSUR) 41(3), 15 (2015)

[19] Cheng, K.W., Chen, Y.T., Fang, W.H.: Video anomaly detection and localization using hierarchical feature representation and gaussian process regression. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2909–2917 (2015)

[20] Cong, Y., Yuan, J., Liu, J.: Sparse reconstruction cost for abnormal event detection. In: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pp.3449–3456. IEEE (2015)

[21] Cong, Y., Yuan, J., Liu, J.: Abnormal event detection in crowded scenes using sparse representation. Pattern Recognition 46(7), 1851–1864 (2016)