Detection of Suspicious Behaviour in ATM Site Using Computer Vision Techniques

[1]Sharath babu CG, [2]Dr. Anitha Devi.M.D, [3]Dr.M Z Kurian

[1]PG, [2]Associate Professor, [3]Head of Department
[1]Electronics and Communication Engineering, Sri Siddartha Institute of Technology, Tumkur, Karnataka

Abstract: Many individuals find that using an automated teller machine (ATM) provides them with a great deal of convenience since it enables them to do bank transactions and cash withdrawals in a much shorter amount of time. However, if an ATM is left unattended outside of normal business hours or on a public holiday, it leaves itself open to the possibility of being attacked. In recent months, it has been reported that a number of automated teller machines have been stolen from their locations or deliberately destroyed in order to get access to the cash they contain. Because of this, many automated teller machine locations actually have video surveillance systems installed to keep an eye on the surrounding area and deter criminal activity. When there are a large number of surveillance cameras, however, it becomes more difficult for security officers to locate the location of an ongoing crime in real time. The goal of this research is to build real-time video analysis algorithms based on computer vision methods and deep learning technologies for artificial neural networks, with the goal of recognizing anomalous human behavior near essential facilities, such as automated teller machines.In this article, a real-time security expert video surveillance system that utilises image processing methods to identify suspicious behaviour was presented as a means of combating the problem. The paper begins with a discussion of the typical approach to the detection and recognition of suspicious activity. The supervised and unsupervised machine learning approaches, which are generally based on SVM, HMM, and ANN classifiers, are then summarised. These methodologies were adopted by the researchers in the past, and they range from modelling single human behaviour to modelling crowded scenes.

Keywords: ATM surveillance, Deep Learning, Machine Learning, Computer Vision

1. Introduction:

Face recognition [1, 2], traffic monitoring [2, 3], the detection of suspicious objects, and the investigation of crimes are just some of the functions that can be significantly expanded with the help of computational analytics when video surveillance is combined with it. Video surveillance is an essential part of the security infrastructure of contemporary cities. The transmission and processing of huge volumes of video data in real time is required for key video surveillance applications for public safety, such as monitoring and tracking, which results in considerable delays for communication networks [4]. In order to properly understand the data, there is an increasing need for human resources. Because of this, standard video surveillance systems are now built for offline forensic analysis rather than being utilised as a preventative measure to stop suspicious behaviour before it does harm. Second-generation video surveillance systems, which have recently been developed, aim to reduce the need for human operators. Instead, clever machine learning algorithms in the cloud analyse the gathered video frames to recognise, track, and report any unexpected events [5]. Human operators are minimised in [second generation] video surveillance systems. The introduction of video analytics technology has led to success in a range of areas in recent decades, including the detection of emergency situations [6, 7], traffic control and management [8, 9]. The usage of video analytics tools was aided by the accessibility, progress, and cost of depth vision processors and sensors, as well as the production of security cameras with intelligent features [5].

We present a strategy in this research that blends computer vision and pattern recognition techniques with recent video recording devices. This strategy, which uses automated teller machines as an example, has the potential to enhance the development of algorithms for recognising people's physical movements in real time and notifying security professionals in the event of odd behaviour near vital facilities. In addition to that, this method may be of assistance in the creation of algorithms that can recognise the bodily motions of humans in real time. In a recent post, we discussed a method for recognising static hand gestures that makes use of a deep convolutional neural network and the idea of learning transfer. This method is explained here. Publications [10] and [11] describe this strategy. The suggested model has been tested extensively, and the results show that it achieves a high degree of recognition accuracy.

2. Related Works

These days, automated teller machines (ATMs) have evolved into one of the fundamental requirements of contemporary living and are frequently used for a variety of financial transactions. On the other hand, there has been a rise in the number of crimes connected to ATMs. In general, crimes committed against automated teller machines may be broken down into three distinct categories. In addition, it is essential to bring attention to the many fraudulent and unlawful activities that are directed at persons who use an automated teller machine. Every kind of criminal act is followed by a certain set of movements on the part of the perpetrator as well as a specific set of suspicious or aberrant conduct. Smart motion detectors are incorporated into contemporary video surveillance systems for the purpose of enhancing public safety. These motion detectors can only distinguish between the most basic sorts of motion, such as "loitering," crowding, a person falling, carrying a weapon in their hands, and abandoning a suspicious object. A large number of research have been conducted in the field of human activity recognition (HAR), all of which show that the observation environment in issue must meet a set of requirements [12].

These criteria helped to pave the way for the development of specific HAR instruments, both in terms of hardware and software (such as CCTV cameras) (OpenCV libraries, etc.). Face recognition [1, 13], hand gestures [10-11,14], crowd behaviour [15, 16],

and abnormal conduct [17, 18] are some of the most used ways. The majority of security measures concentrate on preventing crimes from happening. To arrive at a conclusion, the authors of the research [17] propose an algorithm that includes the phases of capturing motion sequences from a sensor. On the other hand, the system may only sound an alarm once an individual has already misplaced their belongings.

The study [18] gives a practical method for the identification of thirteen distinct suspicious behaviour patterns, including break-ins, hand-to-hand fighting, shooting, and vandalism, among others. They categorise patterns as either normal or abnormal and then utilise a three-dimensional convolutional neural network (CNN) to extract characteristics from the patterns. Recent methods for detecting wireless activity have been shown to be effective, despite the fact that they need for the use of specialist gear. A model was given by the authors of [19] for the propagation of WiFi signals in physical space when there is also human activity present in the environment. Human motor activity has been found to impede wireless signal transmission from the access point to the receiver, which results in a significant shift in the phase and amplitude information that is referred to as the Channel State Information (CSI). These percentages are broken down as follows: in the line of sight, the accuracy is 89.2 percent; outside the line of sight, it is 85.6 percent. In a recent work [20], we investigated the issue of navigation inside contemporary structures that had a complicated structure by making use of a model for the propagation of WiFi signals. As part of this project, we came up with our own unique technique for locating things, as well as an augmented reality application that can be used for navigating inside spaces.

In addition, there are a number of studies that were done in the past that are relevant to the characterization of aberrant conduct of persons who are near ATMs. The authors of the paper [21] advocated the installation of a monitoring system at ATM sites in order to identify any unusual behaviour. The technology makes use of Kinect 3D cameras and has the ability to differentiate between an aggressive and regular stance in a person. Work [22] provides a system that classifies routine and possibly illegal actions related with theft in the vicinity of ATMs. A system that recognises aberrant behaviours targeted at users of ATMs that take place in rooms with ATMs is suggested in [23]. This system would be located in rooms with ATMs. In order to determine aberrant human behaviour, the system implements a method that incorporates the application of motion history images (MHI) and Hu moments to the video in order to extract significant elements. The histogram of directed gradients (HOG) and the Random forest machine learning approach are used in the article [24] to categorise human behaviours into normal and unhealthy. This classification is based on the findings of a study.

The ArchCam expert system, built by the study's authors [25], can detect suspicious behaviour in the vicinity of an automated teller machine (ATM). This assumption is based on the fact that it is possible to assume that such an approach will allow for the recognition of suspicious human movement.

3. ACTIVE FIELDS IN SUSPICIOUS HUMAN ACTIVITY DETECTION FROM VIDEO SURVEILLANCE

Few applications of recognising abnormal human activity that have primarily captured the attention of researchers are the following: the ATM centre; the detection of crowd anomalies;

3.1 A. Anomaly in ATM Center

Even though they are monitored around the clock, Automatic Teller Machine (ATM) facilities remain one of the most susceptible locations for illicit activity. According to the findings of the review study [32], the majority of the research effort that has been done has focused on identifying concealed faces and unlawful items inside the ATM facility. Identifying odd or suspicious behaviours has only been the subject of 4% of the study conducted so far. The most typical technique is known as supervised learning, and it makes use of the SVM classification strategy. In the study referenced by [33], the scientists extracted skeletal data from the depth picture using a 3D camera similar to the Kinect. The authors of article [34] suggested a unique model that detects anomalous activity by using a Convolutional Neural Network (CNN) with Long Short-Term Memory (LSTM). This model was described in the study. Videos from the security camera and CNN's equipment that were equipped for recognising and extracting the relevant elements from the frames of the video were incorporated as input to the model.

3.2 Fall Detection

One of the most important research areas in computer vision is the automatic detection of human falls. A large number of articles have been produced and published on the topic of fall detection in the senior home care system. More recent studies have shifted away from this device-based strategy. As a result, a computer vision-based technique is gaining traction not only for detecting a pedestrian fall in an indoor context, but also for detecting a pedestrian fall in an outdoor setting [35, 36]. This is because computer vision is becoming more useful in both of these settings. The camera location is arbitrary, the people are free to walk around, and there are a variety of covariate elements such different view angles, lighting, and clothes. These are the main problems with both scenarios. This RNN forecasts the temporal dynamics of a person's 2D posture data after they have fallen.

3.3. Crowd Anomaly Detection

The identification of crowd anomalies is a significant research problem in the fields of both computer vision and video analysis. Its uses include areas such as public transit hubs, pilgrimage sites, public or private gatherings, sporting events like cricket or football, crowded streets and marketplaces. It is reasonable to presume that those in a crowd who are participating in fighting, shoving, or collapse, as well as crowd panicking, are engaging in aberrant actions. A crowd running in a marathon may be considered to be a normal activity, however if people suddenly started running in an open market, it might be considered an emergency situation [38]. This is an excellent illustration of the difference between a regular activity and an abnormal activity. In certain instances, anomalous occurrences may also involve a person hovering at a location for an unusually extended period of time. Human behaviour that involves idling often results in unusual occurrences at bus or train terminals, including but not limited to pickpocketing, snatching chains, robbery, and abduction in the surrounding residential neighbourhood, amongst other things. The authors of [40] suggested a Markov random walk model that is able to reliably identify lingering persons in any outdoor public space.

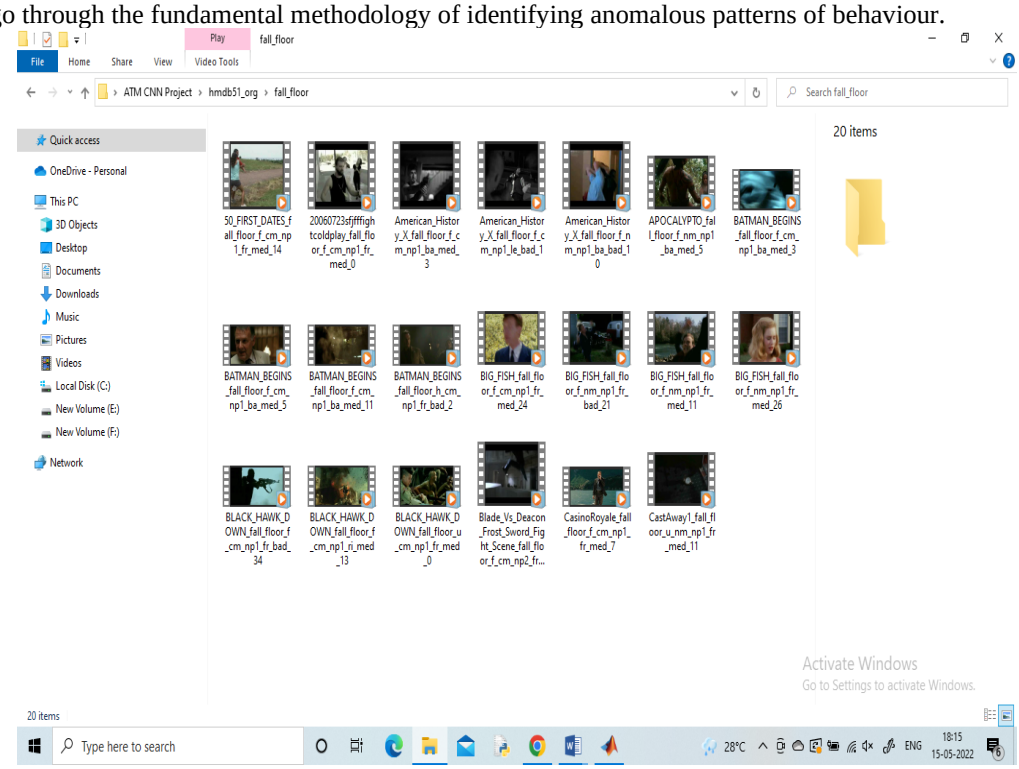
3.4 Detecting suspicious activities in the Examination hall

Monitoring of examination rooms by human invigilators is standard practise in every region of the globe. This includes activities such as hand signalling. The monitoring of an examination room is a difficult undertaking in terms of the amount of staff required.

It is essential to design and implement a surveillance system that will provide the educational institution with assistance in monitoring the examination room. The identification of several faces, head and hand movement, gesture, and eye gazing are all necessary components in the automatic detection of suspicious conduct. The most effective method is supervised learning, such as classification, since it is straightforward to categorise specific behaviours as typical or unusual. The authors of [41] employed a training model such as an Artificial Neural Network to extract automatic facial recognition and hand identification from surveillance films with the assistance of skin colour. The authors of [42] regarded eye gaze and head position information to be the sole possible signals for detecting suspicious behaviours shown by the pupils.

4. Proposed Methodology

The fundamental concept for identifying the anomalous occurrence or behaviour in video sequences is shown in Figure 2. For the purpose of detecting aberrant behaviour identification in video, the majority of researchers in previous study literature employed the following methods. The phase of selecting and extracting features is a crucial one because, unlike other steps, it does not alter the actual representation of the object and, at the same time, it maintains the object's true semantics. These approaches have the additional benefit of being applicable to both supervised and unsupervised learning, which is still another advantage of using them. In this part, we will go through the fundamental methodology of identifying anomalous patterns of behaviour.



2.1 Preprocessing

The first stage in the process of processing video is called preprocessing. To improve the quality of the frames that have been extracted from the video, preprocessing is done on them. The open function in MATLAB allows for the application of a morphological operation, which may be used to increase the quality of blob recognition of moving objects in video. In order to enhance the quality of blurred images, We might use filters like the Gaussian filter, the average filter for grain-like noise, or the Wiener filter.

2.2 Motion detection algorithm After The next step in this process is the preprocessing phase, which is followed by a motion detection algorithm that figures out the area where motion is happening or the area of interest. There are four main algorithms for finding motion [15–18]. Background subtraction is the first one [5, 15, 16], and it is the method that is used the most often for foreground detection. One more typical technique for background removal is referred to as temporal differencing [49]. Following that is the use of statistical techniques [59], and finally, flow analysis [18, 20, 49].

2.3 Feature detection and extraction Feature detection is a term used in the fields of pattern recognition and image understanding to describe how to find important spots or points of interest that might tell you what a picture is about. Examples of such features include edges, corners, texture, and blobs, among other things. Because big input volumes contain more data but less information, feature extraction is essentially a method for decreasing the dimensionality of the data. This is accomplished by reducing the number of features, which are referred to as feature vectors. Therefore, it is necessary to translate enormous amounts of data into some kind of collection of characteristics. Now that the main issue has been identified, one of the most crucial tasks is to pick the appropriate feature. Table 1 contains a list of several feature detection and extraction techniques, together with the years in which they were first published.

2.4 Classification

The process of splitting the pattern into several groups is called classification. The input data are sorted into the specified number of categories as a result of this procedure. Most of the time, figuring out what is normal and what isn't is a two-class problem. In this literature review, we talked about the different types of classifiers that researchers have used, like the HMM, which is a sequential classifier, neural networks, decision trees, rule bases, and fuzzy classifiers and so on. Detection of anomalies using supervised learning techniques (classification) The initial step in the process of supervised learning is to train the model on the basis

of labelled training data. Once the model has been trained, the next step is to classify additional test data so that it may be assigned to one of the available classes. The process of learning via supervision consists of two stages. In the first step, called "training," participants use the labelled data to create a classifier. In the second step, which is called "validation and verification" or "testing," participants compare the new data they have entered to the model they have been taught. The issue is divided into normal and abnormal categories using a variety of different classifiers in the literature. For example, support vector machines are designed to solve problems with two classes. The number of classifiers that are used in abnormal event detection is the topic that will be covered in the next subsection.

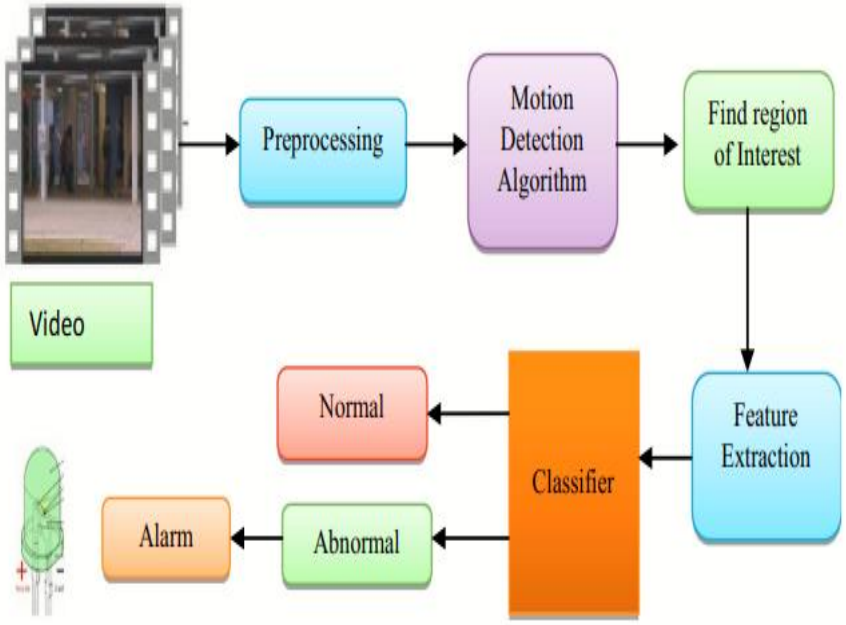


Figure 1 Proposed Methodology

3.2 Neural network based approaches

A neural network is a classifier strategy that can be employed in either a single class issue or a multi class problem. This allows neural networks to be used for anomaly detection in movies. The first step, which is called the training or learning phase, consists of training the network based on the labelled training data. The second step, which is also called the testing phase, consists of testing, validation, and verification. Many distinct varieties of work dhas been one on neural networks. In [27], the author discusses the hardware and software prerequisites that are necessary. The Bayesian approach is used to construct the phase variable model for use in software. The categorization of an object's target is accomplished with the use of back propagation. For the purpose of observing a target's behaviour, an expert system is used, which not only maintains numerous rules for decision making but also sounds an alarm. A strategy based on FAM (fuzzy associative modelling) was presented by Wang et al. [17].

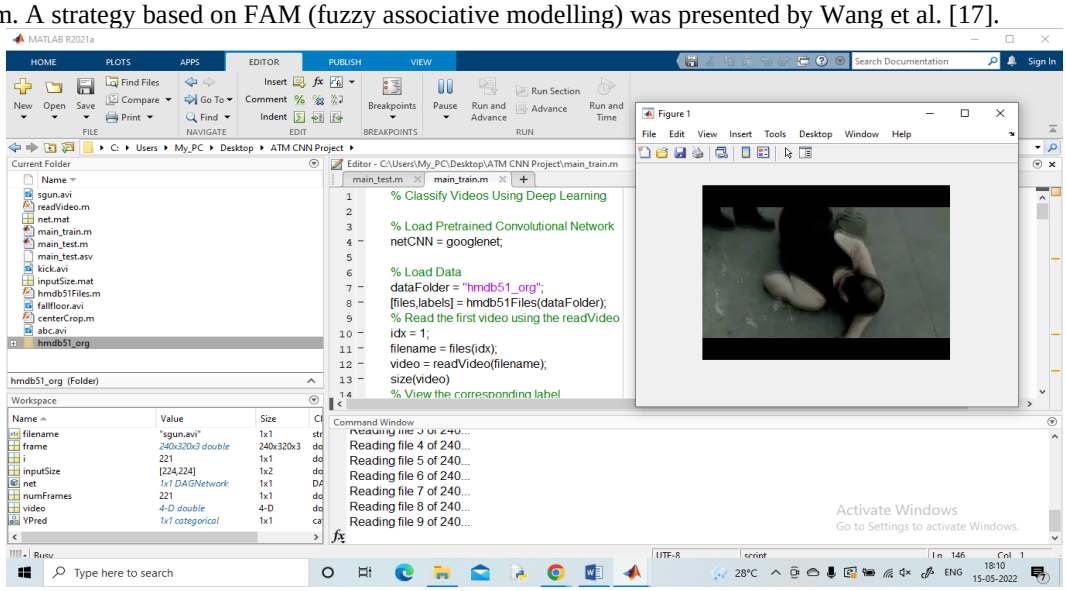


Figure 2 Single Frame Displayed

Conclusion

For this article, we did in-depth research on high-quality research papers in the field of computer vision, specifically in the area of detecting abnormal behaviour in a single object or in a crowd. We have provided a synopsis of each work in Section 2. The term "anomalous activity detection" refers to the process of identifying uncommon occurrences, such as suspicious behaviour, violence detection, mask detection, and so on. But these techniques don't work well with backgrounds that are busy, don't stay still, or have a lot of moving parts, like tree branches or a background with the wind blowing. Many of the currently available algorithms have

been accurately implemented by taking into account just some features of single human behaviour modelling, such as changes in the position of the mouth and nose, eye movement, and so on. They might be put into practise by taking into consideration a greater number of components, such as a person's head movement and their stride. Because it sets up the scene's structure, designing real-time recognition for multicamera systems and calibrating the cameras will continue to be big problems for 3D computer vision research in the future. In the future, the already-existing algorithm will need to be tested with a huge set of real-time video data to make sure it works the same way every time and improve its accuracy. The time allotted for the validation process has to be increased to a number of minutes, up from a few seconds. The use of solely structured video scenes is a requirement for the majority of anomaly detection algorithms. Because of this, it is very necessary for the newly developed algorithms to be compatible with unstructured video situations.

References

1. Ji P, Kim Y, Yang Y and Kim Y S 2016 Face occlusion detection using skin color ratio and LBP features for intelligent video surveillance systems FedCSIS (Gdansk) ACSIS (Polish Information Processing Society) V 8 pp.253-259 https://doi.org/10.15439/2016F508
2. Zhou Y, Liu L, Shao L and Mellor M 2018 Fast automatic vehicle annotation for urban traffic surveillance IEEE Trans. Intell. Transport. Syst. (IEEE) vol. 19 no. 6 pp.1973– 1984
3. Puvvadi U L N, Benedetto K D, Patil A, Kang K-D and Park Y 2015 Cost-effective security support in real-time video surveillance IEEE Trans Ind. Informat. (IEEE) vol. 11, no. 6 pp.1457–1465
4. Prakash P, Suresh R and Pn D. K 2019 Smart city video surveillance using fog computing Int. J. of Enterprise Network Management (Springer, Cham) vol. 10, no. 3-4 pp. 389-399 [5] Shidik G. F. et al. 2019 A Systematic Review of Intelligence Video Surveillance: Trends, Techniques, Frameworks, and Datasets IEEE Access (IEEE) vol. 7 pp.170457-170473
5. Filonenko A, Hernández D C and Jo K-H 2018 Fast smoke detection for video surveillance using CUDA IEEE Trans. Ind. Informat. (IEEE) vol. 14, no. 2 pp. 725–733
6. Finogeev A, Finogeev A, Fionova L, Lyapin A and Lychagin K A 2019 Intelligent monitoring system for smart road environment J. Ind. Inf. Integr (Amsterdam: North-Holland/American Elsevier) vol. 15 pp. 15–20
7. Murugan A S, Devi K S, Sivaranjani A and Srinivasan P 2018 A study on various methods used for video summarization and moving object detection for video surveillance applications Multimedia Tools Appl.(Springer, Singapore) vol. 77, no. 18 pp. 23273–23290
8. Zhao L, He Z, Cao W and Zhao D 2018 Real-time moving object segmentation and classification from HEVC compressed surveillance video IEEE Trans. Circuits Syst. Video Technol. (IEEE) vol. 28, no. 6 pp. 1346–1357
9. Satybaldina D., Kalymova G. and Glazyrina N. 2020 Application Development for Hand Gestures Recognition with Using the Depth Camera Communications in Computer and Information Science (Springer, Cham) V.1243 pp. 55-67
10. Satybaldina D Zh, Ovechkin G V and Kalymova G A 2020 Static hand gestures recognition system with using depth camera Vestnik of RSREU (Ryazan: Ryazan State Radio Engineering University Press) N. 72 pp.93-105. https://doi.org/10.21667/1995-4565-2020-72-93-105
11. Martínez-Mascorro G A et al. 2005 Suspicious Behavior Detection on Shoplifting Cases for Crime Prevention by Using 3D Convolutional Neural Networks (arXiv preprint arXiv:02142v1[cs.CV])
12. Wang Y, Bao T, Ding C and Zhu M l 2017 Face recognition in real-world surveillance videos with deep learning method IEEE 2nd International Conference on Image, Vision and Computing (Chengdu) (IEEE) pp. 239-243
13. Satybaldina D Zh Kalymova G A Stepanov V S 2020 Application for recognition of static gestures by using deep images Certificate of state registration of the right to the computer program, № 10319 (National Institute of Intellectual Property)
14. Direkoglu C, Sah M and O'Connor N E 2017 Abnormal crowd behavior detection using novel optical flow-based features 14th IEEE International Conference on Advanced Video and Signal Based Surveillance(Lecce) (IEEE) pp. 1-6
15. Rabiee H et al. 2018 Detection and localization of crowd behavior using a novel tracklet-based model International Journal of Machine Learning and Cybernetics (Springer Nature)V. 9, №. 12 pp.1999-2010
16. Tsushita H and Zin T T 2018 A Study on Detection of Abnormal Behavior by a Surveillance Camera Image Int. Conference on Big Data Analysis and Deep Learning Applications (Miyazaki) (Springer, Singapore) pp. 284-291
17. Sultani W, Chen Ch and Shah M 2018 Real-world anomaly detection in surveillance videos IEEE/CVF Conference on Computer Vision and Pattern Recognition (IEEE) pp.6479-6488.
18. Zhu D et al. 2017 A ubiquitous WiFi-based abnormal activity detection system IEEE International Joint Conference on Neural Networks pp.1766-1773
19. Verma, K.K., Singh, B.M. & Dixit, A. A review of supervised and unsupervised machine learning techniques for suspicious behavior recognition in intelligent surveillance system. *Int. j. inf. tecnol.* **14,** 397–410 (2022). https://doi.org/10.1007/s41870-019-00364-0