Video Surveillance for Dangerous Situations in Public Spaces

Nikita Bazhenov, Egor Rybin, Dmitry Korzun
Petrozavodsk State University (PetrSU)
Petrozavodsk, Russia
{bazhenov, rybin}@cs.petrsu.ru,
dkorzun@cs.karelia.ru

***Abstract*—Video surveillance system (VSS) tracks situation in certain area, including recognition of located objects. Multi-camera computer vision can be used to achieve high precision of situation recognition. In this paper, we consider VSS for recognition of dangerous situations in public space. The real-time video analytics is required. Our VSS prototype implements the following recognition functions. 1. Detecting a dangerous person. 2. Identifying an unattended object. 3. Recognizing a fallen person. The recognition is event-based, which supports tracking a dangerous situation from its start to the end. Our experiments show that the proposed recognition algorithms are able to run on relatively low performance devices and cameras, and the existing digital infrastructure of public spaces can be utilized.**

I. Introduction

Video surveillance system (VSS) tracks situation in certain area, including recognition of located objects as well as events where the objects participate [1]. Multi-camera computer vision can be used to achieve high precision of situation recognition. The recognition algorithms operate in real-time and require high performance digital infrastructure.

A particular VSS class is recognition of abnormal situations (dangerous situations, safety risks) in public spaces [2]. In large cities, public spaces have advanced digital infrastructure that can be effectively used in detection of violence and other abnormal behavior [3]. In regions, the digital infrastructure exists in many public spaces. Nevertheless, its computation and networking capabilities are relatively low (for real-time video analytics).

In this paper, we show the results of our experimental study with our VSS prototype. The experiment considers the following recognition functions.

1) Detecting a dangerous person.
2) Identifying an unattended object.
3) Recognizing a fallen person.

The proposed recognition is event-based. Tracking a dangerous situation is implemented from its start to the end. Our experiments show that the proposed recognition algorithms are able to run on relatively low performance devices and cameras, and the existing digital infrastructure of public spaces can be utilized, with no introduction of many advanced equipment elements.

The rest of the paper is organized as follows. Section II describes modern solutions in video analytics systems based on abnormal activity. Section III presents the developed VSS prototype for recognition of dangerous situations in public spaces. Section IV shows the implemented recognition algorithms and the results of their experimental performance. Section V summarizes the key findings of this R&D study.

II. Video Analytics for Safety of Public Space

Video analytics systems for recognizing dangerous situations have gained significant importance in the field of AI development. The primary objective is to enhance surveillance technologies by incorporating AI analysis algorithms that can effectively identify potential threats and security incidents. By implementing these advanced technologies, public spaces can be made safer and more secure. The use of video analytics systems in public safety has its own advantages.

One of the main features is that it can help to reduce the level of crime by detecting and preventing dangerous situations before they can be escalated [4]. These systems are capable of monitoring and analyzing video footage in real-time, enabling quick response to potential threats. Moreover, video analytics systems can assist in criminal capture by identifying individuals involved in illegal activities [5]. These systems are designed to detect suspicious behavior, unauthorized access, or the presence of known criminals, enabling law enforcement agencies to take prompt action and apprehend offenders. Additionally, these systems play a crucial role in collecting incriminating evidence [6].

AI algorithms can analyze video footage to extract relevant information, such as the identification of suspects, license plate numbers, or the sequence of events leading to a security incident [7]. This evidence can be utilized in investigations [8] and legal proceedings, aiding in the prosecution of criminals. Ultimately, the application of video analytics systems ensures a more secure daily life for citizens. By incorporating AI technologies, public spaces can be constantly monitored, enabling the prevention of potential threats. This creates a sense of safety and reassurance among the public, enhancing their overall well-being. In conclusion, the development of video analytics systems for recognizing dangerous situations is a significant area of focus in AI development. These systems have the potential to improve public safety, reduce crime, assist in criminal capture, collect incriminating evidence, and provide a more secure daily life for citizens. By leveraging the capabilities of AI analysis algorithms, surveillance technologies can be enhanced to effectively identify and mitigate potential threats and security incidents based on data protection [9].

Video analytics technologies, computer vision algorithms, and machine learning play a crucial role in proactively detecting dangerous situations in today's world. Detection of dangerous situations in crowded places is not only an important public problem but also a security problem in various educational institutions. An analysis of the current market situation shows that security in public places is the most in demand due to regular incidents of theft, hooliganism, and murders in schools and universities. The collection of the next articles presents a comprehensive illustration of the recent advancements in violence detection and anomaly recognition, all of which are critical aspects of video surveillance analytics.

Multiple works based on the topic of recognizing dangerous situations in public places using cutting-edge techniques with video analytics algorithms, machine learning, and neural networks were analyzed. In [3] authors propose a system that leverages computer vision algorithms to detect and track people, followed by recognizing suspicious activities. The system fuses various parameters, such as movement, objects, and behavior, to address challenges in dynamic environments. Add: adopt deep learning techniques, including Convolutional Neural Networks (CNN). Such an approach harnesses the power of advanced neural networks in detecting, classifying, and predicting violent events in real-time, thus augmenting the value of video surveillance systems. The work emphasizes the need to update and train models on real-world, sometimes unclean, data to ensure their robust performance.

In [10] authors present an overview of a large number of methods for detecting dangerous situations and anomalies, including public places. This paper also provides an overview of existing datasets with abnormal human behavior. Such experience can be very useful when creating AbHAR (Abnormal Human Activity Recognition) services. The work covers various forms of deep learning, including convolutional neural networks (CNN), recurrent neural networks (RNN), and generative adversarial networks (GAN), and how these have been utilized and optimized for detecting anomalies in video surveillance. The authors highlight the potential of deep learning in improving surveillance systems and point out key challenges and prospective research directions.

General methods are described in [11]. A systematic review of current violence detection techniques in the field of video surveillance systems is provided. The exploration of current methodologies and state-of-the-art techniques presents a comprehensive understanding of the field's advancement. The authors underscore the shift towards machine learning and deep learning methods over classical techniques which can be very useful when working with video surveillance systems.

In [12] a framework to uncover real-world anomalies in surveillance videos is introduced and evaluated. The authors address the limitations of existing datasets by creating the dataset, reflecting real-world circumstances more accurately. The work emphasizes the need for authentic datasets and their impact on enhancing the performance of anomaly detection systems.

Next, we propose to consider a list of classes of detected situations, where video surveillance and video analytics are the most important parts of recognition.

Detecting Dangerous Persons: traditionally, video surveillance was manually monitored, which relies heavily on human operators' vigilance and quick response. However, humans can easily get fatigued and miss important details, especially in mundane and repetitive tasks like surveillance. By applying advanced video analytics and machine learning, systems can be trained to recognize potential threats such as an aggressive person or someone carrying a weapon like a gun or a knife. These technologies can continuously monitor multiple video feeds, identify suspicious behavior, and promptly alert security personnel, increasing the efficiency and effectiveness of security measures.

Identifying Unattended Objects: unattended objects like bags, briefcases, or packages in public places could potentially indicate a security threat, such as a bomb. Again, depending on human operators to notice these objects amidst bustling environments can be tremendously challenging. Computer vision and machine learning algorithms can be designed to automatically detect such objects, even in crowded scenes or unusual places, triggering a timely response from relevant authorities.

Recognizing a Fallen Person: oftentimes, a person falling and becoming unconscious can go unnoticed, especially in less frequented or remote areas. Timely assistance in such cases can be lifesaving. Video analytic algorithms can analyze patterns of human motion and understand abnormal situations, such as a person falling and not moving. If a person stays immobile for a longer time than usual, the system can send an immediate alert to security personnel or medical teams, significantly reducing the response time and potentially saving lives.

In conclusion, the application of video analytics technologies, computer vision, and machine learning algorithms vastly enhances surveillance systems' capability to detect various dangerous situations promptly and accurately. These technologies not only lessen the burden on human operators but also assist in proactively mitigating risks and enhancing overall public safety.

Edge Video Analytics provides the benefit of processing these data points directly at the source (like the security camera), thus reducing latency, preserving bandwidth, and enabling real-time analysis and decision-making. This methodology is considered to be a transformative approach to public safety, law enforcement, and security management [13]. Thereby, the key scientific contribution of our study is an experimental extension of our previous works [1] show that such approach [14] can be extended to several video cameras that monitor the condition of athletes from several simulators at the same time using existing algorithms and technologies in human activity recognition [15] and work in real-life conditions.

III. Prototype for Recognition of Dangerous Situations

The creation of a video analytics system for dangerous situations is an imperative need in the contemporary landscape of public safety. This process begins with creating a systematic collection and processing of video images, which entails real-time acquisition of video feeds from myriad sources, such as CCTV cameras and surveillance systems located in diverse public spaces. This identification subsequently feeds into the classification of dangerous situations, where the system determines the level and nature of a potential threat. Sophisticated machine learning algorithms analyze the object interactions and movements, classifying them into pre-set categories based on their potential danger, such as gun handling, unusual object placement, or sudden falls.

In particular, as part of the implementation of the algorithm for detecting dangerous situations, the following requirements were fixed:

- Collection and processing of video images. Video images from cameras must be collected and stored for further processing. Images can be compressed and converted into a format that can be used for processing.
- Data preprocessing. To optimize the algorithm, it is necessary to pre-process the data. This may include resizing images, reducing noise, or adjusting the color gamut.
- Identification of objects. With the help of algorithms for selecting and tracking objects on a video image, it is necessary to track people, cars, and abandoned objects that can be dangerous.
- Classification of dangerous situations. Various dangerous situations can be classified using neural networks and machine learning. Models can be trained on labeled data that contains information about dangerous situations and can identify such situations in images in real-time.
- Alert. After detecting a dangerous situation, you can send an emergency alert to the computer or device of the security officer. Alerts can also be based on artificial intelligence, which can recommend actions to eliminate a dangerous situation.
- Data storage. An important part of the structure of the algorithm is the development of infrastructure for storing and accessing data. This may include storing images stored on the server, as well as generating reports and statistics, such as the number of items left per day/week/month.

The architecture of the system is shown in Fig. 1. It describes the design of the software and hardware part of the system, presented in the form of a microservice model and a system architecture for detecting dangerous situations. A particularly important part of this design is performed by a set of components, each of which performs its own functional role. One of the initial tasks of designing the system architecture of a video analytics system prototype is to describe its main components. Containerization technology is expected to be used to package, ship, and deploy the prototype. This approach will make it easier and more detailed to configure the launch of the full version of the system: isolating all module dependencies from the rest of the system, along with the module itself, will help deploy more modules without introducing conflicts between them; the ability to launch multiple instances of the same image, as well as automatic scaling, will help you competently increase system performance by increasing the number of working modules if you have the appropriate computing power; simple operation with a cluster of many physical machines will make it easier to increase capacity and horizontally scale the system. An example of an architecture with interaction between components can be a microservice architecture. In this architecture, the application is divided into small services, each of which performs a separate function. In particular, a model of a service-oriented system oriented to the periphery (Edge computing) and its event-driven control for local processing of video data can be implemented. An example of such a system is shown. Within the framework of this architecture, an example of connecting several video cameras to a local handler is given. Each of the dangerous situations can be defined on any of the cameras. In particular, the number of connected cameras within this architecture may not be limited. This approach uses different computational paradigms depending on how the computations are to be organized.

The next step is to create algorithms (detectors) that could identify the objects, a demanding step that employs computer vision and AI techniques to detect and distinguish various elements within the video frames from human figures to specific objects like guns or abandoned packages.

IV. Experiment with Recognition Algorithms

A Triad of detectors was engineered and developed to keep public places secure: the Stand-or-Fall detector, the Lost-LongLost items detector, and the Gun-Near-Human detector. The Stand-or-Fall detector is designed to promptly identify unusual movement patterns that may indicate a person in distress or the occurrence of an accident. The Lost-LongLost items detector focuses on identifying and alerting authorities about unattended objects, aiding in recognizing potential threats of harmful or lost items swiftly. The Gun-Near-Human detector lends an invaluable layer of safety by autonomously identifying situations where a firearm is in close proximity to a human, a scenario that could potentially escalate into dangerous circumstances. This integrated system heralds a significant advancement toward proactive and efficient public safety solutions. Each of these detectors, while potent independently, function seamlessly within an integrated system, collectively providing eyes on the ground that never blink, comprehensively augmenting public safety.

The algorithms of how the system works and its details were also described. After the frame is retrieved from the camera by the camera connection module, it is sent to the Neural Networks module for recognition. At first Neural Networks module gets executed. Two networks are used there: of the shelf-pretrained neural network for human recognition (YoLov8m), that detects human bounding boxes and key points; and custom-trained version of YoLov8M, trained on threat objects, to detect such items as guns or list cargo (luggage, bags).

The result for this module is transferred to detectors, depending on which data detectors need.

System services are separated into corresponding modules as shown in Fig. 2:

- Camera connection module;
- Neural Network Detectors module;

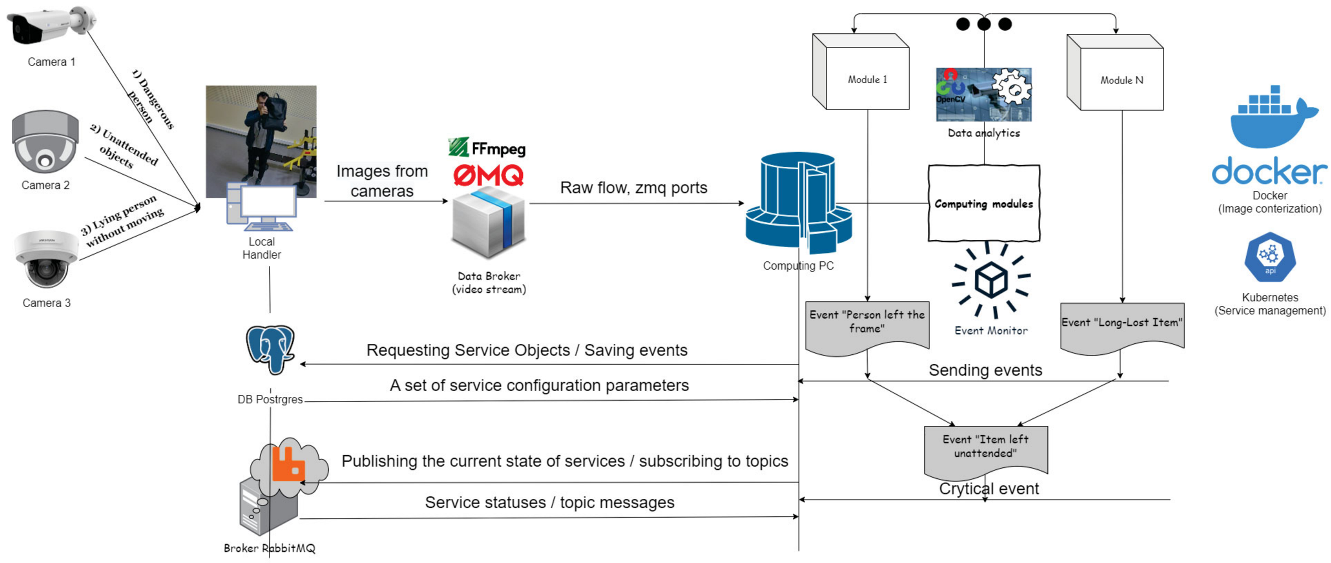

Fig. 1. Architecture of the Video Analytics System for Dangerous Situations

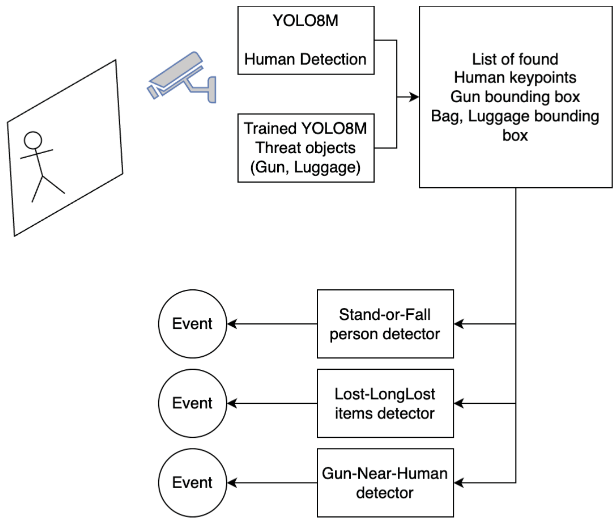

Fig. 2. Data flow from camera through Neural Network modules into detectors modules that produces events for certain situation.

Fig. 3. Concept visualisation that standing person (A) is more tall and less wide that person who fall (B)

- Stand-or-Fall detector;
- Lost-LongLost items detector;
- Gun-Near-Human detector;

A. *Stand-or-Fall detector*

Stand-or-fall detector dedicated to determining the state of a person in the frame, given key points.

The method takes far left, right, top, and down key points and checks distance. If the distance between vertical points is larger than the horizontal then the person stands otherwise falls (Fig. 3).

Points were chosen as a promising way of detecting a person's state, potentially allowing one to move to a better method in the future. Current implementation assumes that a standing person would be doing it in the vertical plane and a person that falls would fall in the horizontal plane, which is not guaranteed to be true. Person orientation thus depends on camera placement.

Algorithm:

1) Calculate the distance between the far-most left and right points;
2) Calculate the distance between the far-most top and bottom points;
3) If the top-bottom distance is higher than left-right then the human is standing and falls otherwise;

B. *Lost-LongLost items detector*

Lost-LongLost items detector dedicated to detecting items that were lost or no longer have their original owner presented nearby for some time, given bounding boxes of found objects. An example is shown in Fig. 4.

The method takes found luggage and tracks how long it was presented. If the luggage is presented but there are no humans around it count the item as Lost. When time exceeds the

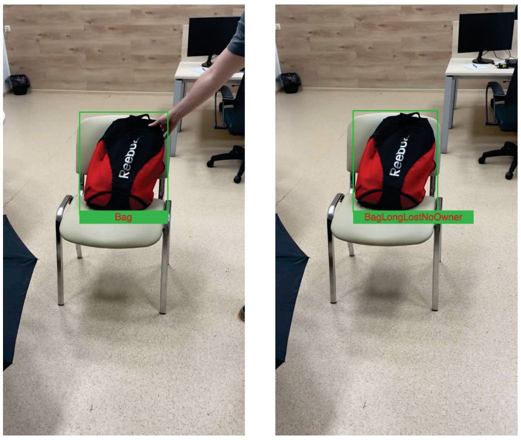

Fig. 4. Example of how bag can be detected in normal state and after some time transferred to lost when there is no owner nearby

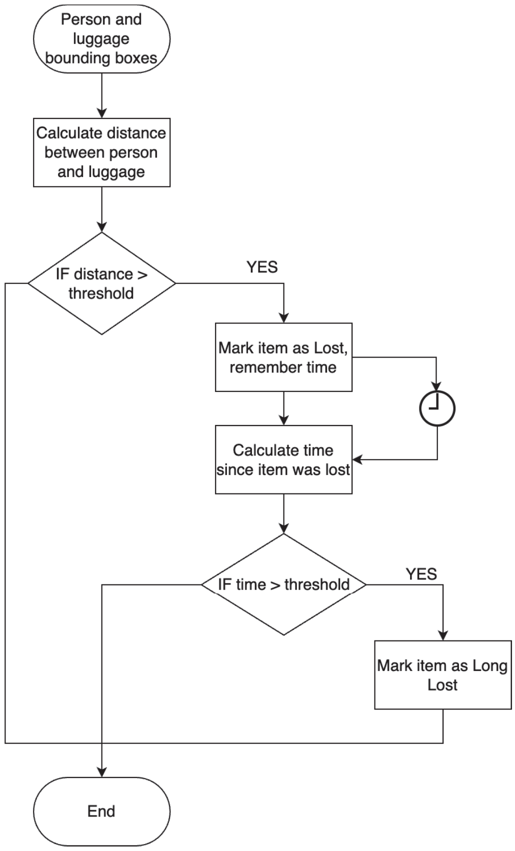

Fig. 5. Diagram of how module works.

threshold value it marks the item as LongLost. Visualization is presented in Fig. 5.

The algorithm is presented below:

1) Calculate distance between luggage and person;
2) If distance less than the threshold, luggage near a person, back to step 1;
3) Otherwise mark luggage as lost;
4) Start tracking time from the moment luggage becomes lost;

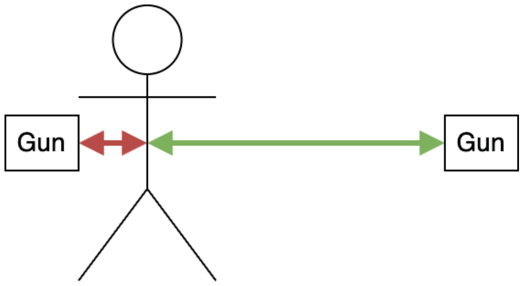

Fig. 6. Concept visualization of the relative distance between human and gun.

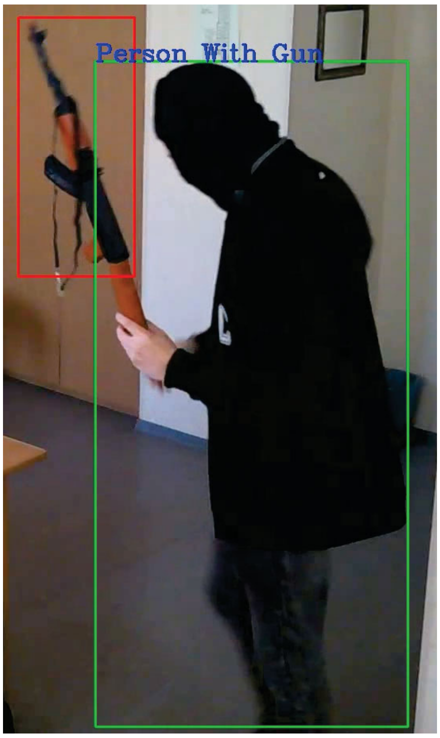

Fig. 7. Example of a Highlighted Person with a Weapon

5) If the time exceeds the threshold, mark luggage as a long list;

C. Gun-Near-Human detector

Gun-near-human detector dedicated to finding dangerous peoples (as in peoples that holds gun), given bounding boxes of people and guns.

If a gun is found near a human it will be marked as a dangerous human (Fig. 6).

The algorithm is presented below:

1) Calculate distance between gun and person;
2) If distance less than the threshold, mark the person as dangerous (a highlighted person with a weapon (toy gun was used as an example) is shown in Fig. 7);

D. Results

These detectors were analyzed and tested based on our experimental stand with the following characteristics:

- ”Security CCTV-camera”: IP-camera Hikvision (2.8 mm) installed at 3m in a corridor;
- Camera mounted on a ”security guard” to simulate real activity in the room: Web camera Logitech installed at 1.5-2m (human height) to capture live-action videos;
- Abnormal Human Activity Recognition occurs using the YoloV8 neural network;
- Each camera (IP, Web) had its own set of images that were used in training datasets;
- A conclusion is made about the current human’s position (stand/fall or with a gun) or objects that were left unattended.

The server for data processing was based on the following specifications:

- Hardware: Intel Core i9-9900K 3.6GHz, Nvidia RTX 2080 8GB, 32 GB RAM, 1TB SSD.
- Detectors were based on the existing technologies: Flask, PyTorch, YoloV8, OpenCV, FFmpeg, ZeroMQ, and GPU-based recognition.

The results showed that the developed detectors and algorithms make it possible, using various types of video cameras in a mode close to real-time, to detect various dangerous situations: abnormal human activity or objects left by a person. The development was based on publicly available datasets and our own datasets, which were filmed indoors as part of experiments. However, in cases where the person occupied various positions that were poorly detected by the camera, the accuracy decreased sharply. These detectors can be improved through initial image processing (for example, removing noise or preventing the camera itself from shaking), as well as by collecting new datasets, which will allow achieving the greatest accuracy in recognition.

V. Conclusion

This study experimented with recognition algorithms for dangerous situations in public spaces. The three types of dangerous situations were considered: a dangerous person, objects left unattended, and a person who has fallen without movement. The recognition is event-based and operating in real-time. The proposed VSS architecture includes several layers for extracting features from images and classifying dangerous situations. The proposed recognition algorithms are based on deep learning using convolutional neural networks. Our experiments show that the proposed recognition algorithms are able to run on relatively low performance devices and cameras. As a result, the existing digital infrastructure of public spaces can be utilized, with no introduction of many advanced equipment elements.

Acknowledgment

This R%D study was made possible in part by R&D Support Program for undergraduate and graduate students and postdoctoral researcher of PetrSU, funded by the Government of the Republic of Karelia. The experimental work is in collaboration with Artificial Intelligence Center of PetrSU.

References

[1] N. Bazhenov, E. Rybin, and D. Korzun, “An event-driven approach to the recognition problem in video surveillance system development,” in *2022 32nd Conference of Open Innovations Association (FRUCT)*, 2022, pp. 65–74.

[2] S. Priya and R. Minu, “Abnormal activity detection techniques in intelligent video surveillance: A survey,” in *2023 7th International Conference on Trends in Electronics and Informatics (ICOEI)*, 2023, pp. 1608–1613.

[3] R. Vijeikis, V. Raudonis, and G. Dervinis, “Efficient violence detection in surveillance,” *Sensors*, vol. 22, p. 2216, 03 2022.

[4] T. J. Nandhini and K. Thinakaran, “An improved crime scene detection system based on convolutional neural networks and video surveillance,” in *2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, 2023, pp. 1–6.

[5] S. G. Shivaprasad Yadav, S. Itagi, B. V. N. V. Krishna Suresh, H. K.L, and R. A C, “Human illegal activity recognition based on deep learning techniques,” in *2023 IEEE International Conference on Integrated Circuits and Communication Systems (ICICACS)*, 2023, pp. 01–07.

[6] N. Ottakath and S. A. Al-Maadeed, “Reliable video forensics evidence cataloguing using video source device identification on the blockchain,” in *2022 International Conference on Emerging Trends in Smart Technologies (ICETST)*, 2022, pp. 1–6.

[7] H. E. A. El Abdallaoui, A. El Fazziki, F. Z. Ennaji, and M. Sadgal, “A crowdsourcing based framework for e-government: Suspect identification and investigation,” in *2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, 2017, pp. 380–387.

[8] J. Xiao, S. Li, and Q. Xu, “Video-based evidence analysis and extraction in digital forensic investigation,” *IEEE Access*, vol. 7, pp. 55 432–55 442, 2019.

[9] M. N. Asghar, N. Kanwal, B. Lee, M. Fleury, M. Herbst, and Y. Qiao, “Visual surveillance within the eu general data protection regulation: A technology perspective,” *IEEE Access*, vol. 7, pp. 111 709–111 726, 2019.

[10] H.-T. Duong, V.-T. Le, and V. T. Hoang, “Deep learning-based anomaly detection in video surveillance: A survey,” *Sensors*, vol. 23, no. 11, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/11/5024

[11] B. Omarov, S. Narynov, Z. Zhumanov, A. Gumar, and M. Khassanova, “State-of-the-art violence detection techniques in video surveillance security systems: a systematic review,” *PeerJ. Computer science*, vol. 8, p. e920, 2022. [Online]. Available: https://europepmc.org/articles/PMC9044356

[12] X. Zhang, J. Fang, B. Yang, S. Chen, and B. Li, “Hybrid attention and motion constraint for anomaly detection in crowded scenes,” *IEEE Trans. Cir. and Sys. for Video Technol.*, vol. 33, no. 5, p. 2259–2274, nov 2022. [Online]. Available: https://doi.org/10.1109/TCSVT.2022.3221622

[13] Q. Zhang, H. Sun, X. Wu, and H. Zhong, “Edge video analytics for public safety: A review,” *Proceedings of the IEEE*, vol. 107, no. 8, pp. 1675–1696, 2019.

[14] N. Bazhenov, E. Rybin, S. Zavyalov, and D. Korzun, “Human activity recognition for sport training machines,” in *2022 32nd Conference of Open Innovations Association (FRUCT)*, 2022, pp. 328–331.

[15] J. A. Patiño-Saucedo, P. P. Ariza-Colpas, S. Butt-Aziz, M. A. Piñeres-Melo, J. L. López-Ruiz, R. C. Morales-Ortega, and E. De-la-hoz Franco, “Predictive model for human activity recognition based on machine learning and feature selection techniques,” *International Journal of Environmental Research and Public Health*, vol. 19, no. 19, 2022. [Online]. Available: https://www.mdpi.com/1660-4601/19/19/12272