Video Monitoring of Personnel in Manufacturing Equipment Activity

Vsevolod Averkov, Kirill Kulakov
Petrozavodsk State University (PetrSU)
Petrozavodsk, Russia
{averkov, kulakov}@cs.petrsu.ru

***Abstract*—The use of robotic and computerized machines in the enterprise requires constant monitoring of personnel actions. Working with machines requires qualifications and safety precautions. The use of technical means of control (covers, locks, etc.) allows you to solve the problem of access, however, it is not always possible to install technical means of control. The paper presents an approach to the use of surveillance cameras for automatic control of personnel actions. The software system determines the presence of an employee in the camera visibility zone, the employees transition from one zone to another, the employees proximity to critical equipment nodes (open mechanisms, cabinets with electrics, etc.). Using a software system allows you to automate surveillance in conditions of increased pollution.**

I. Introduction

Modern production uses high-tech equipment. As a rule, a production workshop contains several machines with auxiliary equipment. A large number of personnel are involved in the production, for example, design engineers, loader operators, machine operators, repairmen, and quality control specialists. Each employee has a specific qualification, responsibilities and level of access. In this regard, for production, the main and main task comes to the fore-ensuring the safety and control of automated technological processes; preventing and eliminating possible problems or preventing incorrect actions of employees that entail damage; monitoring compliance with safety rules. Thus, the problem arises of monitoring the actions of personnel in the enterprise. This task is also related to the issue of compliance with safety regulations.

There are two ways of observation: mechanical (contact) and visual (non-contact). Mechanical surveillance involves the installation of fences, frames, locks, covers and other devices that prevent unauthorized entry. For example, electrical equipment is placed in special cabinets with doors and a lock. Access to such cabinets is provided, for example, to electricians, but is closed, for example, to machine operators. As one of the methods of mechanical observation, radio frequency identification (RFID) is used to track the movement of physical objects [1]. However, this method requires constant availability of RFID tags for staff.

At the present stage of society's development, video surveillance at production facilities, where control and security are of fundamental importance, is becoming a necessity [2]. Installation of video surveillance systems is designed to solve the following tasks for the management of enterprises: to save time for performing key tasks; to control deadlines; to prepare operational planned reports of enterprise divisions; to manage the unloading and shipment of materials and products; to effectively use the working space, which makes the work of production personnel transparent. Video surveillance systems in the enterprise allow you to automatically solve many other, no less important tasks. This is a simple motion detection in surveillance; accurate count passed into the territory of vehicles and the number of people who passed through the checkpoint of the enterprise; the movement of goods; control of access areas; the efficiency of the staff analyzing.

Today, the use of video surveillance systems in hard-to-reach places (objects located outside the head office) is able to remotely control complex automated technological processes and, if necessary, promptly take appropriate measures, preventing a serious failure in the technological process. This remote real-time monitoring allows you to see what employees are currently doing, what technological operations are being performed, and automatically signals violations, for example, of the work schedule. This can be delays, delays, early departures, absenteeism, walks, and other distractions of the staff from the work process. They are automatically reflected in the reports and the time sheet.

Intelligent elements built into the video surveillance system allow you not only to observe events and collect them in the archive, but also to notify the operator of various incidents, to give a signal about the presence of strangers in a certain production area. Accordingly, the use of video detectors will significantly simplify the control of the technological process at the enterprise's facilities and will contribute to the effective operation of the entire production as a whole.

The paper considers the task of monitoring the movement of personnel and determining the contact of a person with infrastructure using computer vision. Monitoring is carried out using multiple cameras. To determine the contact, an estimate of the distance between the person and the object is used.

The task of machine vision is well known [3]. There are software platforms, methods, and algorithms for identifying visual objects. For example, the OpenCV library [4] presents an algorithm for identifying the presence of a person [5]. At the same time, a number of tasks have no universal solutions due to the peculiarities of machine vision and the conditions for using machine vision in practice. The paper presents a solution to the problem of personnel monitoring based on convolutional neural networks (CNN) and the OpenCV library.

II. Related work

The integrated security system "SecurOS Enterprise" [6] is another solution for organizing highly loaded video management systems at geographically separated facilities. The system supports an unlimited number of servers and cameras on the network, has no restrictions on the number of video analytics channels. The solution, built on the basis of "SecurOS Enterprise", provides the functionality of a professional high-end video management system.

The editors of SecurOS Enterprise are designed to ensure the safety of transport, municipal and departmental infrastructure facilities, as well as the security of large enterprises or groups of industrial complex facilities. Today, "SecurOS Enterprise" is widely used for the implementation of video surveillance systems and video analytics in the decisions of the state program "Safe City" [7].

Since 2016, the introduction of the face recognition system "Smart Tracker FRS" [8] has begun at airports in the Russian Federation. Thus, the author of the system is the Russian company "Center for Speech Technologies", is testing its development of biometric video identification at the airport of Yuzhno-Sakhalinsk. The system helps to identify criminals, terrorists, extremists by the image from the video online. The project is also being implemented as part of the state program "Safe City".

The essence of the system is as follows: video cameras are installed around the entire perimeter of the airports transport security zone, at the entrances and exits. The "Vizier" automatically recognizes and identifies incoming faces. The system is integrated with a comprehensive automatic search system, which contains a database of wanted persons. If a person is found among the wanted persons, the message is sent through a special closed channel to police workstations and mobile devices. Such a message contains a frame from a video camera, as well as information about the time and place of detection of the offender.

III. Description of Approach

A. Person detection algorithm

Determining the presence of a person in the frame is a well-known task [3]. As a rule, they consider a narrower problem associated with determining the face of a person [9]. In our case, it is necessary to determine the frames at which a person is present taking into account the characteristics of production: the use of work clothing, the presence of items carried by a person (for example, tools), poor lighting, high pollution and long distances. This leads us to the need to adapt the algorithms used.

In work [9] authors presents algorithms for detecting humans based on neural networks. We use the convolutional neural network (CNN) VGG16 [10], [11]. The idea of the algorithm is the sequential use of classifier layers to filter the image, highlight features and reduce data size. As a result, the last layer determines the presence of a person based on the selected features.

To train the neural network, a set of photos of a person taken from different sides and a set of photos with the absence of a person in the frame are used. During training, the neural network tries to highlight some features of the image in order to be able to compare the incoming image when using it and connect it to any class. For example, based on the original image (Fig. 1), training the convolutional network VGG16 produces the following intermediate results (Fig. 2). The result of applying filters in the first convolutional layer is a multitude of image versions highlighted by various elements: some lines focus on the background or foreground.

Fig. 1. Original image for training

Fig. 2. Result of applying the filters in the VGG16. We can see a lot of versions of same image with different features in first convolutional layer

In Fig. 3 we can see that closer to the input of the model, many small details are fixed in the image and that as we go deeper into the model, less and less details are visible. this is because the model abstracts image elements into more general concepts that can be used for classification. Although it is not clear from the final image that this is actually a person. But it is these characteristics from the part that help to determine in the photo of a person or object.

More promising is the neural network MobileNet SSD COCO [12]. The network uses the VGG16 algorithm and the Single Shot Detector algorithm as a network to extract features. MobileNet SSD neural network training is based on a variety of images with support for XML annotations (LabelImg

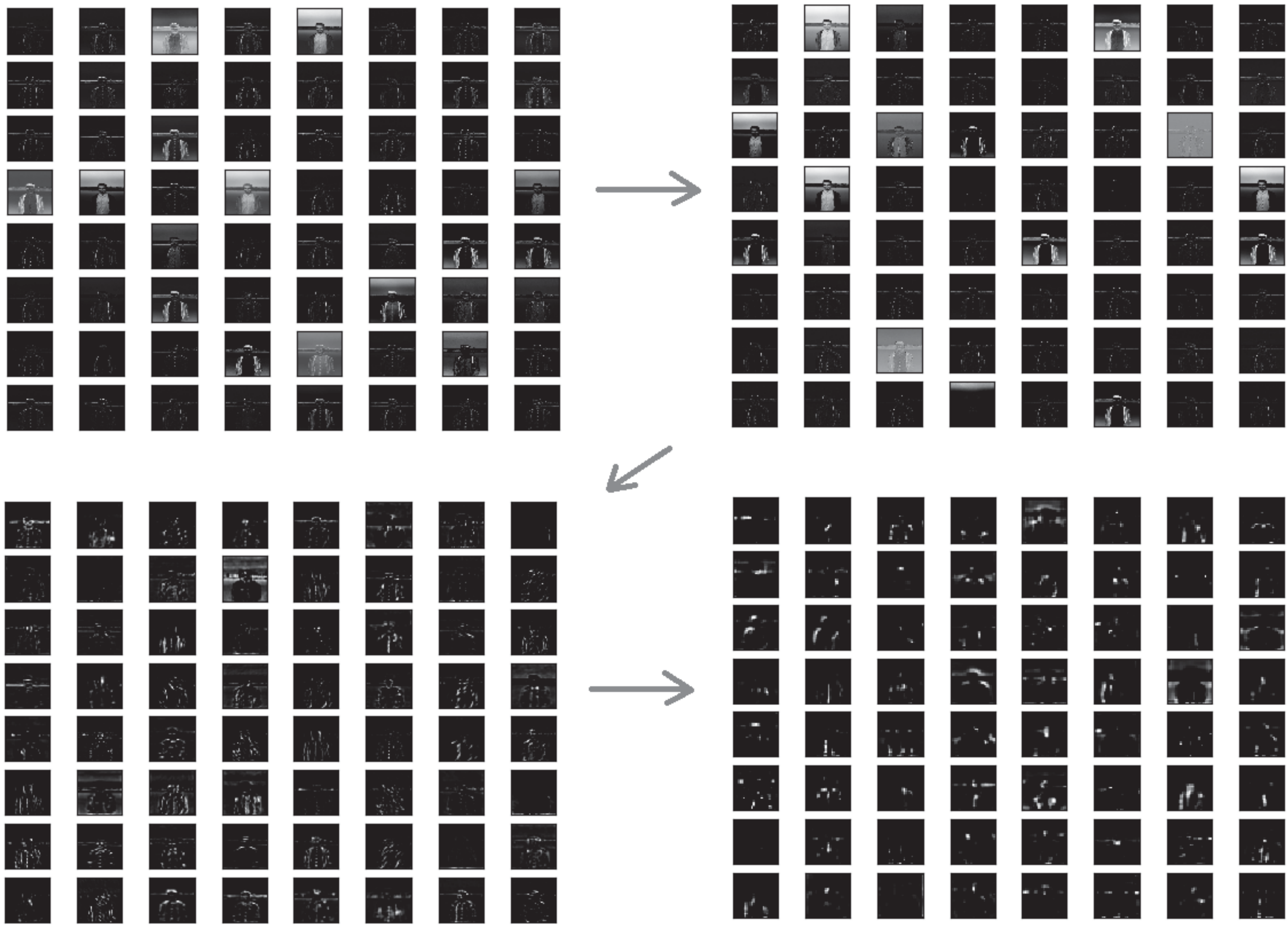

Fig. 3. The process of finding features

program [13]). A key feature of the MobileNet SSD is higher recognition accuracy and faster results with less weight.

B. Object detection algorithm

Unlike a person, where we know his proportions, statistical sizes and posture options, the determination of the equipment nodes required for observation cannot be resolved in the general case. To determine the presence of objects on the screen, we use tracking algorithm [14]. The operation of the pattern detection algorithm is as follows: at the setup (training) stage, the desired object is highlighted on the image and its image is stored. In the process, the algorithm tries to find a similar image of the same object in the same place. If the image matches the pattern, then the object is found.

In the process of work, the following events can occur: camera displacement, moving an object, temporarily closing an object with another object or person, and changing the object. In this case, the pattern detection algorithm will not find the object. To solve these problems, a movement tracking algorithm is used. The algorithm attempts to track small changes to an object in adjacent images. If the changes are small, then the template is updated. As a result, a combination of algorithms allows you to determine the location of a given object for a long time.

C. Measuring the Distance from Person to Object Algorithm

The task of determining the distance between a person and an object can be divided into three subtasks: determining the location of a person, determining the location of an object and determining the desired distance between the obtained locations. To solve the last subtask, it is enough for us to know the distances between the person / object and the camera. (see the diagram in Fig. 4). Thus, it is necessary to be able to determine the distance between the camera and the object.

To determine the distance from our camera to a known object, we need to use the likeness of a triangle (see Fig. 5). The similarity of the triangle looks something like this: there is an object with a known width W. Then we place this object at a certain distance S from our camera. We select the object using our camera and then measure the apparent width in pixels Pw. This allows us to get the focal length of the cameras Focus: $Focus = (Pw * S)/W$. Continuing to move the camera closer and further from the object, again we can use the similarity of triangles to determine the distance of the object to the camera: $Sm = (W * Focus)/Pw$.

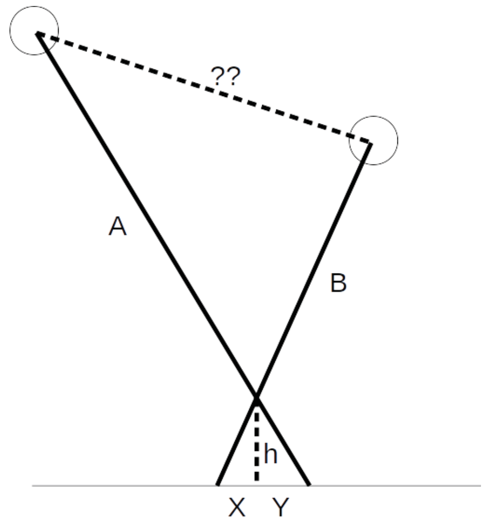

Fig. 4. Measuring the Distance from Person to Object (A, B — distances between the person / object and the camera; X, Y — offset from the center of the image; h — focal length)

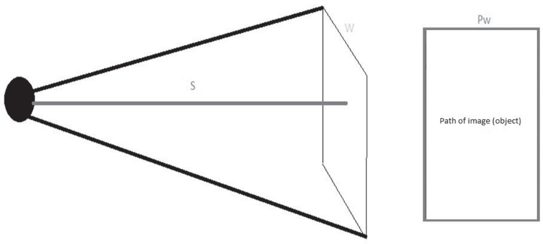

Fig. 5. Scheme for calculating the distance to the object

Knowing the distance from the camera to the object and its location in the picture, you can get its three-dimensional coordinates relative to the camera. Thus, to detect the distance from a person to an object, it is necessary to know the location of the person and the object in the image and their sizes.

An object detection algorithm is used to determine the location of an object. The size of the object can be measured manually, as objects are individual in their characteristics.

To determine the position of a person in the frame, the Haar cascade classifier [15] is used. An algorithm based on a list of primitives (for example, a persons face or physique) determines the appropriate combination of pixels in the image. Because Haar cascade classifier is resource-intensive, then before using it you need to know that the person is in the frame. And the determination of the presence of a person in the frame is carried out using the Person detection algorithm. As the size of a person, average statistics are used: head width 15 cm, person height 170 cm. For a more accurate determination of the distance, the use of two cameras and the organization of ”stereoscopic vision” are required.

IV. Implementation and approbation

A. Implementation

The implementation of the personnel control service is carried out as part of the Opti-Repair software system in the form of a set of modules (see Fig. 6). The “Person Detection” module analyzes the video stream from the camera and determines the fragments where a person is present. The “Contact Detection” module analyzes the fragments found and calculates the distance between the person and the tracking objects. Information about events (“a person has appeared”, “a person has disappeared”, “a person is close to the object”) is placed in the event database. Video fragments and confirmation screenshots are placed in the file system. View the current status and past events through the web interface.

To implement and train the neural network, several software tools were analyzed: TensorFlow object detection API [16], [17], Keras [18], Caffe [19], PyTorch [20]. The analysis showed the possibility of using all of these tools. However, tensorflow is very common when used in practice. As a result, Tensorflow was selected. To select the desired object, the LabelImg [13] program was used. Before training, the entire dataset will be processed with OpenCV filters [14] in order to highlight the characteristic properties of the photo or video material that will be given to the neural network.

Setting up the “Contact Detection” module includes the definition of monitored objects. To do this, a screenshot is taken and the object (s) are selected using rectangles (see Fig. 8). The selected fragments are remembered in the module settings and are updated while tracking the movement of the object. To determine the presence of an object on the screen and for tracking movement OpenCV Tracker [21] (see Fig. 7) is used.

To determine the contact between a person, the module determines the distance from the camera to the object, the distance from the camera to the person (see Fig. 9), the location of the person (Haar cascade). The location of the person and the object, their sizes and calculated and distances are used to determine the spatial coordinates relative to the camera and, ultimately, the distance between the object and the person.

Based on the results of the ”Person Detection” and ”Contact Detection” modules, final reports are generated on the presence of a person in the camera visibility zone and on the persons contact with the monitored objects. An example report is shown in Fig. 10.

B. Approbation

The personnel monitoring system was tested using an experimental stand (see Fig. 11). The stand consisted of a computer, a switch and two cameras Hikvision DS-2CD2123G0-IS. Computer characteristics are presented in table I, camera characteristics are presented in the table II.

TABLE I. Characteristics of the computer stand

Characteristic	Value
Processor	Intel Core i5-9400F 2.9 / 4.1 GHz, 6C / 6T, 9 Mb L3, DDR4-2666
RAM	DIMM DDR4 32768 Mb 2666 Kingston HyperX Fury
Hard disk	2 x SSD 2.5 ”SATA-3 1 Tb WD Blue 3D NAND
Motherboard	GIGABYTE LGA1151-v2 Z370 Z370P D3
Video card	PCI-E AsusGeForce GT 710 Silent LP 1024 Mb 64-bit GDDR3
Operating system	Ubuntu 18.04.2 LTS 64-bit

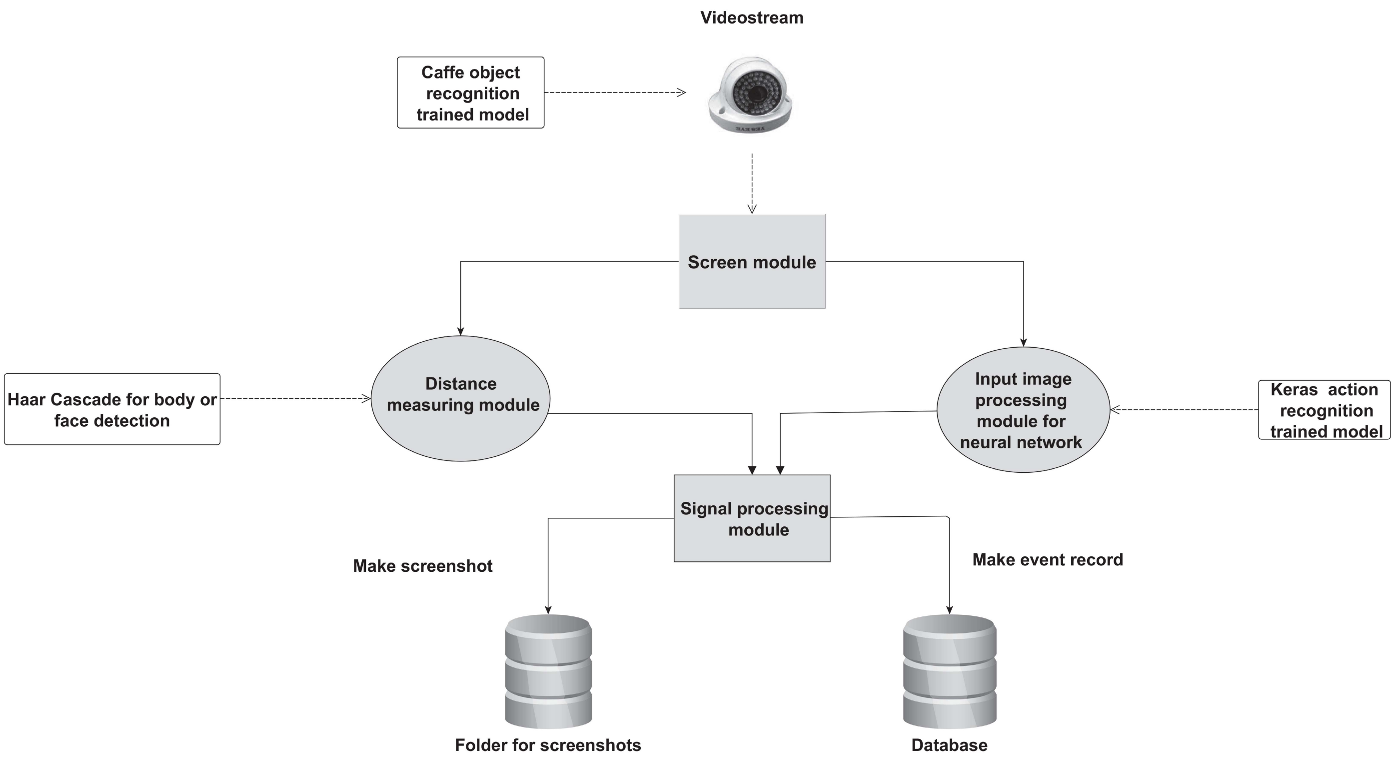

Fig. 6. General architecture of the module for recognizing a person and measuring the distance to an object combined with human recognition

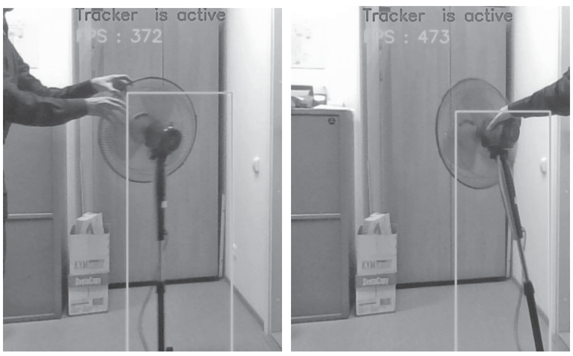

Fig. 7. OpenCV Tracker in action

TABLE II. CHARACTERISTICS OF THE HIKVISION DS-2CD2123G0-IS (4MM) IP CAMERA

Characteristic	Value
Matrix type	CMOS Progressive Scan
The number of megapixels	2Mp
Minimum Illumination	0.028 Lux
Focal length	4 mm
Viewing angle	102 (diag.), 46 (vert.), 86 (horizontal)
Maximum resolution	1920 x 1080
Video recording formats	MJPEG, H.264 +, H.265, H.265 +, H.264

As part of testing, the systems operation was tested on various options: there is no person in the cameras coverage area, the person is far from the object, the person is close to the object, the person is closing the object from the camera. As a result of the experiments, the following parameters were evaluated:

- accuracy of determining a person and an object;

Fig. 8. Object selection for observation

- accuracy of determining the distance from the camera to the person and to the object;
- accuracy of determining the distance between a person and an object.

An example of the operation of the system during the experiments is presented in Fig. 12.

V. CONCLUSION

Year after year, video analytic turns into a phenomenon that each of us faces daily, penetrating into all areas of

Fig. 9. Measuring the distance between a person and camera

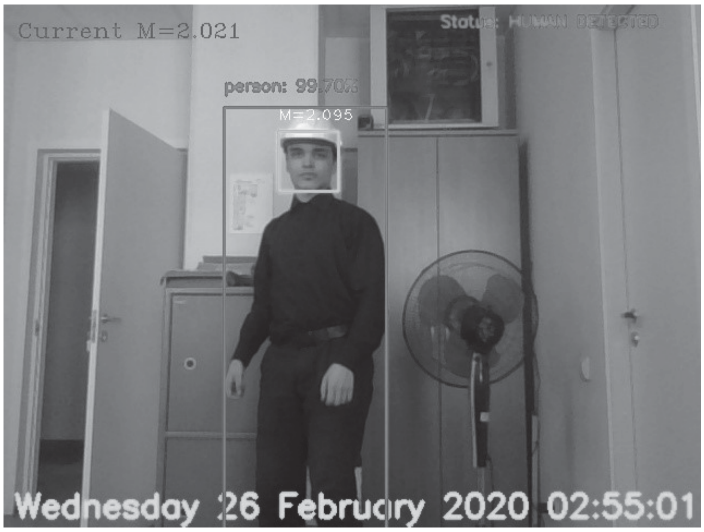

Fig. 12. Full service screenshot

Fig. 10. Person presence report

Fig. 11. It-park stand

human civilization. We see its use not only in scanners of 2D codes and fingerprints, but also in the counters of visitors to outlets; devices for tracking the speed and fatigue of drivers of vehicles; recognition of car license plates, the authenticity of banknotes. According to scientists, despite the successes in the development of artificial intelligence, at the moment, society is still at the very beginning of progress in the development and use of intelligent systems in video analytics.

Acknowledgment

This research is financially supported by the Ministry of Science and Higher Education of Russia within project no. 075-11-2019-088 (Creating the high-tech production of mobile microprocessor computing modules based on SiP and PoP technology for smart data collection, mining, and interaction with surrounding sources).

References

[1] B. Mayorov, “Detection of personnel locations with rfid-system in premises of special objects,” *Military Enginery. Scientific and Technical Journal. Counter-terrorism technical devices. Issue 16.*, no. 5-6, pp. 28–32, 2013.

[2] M. Sonka, V. Hlavac, and R. Boyle, *Image processing, analysis, and machine vision*. Cengage Learning, 2014.

[3] D. T. Nguyen, W. Li, and P. O. Ogunbona, “Human detection from images and videos: A survey,” *Pattern Recognition*, vol. 51, pp. 148–175, 2016.

[4] “Opencv library,” Feb. 2019. [Online]. Available: https://opencv.org/

[5] K. Goyal, K. Agarwal, and R. Kumar, “Face detection and tracking: Using opencv,” in *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*, vol. 1. IEEE, 2017, pp. 474–478.

[6] “Iss securos vms — video monitoring, management, and data processing in real-time and archive.” [Online]. Available: https://en.iss.ru/products/securos-vms

[7] “Iss solutions for safe city.” [Online]. Available: https://en.iss.ru/verticals/safe-city

[8] “Smart tracker frs: the best tool for facial recognition in high-traffic places — speech technology center.” [Online]. Available: https://speechpro.com/product/facial-recognition/smart-tracker-frs

[9] X. Sun, P. Wu, and S. C. Hoi, “Face detection using deep learning: An improved faster rcnn approach,” *Neurocomputing*, vol. 299, pp. 42–50, 2018.

[10] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[11] D. Siegmund, A. Prajapati, F. Kirchbuchner, and A. Kuijper, *An Integrated Deep Neural Network for Defect Detection in Dynamic Textile Textures: 6th International Workshop, IWAIPR 2018, Havana, Cuba, September 2426, 2018, Proceedings*, 09 2018, pp. 77–84.

[12] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[13] "Labelimg is a graphical image annotation tool and label object bounding boxes in images." [Online]. Available: https://github.com/tzutalin/labelImg

[14] K. Pulli, A. Baksheev, K. Kornyakov, and V. Eruhimov, "Real-time computer vision with opencv," *Communications of the ACM*, vol. 55, no. 6, pp. 61–69, 2012.

[15] S. Soo, "Object detection using haar-cascade classifier," *Institute of Computer Science, University of Tartu*, pp. 1–12, 2014.

[16] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.

[17] "Tensorflow — an end-to-end open source machine learning platform." [Online]. Available: https://www.tensorflow.org

[18] "Keras: The python deep learning library." [Online]. Available: https://keras.io

[19] "Caffe — deep learning framework," Feb. 2019. [Online]. Available: http://caffe.berkeleyvision.org/

[20] "Pytorch." [Online]. Available: https://pytorch.org

[21] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *2010 IEEE computer society conference on computer vision and pattern recognition*. IEEE, 2010, pp. 2544–2550.