Faces 2D-Recognition and Identification Using the HOG Descriptors Method

Vyacheslav Voronov, Vladimir Strelnikov, Liliya Voronova, Artyom Trunov, Andrey Vovik
Moscow Technical University of Communications and Informatics,
Moscow, Russia
Vorvi@mail.ru, strel-prod@ya.ru, voronova.lilia@ya.ru,greek17@ya.ru , andreyvovik@gmail.com

***Abstract*—The article describes the development of a face recognition system for university security services. Recent facial recognition solutions were analyzed. The software and hardware architecture, the developed system, the methods used in the recognition program are described. Examples of the implementation of two-dimensional recognition of faces with HOG descriptors are given.**

I. INTRODUCTION

There are many intelligent security systems using biometric people identifiers. The identification is possible with using static and dynamic characteristics such as: papillary finger pattern, hand geometry, iris of the eye, face geometry (2D and 3D), vein pattern, handwriting, silhouette, gait, voice etc. [1].

All presented identification methods have their own set of advantages and disadvantages related to the method of obtaining human metrics (needed /not needed physical contact with a person), cost, reliability of results, elaboration of the methodology.

Recently, the number of video cameras in different public places has greatly increased, at the same time the quality of the data obtained from a standard modern surveillance camera extremely improved and converted to digital format. Besides, the average cost of receiving, processing and transmitting the data has seriously decreased. It has led to the creating and development of methods for the effective identification of people from video and photo images, and the most part of methods focusing on the recognition of human faces.

The largest companies in the world are developing their own software solutions that use the functionality of face recognition in photo and in a video stream. Giants such as Amazon (USA), Facebook (USA), Apple (USA), JD.com (Cina), and also startups NTechLab (Russia), Macroscope (Russia), Yitu (China), Megvii (China) implemented intelligent analysis of the video stream in real-life conditions and offer the market ready-made competing solutions.

Amazon Rekognition [2] from Amazon allows to embed image and video analytics based on deep learning into your applications. This service can recognize objects, people, text, scenes and actions, as well as detect inappropriate content. In addition, Amazon Rekognition accurately analyzes and recognizes faces in images and client videos. The service allows to use private image repositories.

Since 2014, Facebook has been implementing a project to recognize and pre-identify people's faces in photographs placed on a social network, and since 2017, it has released a special function [3] that allows to use words to describe everything shown in the photo. For external commercial companies this technology is not available, but it clearly demonstrates the priorities of social network users and the direction of Facebook evolution.

Apple's FaceID technology [4] allows for highly accurate identification of a smartphone user and suggests a gradual replacement of a fingerprint scanner, since it uses a system of high-quality sensors and cameras and a whole range of identification methods, including face recognition. The technology is not available outside the Apple ecosystem, but considering its size, it makes up a considerable share of the identification market.

JD.com is the second (after Alibaba) largest Internet retailer in China. Tough competition conditions force the company to apply innovative techniques when working with its customers and partners. Computer vision and machine learning made it possible to implement a project of a trade kiosk [5], identifying a client through a video terminal, to think about a fully automated shopping complex [6] without cash registers and controllers, and even organize tracking of the life cycle of pigs [7].

In the Russian market, several innovative companies form the sector of services for intellectual image analysis in access control systems (ACS).

For example, the Macroscop service [8] from the eponymous company offers complete (for sale full-fledged software and hardware complex) typical and expandable solutions for video surveillance and access control within organizations. The systems recognize faces and machine numbers.

NTechLab [9] is the leader in the algorithms development and software solutions to control the field of events and personal identification for large groups of people, cities and countries. The FindFace service from this company occupies the top lines of competitive ratings for face recognition algorithms worldwide. Moreover, the service was successfully used in Moscow during the Football World Cup 2018 and showed excellent results. On the basis of its automated recommendations, the bodies of the Ministry of Internal Affairs detained more than 100 criminals.

Mathematical methods and models used in existing facial recognition systems include prediction using regression, time series prediction, Gray prediction, Markov prediction, an artificial neural network, and many others.

All the manufacturers listed above use their own set of techniques and algorithms, but using of neural networks has become an industry standard, since the corresponding methods give the fastest and most accurate results.

The accuracy degree of identification and the speed of analysis depend on the methods of the primary image transformation, the characteristics of the training set, used deep training methods and the database search algorithm. Learning sets from world leaders contain tens and hundreds of millions of images, and various methods of in-depth training of the network give a serious (sometimes many times) difference in the accuracy of identification.

Identification of a person in a video stream is a process that has several stages. The first step is to convert the video stream into a set of prepared pictures, the second step is to search for the fragments with faces, the third one is to assign a unique set of characteristics to all those found, the fourth one is to identify the person based on a comparison with the database of the available sets.

The MTUCI department «Intellectual Systems in Control and Automation» conducts scientific work on the development and application of data mining methods for predicting situations in different subject areas, including solving the problem of image recognition [10], [11], [12]. Mathematical methods and models used for face recognition include regression method, neural networks [13], [14]., SVM and others

In this paper, the authors write about 2D facial recognition system based on *HOG* descriptors.

II. Biometric identification systems

Biometric identification systems are used to identify the user based on his biometric data [15]. Currently, there are eight of the most developed biometric identification parameters: **face**, fingerprints, hand geometry, iris, retina, thermogram, signature, voice.

At the registration stage, the biometric system records the user's biometric sample using a sensor — for example, it captures the face image with the camera. Individual features are extracted from the biometric sample using the feature extractor software algorithm. Extracted facial features are stored as a template in the database along with other identifiers, such as a name or identification number.

During identification, the user presents his biometric sample to the sensor. The facial features extracted from it are a query that the system compares with an identifiable personality pattern using a matching algorithm. Returns a match rating that reflects the degree similarity between the template and the query. The system makes a positive decision if the compliance rating exceeds a predetermined threshold.

The two main characteristics of any biometric system are errors of the first and second kind. In the theory of radar, they are usually called "false alarm" or "skipping a target", and in biometrics the most well-established concepts are FRR (False Rejection Rate) and FAR (False Acceptance Rate) [15], [16].

- The first kind error FRR (False Rejection Rate) - the probability of a false failure (i.e. the legal user of the system has not been identified);
- The second kind error FAR (False Acceptance Rate) - the probability of false acceptance (i.e., someone successfully passed the identification under the name of a legal user)
- The system is the better, the lower the FRR value for the same FAR values. The comparative characteristic EER is sometimes used, which determines the point at which the FRR and FAR plots intersect, but it is not always representative.

Thermogram of a person is a unique distribution of temperature fields on the face. It is built using infrared cameras. Due to the low quality of such systems are not widely distributed.

Face geometry is a scanning method in which the outlines of eyebrows and eyes, lips and nose and other face elements are highlighted. After that, the distance between these elements is calculated and a three-dimensional model of the face is constructed. It takes from twelve to forty specific elements characteristic of a particular person to create and recreate a unique template [17]. There are many methods of recognition of facial geometry based on the individuality of facial features and the shape of the skull of each person [18], [19], [20].

This area is divided into two areas: 2D recognition and 3D recognition. Each of them has advantages and disadvantages, but much depends on the application area and requirements for a particular algorithm [21], [22].

III. 2D and 3D face recognition

2D face recognition is one of the most statistically inefficient methods of biometrics. It appeared quite a long time ago and was used mainly in forensic science, which contributed to its development. Later computer interpretations of the method appeared, as a result of which it became more reliable, but, of course, it was inferior and more and more inferior to other biometric methods of personal identification every year. Currently, due to poor statistics, it is used in multimodal or, as it is also called, cross-biometrics, or in social networks.

The advantages of the 2D recognition method are as follows: 2D recognition does not require expensive equipment, unlike most biometric methods. Recognition at large distances from the camera is possible with the availability of appropriate equipment.

The method's disadvantage is low statistical confidence, 2D face recognition has the highest FAR and FRR, which can cast doubt on the reliability of such an ACS system. There are requirements for lighting (for example, it is not possible to register the faces of people entering from the street on a sunny day). For many algorithms unacceptability of any external interference, such as glasses, a beard, some elements of hairstyle. 2D recognition requires a frontal image of the face, with very small deviations. Many algorithms do not take into account possible changes in facial expressions, that is, the expression must be neutral.

3D face recognition. The implementation of this method is a rather difficult task. Despite this, there are currently many methods for 3D facial recognition. Methods cannot be compared with each other, as they use different scanners and databases.

The method that implements the accumulation of information on a person is the transition from 2D to 3D. This method has better characteristics than the 2D method, but also, like it, it uses only one camera. When the subject enters the database, the subject turns its head and the algorithm connects the image together, creating a 3D template. For recognition uses several frames of the video stream.

The most used method is the template projection method. It consists in the fact that a grid is projected onto the object (person). Next, the camera takes pictures at a speed of tens of frames per second, and the resulting images are processed by a special program. The beam falling on a curved surface is bent - the greater the curvature of the surface, the stronger the bending of the beam. Initially, this used a source of visible light supplied through the "blinds". Then the visible light was replaced by infrared, which has several advantages. Usually, at the first stage of processing, images are discarded in which faces are not visible at all or there are foreign objects that interfere with identification. The obtained images restore the 3D model of the face, on which unnecessary interferences (hairstyle, beard, mustache and glasses) are highlighted and removed. Then an analysis of the model is performed — anthropometric features are highlighted, which ultimately are recorded in a unique code entered in the database. Time capture and image processing is 1-2 seconds for the best models.

The 3D recognition method is gaining popularity from the image obtained from several cameras. This method gives positioning accuracy, according to the developers' assertions, above the template projection method. The advantage of the method is the absence of the need to contact the scanning device. Low sensitivity to external factors, both on the person himself (appearance of glasses, beard, change of hairstyle), and in his environment (light, turn of the head). High level of reliability, comparable to the fingerprint identification method. Disadvantages of the method is the high cost of equipment. Commercially available systems outperformed even iris scanners. Changes in facial expressions and disturbances on the face worsen the statistical reliability of the method. The method is not yet well developed, especially in comparison with the long-used dactyloscopy, which complicates its widespread use.

IV. Setting identification and classification tasks in 2D face recognition

Purpose of the developed software is persons classification and identification. The classification makes it possible to distinguish the faces of people on the photo, and the task of identification (*I*) is to identify a person by photo of his face. To know a person, you must have at least one snapshot of him. The photo is a matrix R^{m*n}each element of which is the intensity of a pixel.

Thus, the identification function takes the form: $I: p(R1^{m*n}, R2^{m*n})\text{->}[0;1]$ or $p(R1^{m*n}, R2^{m*n})\text{->}d$, where d is the distance estimation metric between two photos. If the result is less than d, it means that the people in the two pictures turned out to be similar. Before metric calculating the frames must be converted to a special format.

The *HOG*-algorithm — a histogram of directional gradients — descriptors of singular points for object recognition [23] is used to detect faces in a photograph. *HOG*-descriptor is an image transformation into a multidimensional vector, which allows using the SVM classifier [24] or a neural network. For the classification of *HOG*-descriptors usually used SVM. An example of the *HOG*-structure of the face is shown below in Fig. 1.

The *HOG* implementation uses a sliding window classifier. Since most teaching algorithms with a teacher work in a feature space of a fixed dimension, the *HOG* feature vectors for different images must be of the same length, and therefore the images to be classified must be of the same size. Moreover, to ensure an acceptable quality of the solution to this problem, it is assumed that the image data contains objects of the same (similar) size, which are located on the same image area.

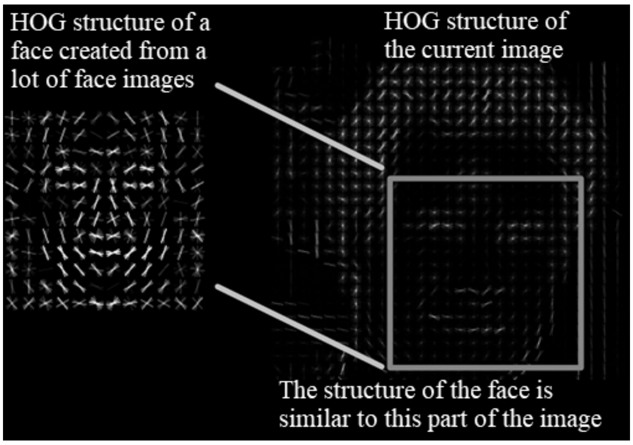

Fig. 1 Example of the *HOG*-structure of the face

The *HOG*-descriptor from the source image is calculated as follows: In the first step, color normalization and gamma correction are performed. Next, the gradient values are calculated vertically and horizontally. After that, the image is divided into a uniform grid of cells. The basic unit of a *HOG*-descriptor is a block — a rectangular area of image pixels of a given size. A block consists of cells that are assigned a histogram of directions (inclination relative to the horizontal) of gradients. The *HOG*-descriptor is the vector of the components of the normalized histograms of cells from all areas of the block. As a rule, blocks overlap, that is, each cell is included in more than one final descriptor.

The solution to the classification problem is to determine the class of an object based on its input characteristics. A training set is a set of n-dimensional vectors, and it is known to which class each vector belongs. When training a classifier, an approximate function is constructed that associates with an arbitrary object the class to which this object belongs. In the case of the SVM classifier, when training in a training set, a hyperplane is constructed, dividing the object space into classes. It is given by the following equation:

The solution to the classification problem is to determine the class of an object based on its input characteristics. A training set is a set of n-dimensional vectors, and it is known to which class each vector belongs. When training a classifier, an approximate function is constructed that associates with an arbitrary object the class to which this object belongs. In the case of the SVM classifier, when training in a training set, a hyperplane is constructed, dividing the object space into classes. It is given by the following equation: *<w, x>= b,* where w is perpendicular to the separating hyperplane, b is a scalar threshold. In this case, the vector w is a linear combination of the support vectors of the elements of the training sample, which are closest to the separating hyperplane. In the case of using *HOG*-descriptors as objects, the reference vectors are some set of *HOG*-descriptors from the training set. Therefore, to obtain SVM classifiers for windows of new sizes from the original classifier, its support vectors can be scaled, treating them as *HOG*-descriptors.

The identification task can be solved by directly comparing the unknown person found in the classification step with all those already marked. If an unknown person is found that looks like a tagged one, then it's the same person. The *HOG*-descriptor allows you to extract all areas on the original image that correspond to the found faces. For each person, it is necessary to find some basic characteristics for comparison with the characteristics of known persons.

The solution is to train the convolutional neural network (CNN) [25]. But instead of learning the network to recognize graphic objects, it is now necessary to teach it to create 128 characteristics for each person. The learning process is valid when considering 3 face images at the same time: loading the learning face images of known persons; uploading another image of the same person's face; loading the image of the face of some other person.

The algorithm considers the characteristics that it currently creates for each of these three images and adjusts the neural network so that the characteristics created by it for images 1 and 2 are a little closer to each other, and for images 2 and 3 - a little further. In case a trained neural network already exists, the rest of the task is to transfer the original images from the database through the previously trained network and obtain 128 characteristics for each person.

The final stage of identification consists in comparing the newly received image with those already existing in the database. In this paper, the Euclidean metric is used as a metric for estimating the distance between two vectors consisting of 128 face characteristics, which is the geometric distance between points x and y in n-dimensional space and is calculated using the Pythagorean theorem as the root of the sum of differences of squares of coordinates of points [26].

$$d(p,q) = \sqrt{(p_1-q_1)^2 + (p_2-q_2)^2 + \cdots + (p_n-q_n)^2} = \sqrt{\sum_{k=1}^{n}(p_k-q_k)^2} \quad \mathbf{<= 0.6}$$

Thus, the Euclidean distance is a similarity metric for a 128-dimensional vector.

V. FACE RECOGNITION SYSTEM IMPLEMENTATION.

The architecture of the prototype developed by the authors of the intelligent software and hardware security complex of the University, which implements facial recognition, consists of a number of components that ensure the operation of the complex.

The hardware architecture of the complex is shown in Fig. 2.

The designed complex has a distributed structure of servers: application server, file server and database server. These servers can be located on one or more physical servers.

Surveillance monitors that will be located at the security Desk, and CCTV via the switches connected to the application server, the programs which provide the processing of the captured video stream, recognizing and identifying individuals and recording data about employees in a database and display the results on monitors. The captured video stream is stored on a file server. Devices and servers can be connected via LAN or WI-FI.

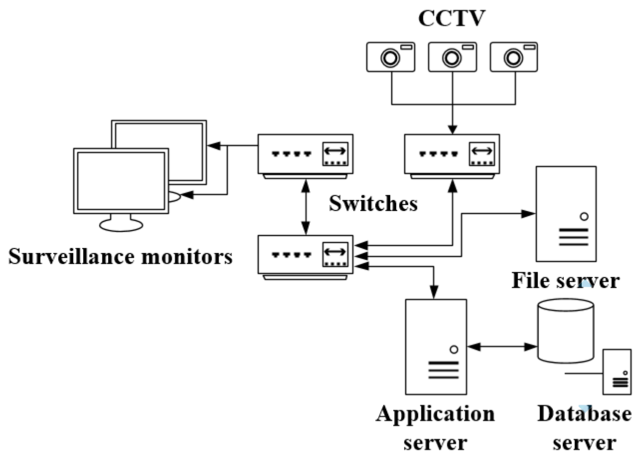

Fig. 2 The hardware architecture of the face recognition complex

The software architecture of the complex is shown in Fig. 3.

The application server contains a set of programs that provide the operation of the entire complex. Each program is a separate component with specific objectives. The application server consists of the following applications:

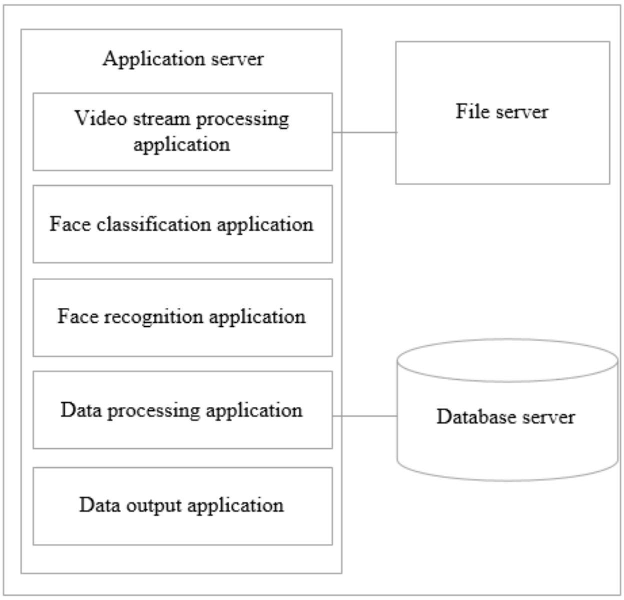

Fig. 3 The software architecture of the face recognition complex

- a video stream processing application that provides video streaming capture from video surveillance cameras and frame preprocessing, in addition to storing data on a file server;
- face recognition application that searches for areas in the frame that contain faces;
- a person identification application that provides a search for matching the recognized persons in the database;
- a data processing application that provides the necessary processing of the information received and then writes it to the database (for example, information about the time of employee detection);
- data output application that provides video output to monitors of the security post and service information, including recognized names of persons and other service data.

The authors carried out the development of the infological and datalogical models and the implementation of a database containing video from the camera, sets of segmented frames for algorithms training and other working information.

Dataset is stored in the database containing 128-dimensional vectors characterizing each face. Each vector is formed from a collection of pre-marked images. Images correspond to persons for whom access to the building is open.

The storage for 128-dimensional vectors is selected based on the requirements and contains a relational database, NoSQL solutions, text files, video and graphics.

Some of the information is stored on a file server in the repository, the structure of which includes:

- folder /new_ph, in which photos of new people are added to provide them access to the system;
- folder /mat, in which all transformed faces are saved in a 128-dimensional vector. This folder stores text files of the ".txt" extension into which the vectors are written. File names correspond to the names of employees.
- folder /ph contains files from the /new_ph folder after processing.

Capturing video from CCTV cameras starts after the launch of the complex. Frame-by-frame processing of the captured video stream occurs on the application server. Processing is the pre-processing of frames for face recognition.

Capturing video from CCTV cameras is done by the *OpenCV* library [27]. *OpenCV* allows frame-by-frame processing.

When faces are detected on the video stream, they are compared with those already present in the database. If the distance between the vectors of the received face image from the video stream and the face image from the database does not exceed the specified threshold value, therefore the person is detected in the database and must be entered into the building. This means that the person has successfully passed the authentication stage.

There are many freely distributed libraries with an already pre-trained neural network for solving the problems of classification and identification of people's faces. One of these libraries is *face_recognition* [28], written in a high-level programming language *Python* [29]. This library uses *dlib* [30] written in *C++*.

The designed face recognition algorithm shown in Algorithm 1.

Algorithm 1 A script on Python, describing face recognition

```
import face_recognition as fr
import cv2
vc = cv2.VideoCapture(0)
ret, frame = vc.read()
frame = cv2.resize(frame,
                (0, 0),    fx=0.25,  fy=0.25)
rgb_frame = frame[:, :, ::-1]
faces = fr.face_locations(rgb_frame)
face_encodings = fr.face_encodings(rgb_frame, faces)
for face_encoding in face_encodings:
        matches = fr.compare_faces(known_face_encodings,
face_encoding)
        name = "Unknown"
        if np.any(matches):
                index=matches.index(True)
                name=known_faces[index]
        face_names.append(name)
```

First, using the *OpenCV* library, one frame is captured from the video stream, which is reduced to 1/4 frame in order to increase the speed of face recognition, and the frame color scheme is converted from BGR to RGB.

In the next step, the *dlib library* classifies all areas using the “sliding window” method using *HOG* descriptors, detecting faces on a pre-processed frame. The detected regions are shown in fig. 4.a., and each area is assigned a weight value, which indicate whether a person is in a given area fig. 4.b. When the weight is more 1, then there is a person in this area.

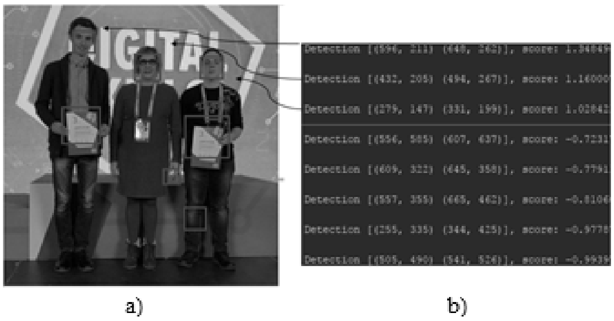

Fig. 4 Face detection on the image

There are 3 areas found on the image shown in Fig. 4 the weight of which turned out to be more than 1, therefore, it is more likely that there are faces there.

You should to select the data after finding faces on the image. The *face_recognition*library using the *face_encoding()* function allows you to encode all the regions found in a 128-dimensional vector. An example of the coding area is presented in Fig.5.

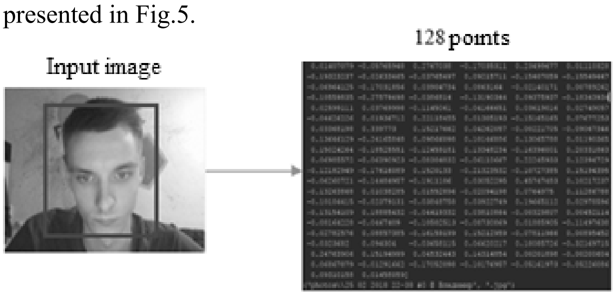

Fig.5. The result of encoding the region into a 128-dimensional vector characterizing the face

At the last step, it is required to find a person in the database of known persons of the program that has the characteristics closest to the characteristics of the test image. To compare two 128-dimensional vectors, use the *compare_faces()* function of the *face_recognition* library, which uses the Euclidean distance as a proximity metric. For each identified person, its name is added to the *face_names* array. As a result, on the basis of free software, you can implement your own system of recognition and identification of persons.

VI. Conclusion

The article discusses the development of a subsystem for the classification and identification of persons for the organization security system. Analyzed solutions from leading IT companies in this field and modern biometric identification systems. Considered in detail the approach to image transformation into a multidimensional vector named “*HOG*-descriptor”, as well as the method of "sliding window", which allows to classify all interested objects- persons to solve the problem of classification and identification of individuals. Designed face recognition subsystem. A program for 2D face recognition has been developed, which preserves 128-dimensional characteristics-features in the freeware database. Examples of using and obtained results are given.

References

[1] “Biometric identification. Definitions and basic methods”, TechPortal website. Web:http://www.techportal.ru/ glossary/ biometricheskaya_identifikaciya.html.
[2] “Technology features of Amazon Rekognition”, Amazon official website,Web:https://aws.amazon.com/ru/rekognition/.
[3] “Facebook Accessibility. Face Recognition Features”, Facebook official website, Web:https://www.facebook.com/ accessibility/ videos/new-face-face-recognition-features/1628143837229335/.
[4] “Apple FaceID technology”, Appleofficial website, Web:https://support.apple.com/ru-ru/HT208108.
[5] “Supermarkets trial facial recognition at checkouts”, China Plus,Web:http://chinaplus.cri.cn/news/business/12/20180502/125397.html.
[6] “JD.com overtook Amazon to introduce automated shopping”, NewRetail.com, Web:https://new-retail.ru/novosti/retail/ jd_com_obognala_amazon_po_vnedreniyu_avtomatizirovannykh_magazinov_5726/.
[7] “JD.com is Testing Facial Recognition for Pigs”, RadiiChina.com,Web:https://radiichina.com/jd-com-is-testing-facial-recognition-for-pigs/.
[8] Official web site of NTechLab company,Web:https://ntechlab.ru/.
[9] Official web site of Macroscope company,Web: https://macroscop.com/.
[10] V. I. Voronov, K. V. Genchel, M. D. Artemov and D. N. Bezumnov, ““Surdotelephone” project with convolutional neural network”, 2018 Systems of Signals Generating and Processing in the Field of on Board Communications, Moscow, 2018, pp. 1-6.
[11] A. Goncharenko, L. I. Voronova, V. I. Voronov, A. A. Ezhov and D. V. Goryachev, “Automated support system designing for people with limited communication”, 2018 Systems of Signals Generating and Processing in the Field of on Board Communications, Moscow, 2018, pp. 1-7.
[12] Voronov Vyacheslav I., GenchelKsenia V., Voronova Lilia I., Travina Maria D. “Development of a Software Package Designed to Support Distance Education for Disabled People”, IEEE-International Conference "2018 Quality Management, Transport and Information Security, Information Technologies" (IT&QM&IS-2018), St.Petersburg, pp.746-751 (in press)
[13] Mikhail D. Artemov, Vjacheslav.I.Voronov, Lilia I. Voronova “Development of the Subsystem for Recognition of Simple Dynamic Gestures Using 3DCNNLSTM”, 24th Conference of Open Innovations Association FRUCT, Moscow, Russia, 8-12 April 2019 (unpublished)
[14] Artem A. Goncharenko, Lilia I. Voronova “Development of a Sign Language Recognition Information System Using Wireless Technologies for People with Hearing Impairments” 24th Conference of Open Innovations Association FRUCT, Moscow, Russia, 8-12 April 2019 (unpublished)
[15] Modern biometric identification methods. Habr.com Web:https://habrahabr.ru/post/126144.
[16] Martynova L.E., Umnitsyn M.Y., Nazarova K.E., Peresypkin I.P. “Research and comparative analysis of authentication methods”, Young scientist, 2016, №19,pp. 90-93.
[17] Petrisor, Daniel & Fosalau, Cristian & Avila, Manuel & Mariut, Felix. (2011). Algorithm for Face and Eye Detection Using Colour Segmentation and Invariant Features. 2011 34th International Conference on Telecommunications and Signal Processing, TSP 2011 - Proceedings. 564-569. 10.1109/TSP.2011.6043666.
[18] Yavuz, Hasan & Cevikalp, Hakan & Edizkan, Rifat. (2016). A comprehensive comparison of features and embedding methods for face recognition. Turkish Journal of Electrical Engineering and Computer Sciences. 24. 313-340. 10.3906/elk-1301-65.
[19] Sharma, Shailja & Gupta, Manish. (2016). An improved connected component based algorithm for face recognition. 4931-4935. 10.1109/ICEEOT.2016.7755659/

[20] Cheng, Eric-Juwei & Prasad, Mukesh & Puthal, Deepak & Sharma, Nabin & Prasad, om & Chin, Po-Hao & Lin, Chin-Teng & Blumenstein, Michael. (2017). Deep Learning Based Face Recognition with Sparse Representation Classification. 665-674. 10.1007/978-3-319-70090-8_67.

[21] A, Vinay & Rajesh Sampat, Pratik & V. Belavadi, Sagar & Pratik, R & Rao, Nikitha & Ragesh, Rahul & Balasubramanya Murthy, Kannamedi & Subramanyam, Natarajan. (2018). Face recognition using interest points and ensemble of classifiers. 1-8. 10.1109/RAIT.2018.8389033.

[22] Wei, Qiuyue & Mu, Tongjie & Han, Guijin & Sun, Linli. (2019). Face Recognition Based on Improved FaceNet Model. 10.1007/978-3-030-03766-6_69. In book: Proceedings of the Fifth Euro-China Conference on Intelligent Data Analysis and Applications, pp.614-624/

[23] N. Dalal, B. Triggs. “Histograms of Oriented Gradients for Human Detection”, Cordelia Schmid and Stefano Soatto and Carlo Tomasi. International Conference on Computer Vision & Pattern Recognition (CVPR '05), Jun 2005, San Diego, United States. IEEE Computer Society, 1, pp.886--893, 2005 Web:http://lear.inrialpes.fr/people/triggs/pubs/Dalal-cvpr05.pdf.

[24] Bartlett P., Shawe-Taylor J. “Generalization performance of support vector machines and other pattern classifiers”, Advances in Kernel Methods, MIT Press, Cambridge, USA, 1998. Web:http://citeseer.ist.psu.edu/bartlett98generalization.html.

[25] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel: “Backpropagation Applied to Handwritten Zip Code Recognition”, Neural Computation, 1(4):541-551, Winter 1989.

[26] M. Deza, E. Deza. “Encyclopedia of Distances”, Springer-Verlag, 2009, ISBN 978-3-642-00233-5

[27] Computer vision library with the source code OpenCV,Web:https://opencv.org/.

[28] Face Recognition and Identification Library “face_recognition”,Web:https://github.com/ageitgey/face_recognition

[29] High Level Python Programming Language, Web:https://www.python.org.

[30] Machine learning library “dlib”, Web:http://dlib.net.