Sign Language Recognition Information System Development Using Wireless Technologies for People with Hearing Impairments

Artem Goncharenko, Lilia Voronova, Mikhail Artemov, Vjacheslav Voronov, Danil Bezumnov
Moscow Technical University of Communications and Informatics
Moscow, Russian Federation
artem@agsys.ru, voronova.lilia@ya.ru, artemov_mikle@mail.ru, vorvi@mail.ru, danbez@yandex.ru

Abstract—**The article presents the development of a sign language recognition information system for people with hearing and / or voice impairments, allowing to improve the quality of life and interaction in society with other people. The device, software, functional blocks and information system subsystems are described. Examples of possible application and placement of the system in various spheres of public life are given. One of the types of implementation of the information system of gesture recognition is described in the local network. The system requirements are listed based on the characteristics of the life of the target audience of people and the life cycle of gesture recognition. The decomposition of information system subsystems was made on the basis of belonging to each of the program blocks: the client and the server.**

I. INTRODUCTION

Modern world information technologies have taken the course to adapt people with disabilities. The global community has more than 360 million people with hearing and / or voice impairments. Deaf people communicate in a special sign language. He distinguishes between two subspecies: a language with the use of a dactyl (each character is shown separately) and a language with the use of gestures denoting words and phrases. To adapt such people in society, there is a special profession - sign language interpreter (a specialist who translates oral speech into sign language). Training this profession is quite difficult and long, and the cost of such specialists is high.

The communication gestures system for the deaf is the most complex in terms of volume, structure, and functions among the well-known kinetic systems combined by motor acts of transmission and perception of information. There are different sign languages in different countries, but all of them are invented by people in the process of adaptation and include many descriptive, imitative and indicating gestures [1].

In addition to national sign systems, there is an international sign language called gestuno. This language arose from the fact that translation from the sign language of one nationality into the sign language of another was carried out in several steps. When translating from a foreign sign language into Russian, it is first necessary to translate sign language into ordinary national, then translate from this national language into Russian, then translate from Russian into the Russian sign language. It took quite a lot of time and also created problems during various international meetings and events of the World Federation of the Deaf. Therefore, the International Committee, which includes several countries, selected and published a vocabulary, which includes about 1,500 gestures grouped by topics: people, interpersonal relations, justice, nature, etc. Gestures included in the international lexical fund are mainly selected from national gestures, and in part - re-created by members of the commission.

Russian sign language can be divided into two types: Russian and tracing [2]. Russian gestural speech (RZHR) is communication using the means of the Russian sign language - an original linguistic system with peculiar vocabulary, grammar, etc. There is an online dictionary for RZHR [3]. Calculating gestures (QOL) - traces the linguistic structure of the verbal language. Calculating in linguistics - borrowing foreign words, expressions, phrases by literal translation of the corresponding language unit. Calculating gesture speech is a secondary sign system that is learned at the base and in the process of studying a verbal speech by a deaf child. Gestures here are equivalents of words, and their order is the same as in the usual sentence.

In the direction of recognition of the sign languages of people with hearing impairments in different countries, active research and development is being conducted. Many scientific papers are devoted to the use of neural networks for gesture recognition.

In the work of George Stresoski, Dario Stojanovski, Ivica Dimitrovski and Georgi Madzharov, presented in their article [4], they conducted experiments with various types of convolutional neural networks, including their own developed model. The performance of each model was assessed using the Marcel dataset [5], which provides a relevant understanding of how the various characteristics of the network architecture affect the performance and quality of recognition. The best results were obtained using the GoogLeNet approach [6] using the Inception architecture [7], and then their proprietary model and the VGG model [8].

Inception is an effective deep neural network architecture designed for computer vision, which got its name from the "Network in the Network" described in the article by Lin et al. [9] in conjunction with the famous "We need to go deeper" Internet meme [10]. In this case, the word "deep" is used in two different senses: firstly, in the sense that a new level of organization is introduced in the form of "Inception module" (initial module), and secondly, in a more direct sense, an increase in the depth of the network.

In the Moscow Technical University of Communications and Informatics (MTUCI) there is a system of internal grants for research work. As part of this system, in 2018, the department "Intelligent Systems in Management and Automation" conducted research work on the project "Developing an Intelligent Information and Communication System for Social Accessibility for People with Disabilities Based on Machine Learning Methods" Ph.D. V.I. Voronov. [11], [12], [13]

The object of R & D are algorithms, technologies and means of transforming information between natural and sign languages based on machine learning methods [14].

The goal of the project: the development of a software package based on machine learning methods that provide the functionality of the communication system of social accessibility for people with hearing impairments [15].

In the process, studies were conducted on models and methods of data mining to convert information from sign language into colloquial, designing tools that ensure communication of hearing impaired people, including fingerprint recognition, gestures, gestures and emotions [16].

As part of the research, a large number of experiments on the recognition of gestures with the help of trained neural networks were conducted. The obtained results give grounds for cautious optimism in connection with the active development of convolutional and recurrent neural networks and other data mining methods used for pattern recognition [17].

The article describes the real-time sign language recognition information system: information system architecture, hardware, operation and data transfer algorithms.

II. Hardware organization of gesture recognition information system

Gesture recognition information system is a project of the Department of Intelligent Systems in Control and Automation at the University of MTUCI, which includes the development of subsystems for recognizing various types of communication between people with hearing and / or voice disabilities, fingerprints, dynamic gestures, gestures using the user's hands and his emotions. The recognition system is a combination of software and hardware tools that allow remotely using the recognition system from various client devices of the user. The information system can be used for various types of hardware placement and for the following purposes: alerting emergency services, used as a translator in real time, with distance learning in educational institutions, in enterprises. Placing an information system can be done locally, using the hardware and computing capabilities of the organization using the system, or remotely, with the placement of the server part of the information system, for example, in a data center, using cloud technologies. In the latter case, the Internet will be used as data transfer between devices of the system. The general hardware organization of the information system is depicted in Fig. 1.

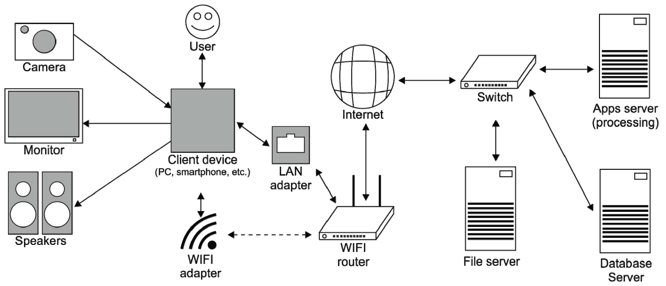

Fig. 1. Information system hardware

Gesture recognition information system (IS) includes the following hardware devices, presented in Table I.

TABLE I. GESTURE RECOGNITION HARDWARE DEVICES

№	Type of device	Functionality
1	Camera	Record video data movement of the user's hands and face
2	Monitor	User interaction with system interface and data output
3	Speakers	Sound reproduction of the sign language translation results
4	Client device	Control of input and output devices, use of an IS client
5	WIFI and LAN adapters	Connecting to the network of the client device and data transfer
6	WIFI router and switch	Combining elements of the system into a single network
7	Internet	Remote device connection
8	Apps server	Provide gesture recognition and server interaction
9	Database server	Providing database storage, access and information processing
10	File server	Providing storage and file access for a training set of neural

As a client device can act: personal computers, laptops, smartphones, tablets and others. The choice of device depends on the private implementation of the information system.

The information system has a distributed server structure: applications, files and databases. Depending on their use, they can be located on one or on several physical servers.

The camera, display, network and acoustic components are connected to the client device and interact with it using drivers. Client (application) of the information system is installed on the client device; it provides recording and preprocessing of video, displaying translation results, receiving and sending data over the Internet to an application server. A wireless connection via WIFI is used as communication lines, or for fixed devices there is a cable connection in the LAN.

When implementing the information system, the authors chose the option of locating the information system in the local network, excluding the Internet, and the router is directly connected to the switch. The client and server side are the router and the switch.

A router provides wireless connectivity for mobile devices and a wired connection for fixed clients, and a switch unites several physical servers into a single local area network. Data exchange in the network takes place over IPv4 and IPv6.

III. Software and Functionality of Hardware of an Information System

The application server (Fig. 2) performs the function of the main server of the information system. It is installed: Linux operating system Ubuntu 16.04, Web server software, FTP server for accessing executable files, interpreters of programming languages PHP, Python and Java.

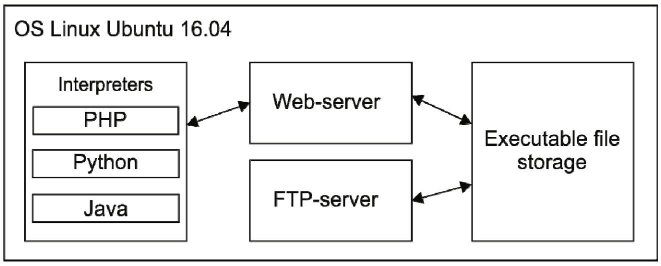

Fig. 2. Application server software

The database server (DB) performs the function of storing and accessing databases of the information system, which contain data from trained neural networks, user data, and statistics on the information system operation. The database server is running Linux Ubuntu 16.04 operating system, MySQL server relational database software.

The file server performs the function of storing and accessing photos and video files for learning neural networks, logging files of the information system. The server is running Linux Ubuntu 16.04 operating system and file FTP server.

The overall hardware and software organization of the servers that make up the server part of the gesture recognition system is shown in Fig. 3.

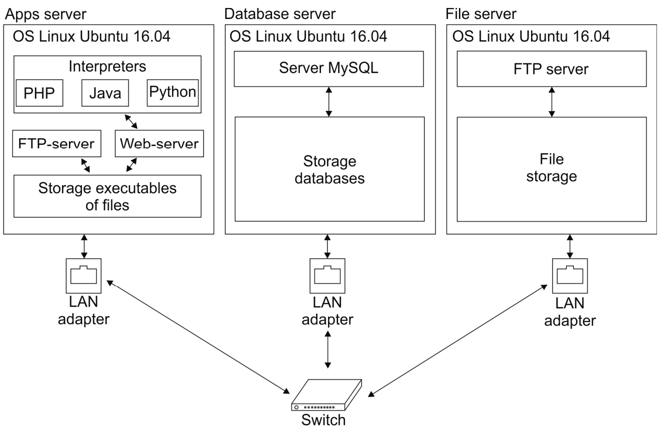

Fig. 3. General hardware and software organization of servers

As a client device, consider a laptop. It should provide video recording using a camera, playing audio files using speakers, displaying text data on a display and exchanging data over the Internet. The operating system installed on the client device is Windows 10 OS. For access of the user through the browser to the recognition system, Google Chrome version 71.0 and higher is used. In addition, for user access to the sign language recognition system, it is possible through an installed client program based on the Microsoft .NET Framework version 4.5 or higher.

Work through a web-browser client of the information system is ensured by the ability of the browser to access the camera, display and audio equipment of the client device through requests to the operating system. This client allows you not to install a special information system application on the user's device and use a gesture translator from any device that has a web browser and a camera.

Using the same technology, the client software installed in the operating system, but having a greater amount of possibilities, compared to a web browser, works. The program has direct access to the OS and can use the distribution of computing resources of the processor cores of a laptop using threads. Another advantage of the client program is access to the hard disk (file storage), in which the client can store and process materials recorded using the camera.

The hardware-software organization of the client device of the gesture recognition information system presented in the form of a user's laptop is shown in Fig. 4.

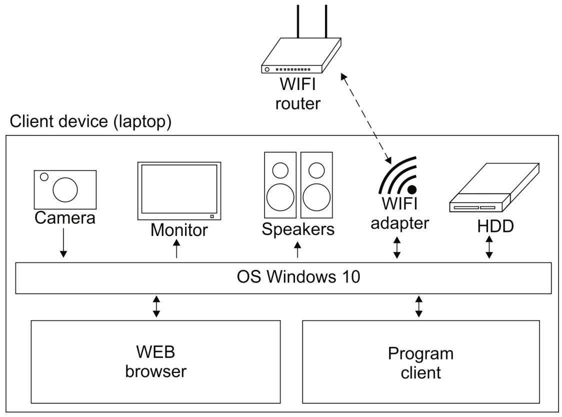

Fig. 4. Hardware and software organization of the client device

IV. Gesture Recognition Software Requirements

Requirements for the gestural speech recognition information system, based on the characteristics of the lives of people with hearing impairments, are given in Table II.

TABLE II. System requirements based on life characteristics

Peculiarity of people's lives	Requirement to the software package
The speed of phrases and words practically does not differ from ordinary speech	High processing speed
All gestures are dynamic	Video stream processing
Emotional expressions occur when mouth and lips move	Having the ability to recognize facial expressions (emotions)
Both hands are used in combination with the upper body	The location of the recording device (camera) at a distance
Do not have special equipment for recognition	Client on affordable hardware (PC, smartphones, etc.)
Lack of voice	Reproduction of translation results in text and audio form
People of different ages, nationalities and races	The possibility of continuous learning of the recognition system
The use of recognition systems in different light conditions	Increase the training set of neural networks due to user data
Limited amount of Internet traffic in the tariffs of operators	Adaptive preprocessing of photo and video data transmitted to the server

In addition to the features of people's daily lives, we consider the life cycle (Fig. 5) of recognizing the gestures of an information system. It describes the interaction and execution of functions within the information system from the moment of launching the client application to exiting it. The life cycle process involves looping functional blocks from the moment they start to the end of gesture recording, since the system works in real time with a continuous stream of video data. Before being sent over the network, the video is pre-processed - cropping and compression to reduce the transmission time and the use of Internet traffic. Once on the server, the video is converted to the required format for the neural network and begins to be processed. After that, the result is formed into the translation message, processed for correctness and literacy of the Russian language, returned back to the application and reproduced.

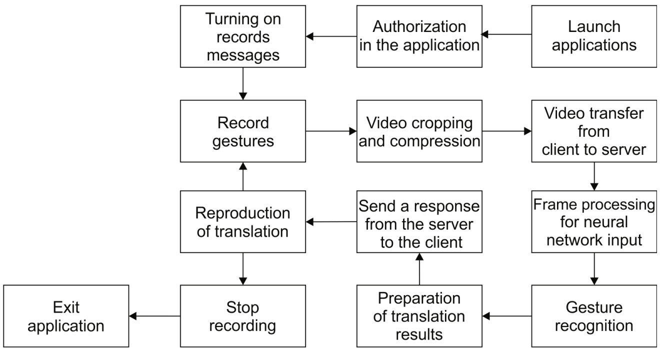

Fig. 5. Gesture Recognition Life Cycle

Based on the requirements for the software package and the gesture recognition life cycle, the following functional systems and subsystems of the sign language recognition information system have been compiled with the separation of the server and client parts (Fig. 6).

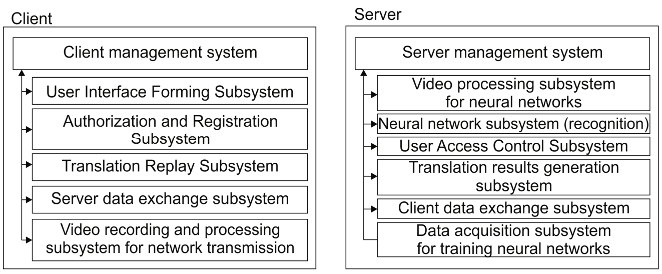

Fig. 6. Functional subsystems of the sign language recognition information system

Each subsystem contains a number of its own modules. Their functional features and interaction with other subsystems are described in the corresponding sections below.

V. ORGANIZATION OF THE CLIENT OF THE SIGN LANGUAGE RECOGNITION INFORMATION SYSTEM

A client is an installed application (on mobile devices), a program (on laptops and PCs), or a web browser application that provides user interaction with the sign language recognition information system. The decomposition into modules of each of the subsystems of the client of the sign language recognition information system is shown in Fig. 7.

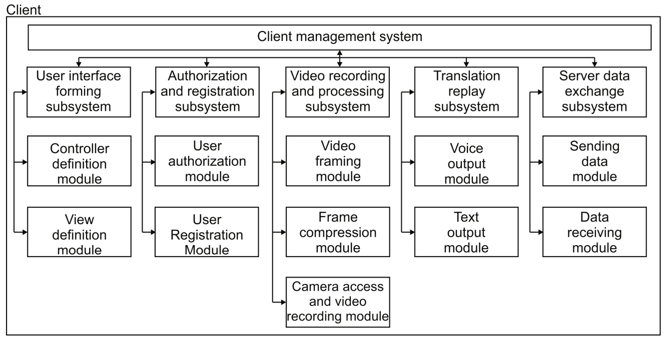

Fig. 7. Client functional subsystems of the sign language recognition information system

Each of the subsystems interacts with each other by referring to the client management system. The control system responds to user actions and performs the appropriate tasks. Table III lists all subsystems, their modules and functional features.

TABLE III. CLIENT FUNCTIONAL FEATURES OF SUBSYSTEMS AND MODULES

Subsystem / Module	Functionality
User Interface Forming Subsystem	Provides automatic generation of visual elements of the client application interface (forms, buttons, windows, etc.)
Controller Definition Module	Responsible for the automatic selection of the required controller that processes the output of data to visual elements in accordance with the selected user action
View Definition Module	Responsible for the automatic selection of views for the formation of the user interface based on controller data
Authorization and Registration Subsystem	Provides protection from unauthorized access to the system
Authorization module	Performs the function of checking data entry to access the recognition system.
Registration module	Performs the function of registering new users and recovering data for entering the system.
Video recording and processing subsystem for network transmission	Provides recording and preliminary processing of video data from the client device camera to reduce the amount of transmitted traffic of the Internet network
Video framing module	Performs splitting video stream into frames in accordance with the established time interval.
Frame compression module	Performs the resize function for cropping data
Camera access and video recording module	Performs the function of connecting to the camera of the client device, start recording video data and storing it for processing.
Translation Replay Subsystem	Provides text or voice output of the results of the translation of the sign language to the monitor or the dynamics of the user's device
Voice output module	Provides audio recording from a set of stored files based on the translation results.
Text output module	Provides access to the client device monitor and text data output via the interface generation subsystem
Server data exchange subsystem	Provides client interaction with the information system server
Data sending module	Performs the function of storing server API access addresses, generating request data in the required format (JSON, XML, Websocket) and sending the client's request to the server
Data receiving module	Performs the function of receiving, translating the format of receiving JSON data into an internal variable and transferring the data to the appropriate subsystem in accordance with the action of the user or system

Depending on the hardware characteristics of the client device, the video processing subsystem may not be used in the client information system if the computing load on the device is high and it cannot provide high performance and process in real time. In this case, video data will be transmitted over the network and processed on the server.

VI. ORGANIZATION OF THE SERVER OF THE SIGN LANGUAGE RECOGNITION INFORMATION SYSTEM

The server is a software implementation of the sign language recognition information system that provides user access control, data processing and preparation for recognition by neural networks, recognition of gestures and emotions, and generation of translation results. Decomposition into modules of each subsystem of the client of the sign language recognition information system is shown in Fig. 8.

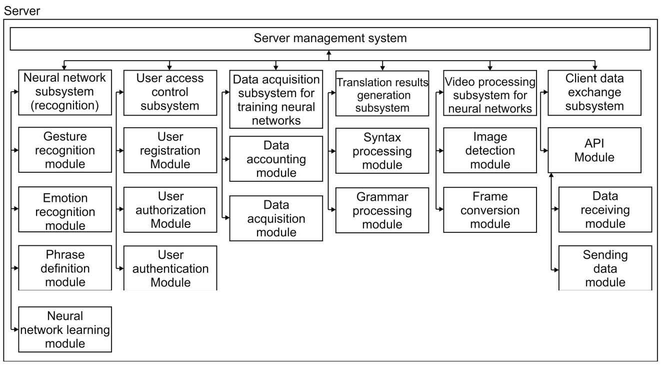

Fig. 8. Server functional subsystems of the sign language recognition information system

Server subsystems interact with each other through the server management system, which in turn responds to the received requests from the client application of the information system. The functionality of each subsystem and its modules are described in Table IV.

TABLE IV. SERVER FUNCTIONAL FEATURES OF SUBSYSTEMS AND MODULES

Subsystem / Module	Functionality
1	2
Video processing subsystem for neural networks	Provides data processing for input to the recognition subsystem
Image detection module	Performs hand detection, faces in the frame and their selection
Frame conversion module	Performs frame conversion based on the required input data into neural network modules.
Neural network subsystem (recognition)	Provides recognition of gestures, emotions, and their combination for translating sign language
Gesture recognition module	Performs the function of recognizing gesture or movement of hands in a frame based on trained neural networks.
Emotion recognition module	Performs the function of recognizing the user's emotions in the frame based on the trained neural networks
Phrase definition module	Performs the recognition of phrases and words based on the data obtained by the modules of the recognition of gestures and emotions
Neural network learning module	Performs the training function based on the accumulated training set transferred from the data acquisition subsystem for training neural networks
1	2
Translation Results Generation Subsystem	Provides competent translation of phrases and words in accordance with the lexical, syntactic and grammatical features of the Russian language
Syntax processing module	Performs syntactic processing of the received data from the recognition subsystem.
Grammar Module	Performs grammatical processing of the received data from the recognition subsystem
Data acquisition subsystem for training neural networks	Provides accounting and data accumulation for the replenishment of the training set of neural recognition networks
Data accounting module	Performs the function of selection and recording of data, for the training of neural networks of recognition of gestures and emotions
Data acquisition module	Performs the function of storing and accessing data received from the data accounting module
User Access Control Subsystem	Protects the system from unauthorized access
User Registration Module	Performs user registration in the system and storing information in the database
User Authorization Module	Authorizes users and provides access to the system
User Authentication Module	Authenticates the received data for access based on the data stored in the database.
Client data exchange subsystem	Provides data exchange between client and server information system
API module	Performs the function of accessing the system and routing requests to specific addresses
Data receiving module	Performs the function of processing received data from the client to the internal data format used
Sending module	Performs the formation of data in formats JSON, XML and transfer to the client

VII. API INTERACTION BETWEEN CLIENT AND SERVER INFORMATION SYSTEM

The sign language information system has several client applications. For Android mobile devices, the application is implemented in the JAVA programming language, C # is used for PCs and laptops on Windows, and PHP is used for the browser client. To merge client applications written in different programming languages with the server, you must use the API. For routing API requests and data access, several types of requests (GET, POST, PUT, DELETE) are sent to the address of the information system server using the HTTP protocol. The API allows you to simultaneously develop and support multiple versions of the server application. When updating on the client, the appropriate version of the API is set.

The work of the user access control subsystem on the client device is provided by contacting the address containing the domain name or IP address of the server, the version of the API, and the method being called. For example, to register a user, the data access address looks like this: https://192.168.1.107/v1.01/Users (POST method).

To get data about the user, the GET method is used, with the identification number of the user in the system: https://192.168.1.107/v1.01/Users/101 (GET method). The DELETE method is used to delete a user from the system, and the PUT method is used to change user information.

In general, the client application to the server of the sign language recognition information system is as follows:

https://<address of the domain name or IP>/<API version>/<called object>

For additional protection of data transmission, the HTTPS protocol is used. This protocol is an extension of the HTTP protocol and allows you to use data encryption using special digital certificates. To use the HTTPS protocol, you must obtain a certificate that is issued by specialized certification authorities on the Internet. A certificate is issued for each domain name and, depending on the type of certificate, its subdomains, thereby providing confirmation of the certificate's membership in the domain. There are self-created certificates, they are called "self-signed", but the responsibility for data encryption and the security of such a certificate lies with the developer and may be vulnerable. To receive a free HTTPS certificate, the global Internet community offers users a non-commercial certificate authority Let's Encrypt. This center issues certificates for a period of 3 months, after which it is necessary to renew the certificate. Its advantage is also that the entire process of obtaining a certificate is automated by installing the appropriate software on the server to which the domain name belongs.

VIII. CONCLUSION

A sign language recognition information system developed by the authors with the use of wireless technologies for people with hearing disabilities will improve the quality of life and the interaction of people with such disabilities in society.

This system can be applied both in education, in enterprises, and in the daily life of people with disabilities as a real-time translator.

In the project, the authors suggest:

- Distribution of computing power between hardware and physical means to improve the interaction between them and support the health of the system as a whole (for example, at the enterprise or at office, by means of the Internet);
- Implementation of several client applications under different types of devices (computers, smartphones, tablets and others).
- Automated addition and updating of the data collection system after collecting additional photo and video data for subsequent retraining of the neural network and improving the quality of recognition of the gesture conversation.

REFERENCES

[1] Stokoe William C. Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf Archived. // *Wayback Machine., Studies in linguistics: Occasional papers (No. 8).* Buffalo: Dept. of Anthropology and Linguistics, University at Buffalo – 2013

[2] Zaitseva G.L. *Sign language. Dactylology: Textbook. for stud. higher studies. institutions.* - M .: Humanit. ed. Center VLADOS, 2000. - 192 p.

[3] *DigitGestus* [Electronic resource]: RSL online dictionary. URL: http://www.digitgestus.com/index.html (access date: 01.29.2019)

[4] Strezoski G. Hand Gesture Recognition Using Deep Convolutional Neural Networks / G. Strezoski, D. Stojanovski, I. Dimitrovski, G. Madjarov // *Advances in Intelligent Systems and Computing.* — 2018. — 665. – Pp. 49-58.

[5] Marcel S. Hand gesture recognition using Input/Ouput Hidden Markov Models / S. Marcel, O. Bernier, J-E. Viallet, D. Collobert. // *Proceedings of the 4th International Conference on Automatic Face and Gesture Recognition.* — 2000. — Pp. 456-461.

[6] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, 2015

[7] Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *CoRR, abs/1207.0580*, 2012.

[8] *Visual Geometry Group* [Electronic resource]: Very Deep Convolutional Networks for Large-Scale Visual Recognition. URL: http://www.robots.ox.ac.uk/~vgg/research/very_deep/ (access date: 01.30.2019)

[9] Min Lin, Qiang Chen, and Shuicheng Yan. *Network in network*. CoRR, abs/1312.4400, 2013.

[10] *Know your meme* [Electronic resource]: We need to go deeper. URL: http://knowyourmeme.com/memes/we-need-to-go-deeper (access date: 30.01.2019)

[11] 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.

[12] A. S. Trunov, L. I. Voronova, V. I. Voronov, D. I. Sukhachev and V. G. Strelnikov, «Legacy applications model integration to support scientific experiment», *2018 Systems of Signals Generating and Processing in the Field of on Board Communications*, Moscow, 2018, pp. 1-7.

[13] Trunov Artem S., Voronova Lilia I., Voronov Vyacheslav I., Ayrapetov Dmitriy P. «Container Cluster Model Development for Legacy Applications Integration in Scientific Software System» IEEE-International Conference *«2018 Quality Management, Transport and Information Security, Information Technologies»* (IT&QM&IS-2018), St. Petersburg, pp.815-819 (in press)

[14] Voronov Vyacheslav I., Genchel Ksenia 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)

[15] A. 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.

[16] Vyacheslav I. Voronov, Vladimir G. Strelnikov, Liliya I. Voronova, Artem S. Trunov, Andrey G. Vovik «Faces 2D-Recognition and Identification Using the HOG Discriptors Method» *24th Conference of Open Innovations Association FRUCT*, Moscow, Russia, 8-12 April 2019 (unpublished)

[17] Mikhail D. Artemov, Vyacheslav I. Voronov, Lilia I. Voronova, Artem A. Goncharenko, Vasiliy A. Usachev "Subsystem for Simple Dynamic Gestures Recognition Using 3DCNNLSTM". *24th Conference of Open Innovations Association FRUCT*, Moscow, Russia, 8-12 April 2019 (unpublished).