Indonesian Journal of Electrical Engineering and Computer Science
Vol. 35, No. 2, August 2024, pp. 1311~1324
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v35.i2.pp1311-1324

Gesture recognition technology: a new dimension in human-computer interaction interface

Nurbol Beisov, Gulnar Madyarova, Nurassyl Kerimbayev
Department of Computer Science, Faculty of Information Technology, Al-Farabi Kazakh National University, Almaty, Kazakhstan

Article Info

Article history:

Received Dec 11, 2023
Revised Apr 1, 2024
Accepted Apr 13, 2024

Keywords:

Alphabet of Kazakh language
CNN
Gesture recognition
Human-computer interaction
Kazakh sign language

ABSTRACT

Development of an interface for intelligent gesture control to improve user experience and increase the efficiency of interaction with a computer. This paper proposes a gesture recognition system based on artificial intelligence using convolutional neural networks (CNN). The system comprises three stages: pre-processing, optimal frame determination, and gesture category identification. The extracted features used are independent of movement, scaling, and rotation, providing greater flexibility to the system. The suggested gesture control technology, known as Kazakh Sign Language (KSL) for Kazakh alphabets, eliminates the need for additional devices, enabling users to interact with the system naturally. Experiments demonstrated that the proposed KSL system can accurately recognize Kazakh language alphabet letters with a high precision of 97.3%, owing to the utilization of artificial intelligence and CNN to enhance the accuracy and effectiveness of gesture control. Gestures, a type of visual formation, are perceivable by computers through machine learning models. The selection of methods and systems for recognizing Kazakh sign language gestures was accompanied by addressing various challenges related to language-specific orthographic and gestural features. The developed gesture control interface for human-computer interaction is applied in the field of inclusive education, aiming to assist deaf and hard-of-hearing children in learning sign language.

Corresponding Author:

Nurbol Beisov
Department of Computer Science, Faculty of Information Technology
Al-Farabi Kazakh National University
Almaty, Kazakhstan
Email: beisov.nurbol@gmail.com

1. INTRODUCTION

Sign language serves as the primary means of communication for individuals with hearing impairments. In order for a hearing person to communicate with the deaf, they typically require an interpreter who translates sign language into a natural language and vice versa. In the modern world, human-computer interaction (HCI) is becoming increasingly crucial as computers play a growing role in our lives. Researchers in the field of human-computer interaction strive to make this process more efficient, convenient, and natural by incorporating methods used in human communication. One of the most productive methods of human interaction is the use of hand gestures to convey ideas.

Currently, there is active development in technologies that enable the use of human gestures to control computer systems. A number of studies focus on hand gesture recognition in the context of computer interaction and its application in virtual reality [1]. One work explores key computer interaction technologies for creating interactive games using virtual reality (VR) technologies [2]. Other research proposes methods for hand gesture recognition, including optimal segmentation, real-time recognition systems using computer vision, and comparative analyses of challenges and solutions in this field [3]-[5]. The development of a hand gesture recognition system to enhance human-computer interaction is also considered [6]-[7]. The integration of deep learning, such as autoencoders and data preprocessing, is also an actively researched area [8]. These studies contribute to the advancement of hand gesture recognition technologies to improve human-computer interaction.

Zhang *et al.* [9], developed a dynamic gesture recognition system using a combination of machine learning algorithms and achieved high accuracy in recognizing dynamic gestures. “Hand gesture recognition for human-computer interaction” is a study that explores various methods of hand gesture recognition as an interface for human-computer interaction [10]. In their paper, “Hand gesture recognition for human-computer interaction,” the authors investigate the use of hand gestures to control computer systems. They explore various methods of hand gesture recognition, including camera and sensor-based methods, machine learning, and deep learning methods. The authors also discuss the application of hand gestures in various fields such as medicine, education, and entertainment, outlining the challenges and prospects of using this technology in the future. Their research indicates that hand gesture recognition can be an effective and convenient means of human-computer interaction, offering several possible avenues for further development of this technology [11].

In recent years, gesture recognition technologies have significantly thanks advanced due to the development of deep machine learning. Researchers are actively working to develop algorithms and neural networks that can recognize hand gestures with high accuracy. However, some challenges, such as adapting to different gesture styles and taking into account context, still remain the subject of research.

Sign language serves as the primary means of communication among individuals with hearing impairments. To facilitate communication between hearing individuals and the deaf, an interpreter is typically required to translate sign language into a natural language and vice versa. HCI is becoming increasingly important due to the growing influence of computers in our lives [12]. Researchers are striving to make HCI faster, simpler, and more natural. To achieve this goal, methods of HCI are being introduced in the field. One of the most rapidly advancing areas of human interaction is the use of gestures to express ideas.

In conclusion, the development of gesture control interfaces requires an interdisciplinary approach, taking into account a broad spectrum of considerations, including gesture recognition, interface design, privacy and security, accessibility, and platform integration. By approaching this technology comprehensively, developers can create gesture control interfaces that enhance user interaction, providing users with new and exciting ways to engage with technology. It is worth noting that the development of gesture control interfaces has the potential to revolutionize human-computer interaction, making it more natural and intuitive. However, it is crucial for developers to approach this technology holistically, with caution, considering the complex issues associated with its implementation. Also, the development of a gesture control interface represents an important step in the development of human-computer interaction. This allows users to interact with devices in a more natural and intuitive way, improving efficiency and opening up new possibilities for using technologies from fields as diverse as medicine, education, entertainment and many more. In the future, we can expect further development and improvement of gesture control technologies, leading to even more intuitive and user-friendly interfaces.

The work of Ionescu *et al.* [13] explores gestural control as a new type of intelligent interface between human and machine. They present their research in a collection of papers describing the application of computational intelligence in engineering and information technology. Elakkiya *et al.* [14] present the development of an intelligent system for human-computer interaction using hand gesture recognition. Huang *et al.* [15] propose a gesture-based system for next-generation, interface creation. The work of Zhao [16] discusses the study of human-computer interaction in the field of intelligent cars. Zhang *et al.* [17] present a study on exploring geology scenes in 3D space using somatic interaction based on hand tracking. The work of Jiang *et al.* [18] explores the design of gestural interaction for smart home control devices. Qi *et al.* [19] present a study on intelligent human-computer interaction, based gesture recognition using surface electromyography. This research contributes to the development of innovative methods of human-machine interaction and the creation of more efficient interfaces.

In the work by Zou and Cheng [20], a knowledge transfer model is presented for gesture recognition based on deep features extracted through convolutional neural networks (CNNs). Rautaray and Agrawal [21] present a review of video recordings based on hand gesture recognition for human-computer interaction. Trigueiros *et al.* [22] describe a general-purpose system using gestures applied to sign language recognition and refereeing robots in soccer. Dubey [23] introduces enhanced hand gesture recognition using an optimized probabilistic neural network based on the “swarm intelligence” optimization algorithm. Waldherr *et al.* [24] describe a gesture-based interface for human-robot interaction. In the work of Periverzov and Ilies [25], a system for generating software and hardware is considered. Elatawy *et al.* [26] uses neurosophic methods and fuzzy c-means to detect and recognise Arabic gesture alphabet. In the work by Al-Shamayleh *et al.* [27], a systematic review of vision-based gesture recognition techniques is conducted. Mirehi *et al.* [28] explore hand gesture recognition using topological features. These studies contribute to the development of gesture recognition methods and their application in the field of human-computer interaction.

The work by Sarma and Bhuyan [29] provides a review of methods, databases, and recent advancements in computer vision-based hand gesture recognition for human-computer interaction systems. Jiang *et al.* [30] describe an optimized real-time hand gesture recognition interface for people with spinal cord injuries. The work by Burger *et al.* [31] discusses recognizing two-handed gestures and combining them with spoken commands to control a robot. Pérez-Mayos *et al.* [32] explore part-of-speech and prosody-based approaches for synchronizing robot speech and gestures. Badi and Hussein [33] describe a technology for recognizing posture and hand gestures. The work of Kowdiki and Khaparde [34] presents an adaptive Hough transform with optimized deep learning accompanied by dynamic time warping for hand gesture recognition. These studies make a unique contribution to the development of hand gesture recognition methods and their application in human-computer interaction systems.

Like spoken language, sign language is not universal; it varies depending on the country or even regions. The Kazakh alphabet expands the possibilities of using the Cyrillic script and includes 42 letters, 9 of which represent specific sounds of the Kazakh language: "*Ә, І, Ө, Ұ, Ү, Ғ, Қ, Һ, Ң*". Here, we will focus on the pronunciation of only those specific sounds in the Kazakh language that have a distinct sound compared to Russian. For example, the sound, although coinciding in writing with the Russian, is pronounced quite differently in the Kazakh variant. In Kazakh, this sound is a back vowel, while in Russian, it is a mid-vowel. When pronounced, the lower jaw is raised, and the tongue is moved backward. It is pronounced harder and extremely briefly. The purpose of the article is to create a more intuitive HCI that is close to the human type of communication and eliminates misclassification, misunderstanding, ambiguity and contradictory interpretations that are often encountered when interpreting Kazakh sign language. This eliminates the characteristic problems of using traditional deep learning methods, leading to misinterpretation of important elements and construction of the Kazakh sign language.

2. METHOD

2.1. The effectiveness of the research approach

In the contemporary research domain, the concept of a computerized interpreter capable of interpreting hand gestures is gaining increasing attention. Within the field of HCI, there are two main classifications of gesture recognition systems. Those based on glove-based systems and those relying on visual perception.

The first type of systems operates with electromechanical devices that gather information about gestures. However, they suffer from a significant drawback users have to deal with uncomfortable and cumbersome devices, reducing the potential for interaction between humans and the system. Users are required to wear gloves with wires connected to various sensors. Then, based on the data collected by these sensors, the computer performs hand gesture recognition. To enhance the efficiency of interaction, an alternative category of HCI systems has been introduced, relying on visual perception.

The second category of vision-based systems involves the use of video cameras, image processing, and artificial intelligence for the recognition and interpretation of hand gestures. Within this category, two approaches stand out. The first entails the use of specially designed gloves with visual markers to determine the hand's position. However, this method does not always provide the required level of interaction, and the use of colored gloves can also increase the complexity of data processing. The second approach aims to achieve optimal comfort by utilizing images of the hands.

Thus, the best representatives from the group of deep learning methods and other modern sign language processing techniques have been considered, which are particularly effective when working with the English language. In this work, experimental data are presented, allowing for an improvement in the overall effectiveness of gesture recognition. In the Kazakh variant, both the static position of the hand and the dynamics of the process leading to the result are essential for the correct recognition of certain letters. Therefore, if the dynamic movement of the hand is not taken into account, errors are inevitable. Hence, we had to use our combination of recognition methods presented in this study.

As part of our research, an analysis of various sign languages was carried out, on the basis of which a program for recognizing gestures of the Kazakh language was developed. This distinguishes our approach from many other works that focus on studying individual sign languages and classifying the data used for gesture recognition. Such an approach brings several advantages, allowing for the generalization of learning and gesture recognition across multiple languages, taking into account variations in data between them. However, it is necessary to consider the unique features of the Kazakh sign language when using a database obtained from the analysis of other languages.

In the Kazakh sign language, there are unique gestures and symbols that are absent in American, Russian, and Indian sign languages. Additionally, there may be differences in the interpretation of gestures and their meanings across different cultures and languages. Therefore, despite using a common database, it is crucial to consider the specific characteristics of the Kazakh sign language for accurate and effective gesture recognition in this context. It may be necessary to conduct additional research and gather additional data for the Kazakh sign language to ensure high accuracy and efficiency in the gesture recognition system for this language. Furthermore, cultural and social aspects of sign language use should be taken into account, as they can vary significantly among different linguistic and cultural groups.

2.2. Computer vision motion detection

While the concept of computer vision has been around for a long time, it has now entered a special stage of development. There is a rapid expansion of public access through computer programs designed for various areas such as color image processing, computer graphics, computer geometry, and robotics. This means that computer vision can be used for a wide range of interdisciplinary research to solve a variety of everyday problems. Computers can non-contacting track facial expressions, hand movements, gestures, eye movements, and emotions. Motion detection allows the identification of specific actions based on computer vision technology and segmentation of the obtained video stream.

Software is capable of recognizing different body shapes. After installing the necessary software and hardware, the collected data is processed through a neural network. Other tools, such as Python, C++, C, Java programming languages, the OpenCV library, Tensorflow, Keras, are used for motion detection using computer vision. For our research, the most suitable programming environment for Python was chosen - open source computer vision library (OpenCV) an open-source library for computer vision and machine learning. OpenCV is designed to provide pattern recognition.

The implemented solution utilizes a single webcam and is based on a set of assumptions defined below:

- The user must be within a specific area in front of the camera.
- The user must be within a certain range due to camera limitations. The user should be at a specific distance. The system-defined values are 0.6 meters for the nearest plane and 1.3 meters for the far plane.
- Hand pose is determined with a bare hand and is not obscured by other objects.
- The system should be used indoors because the selected camera performs poorly due to ambient light noise.

2.3. Training gestures using CNNs

Most gesture recognition methods involve three primary stages. The first stage is the detection of the object representing the user’s hand. To achieve this, it is necessary to consider the environmental conditions and image-related issues to ensure accurate delineation of hand contours or regions, thereby enhancing recognition accuracy. The second stage is gesture recognition, where the selection of differentiated features and efficient classifiers is a primary challenge. The third stage involves the analysis of sequential gestures to determine user instructions or behavior. However, when the system operates in real-world conditions, it is challenging to control the environment and images. Therefore, cameras with better parameters and a suitable environment are necessary for the effective functioning of the system.

Sign language is not universal and can vary depending on the country or region. In the Kazakh world, sign language was recently recognized and documented, with many efforts made to establish standards and promote its use among the deaf and other interested groups. These efforts have led to the emergence of numerous sign languages, almost as many as there are countries, but they all use the same sign alphabets. Figure 1 shows the basic signs used in the alphabets of the Kazakh sign language. A CNN is a specialized multi-layer architecture of an artificial neural network used for classifying results. Various architectures, variations, and combinations of methods using CNN have been explored, with one of the latest being the CNN SA/Self Attention CNN.

First and foremost, algorithms using CNN are invariant to various distortions, such as camera rotation or uneven light distribution in the image, horizontal or vertical shifts, and other issues. Additionally, a CNN does not require a large amount of memory to store the features extracted during operation. Another advantage of using a CNN is its relatively fast learning speed, achieved by reducing the number of parameters used. Thus, the performance of a CNN surpasses by several times the performance of similar neural networks used in recognition tasks.

To capture hand histograms and create our first real-time object detector, it was necessary to use custom settings for the webcam. The main goal is to obtain an image that allows identifying the characteristics of action classifications that need to be processed. The action recognition algorithm is based on the use of CNN properties, which is trained using a “supervised learning” approach. Thus, choosing the most effective image is a factor influencing the future performance and learning of the accumulated

experience of the neural network. One of the main characteristics of a digital image is its color. Based on this feature, an image can be classified into the following categories:
- Binary;
- Half-tone;
- Color.

Figure 1. Alphabet of Kazakh sign language

In binary images, there are only two states for pixels; specific colors are considered in half-tone images, while multicolored ones are in color images. The most efficient type of video that is easy to edit is binary video. When training gestures, the obtained histogram is treated as a graphical representation of the image, carrying information only about color intensity, where the pixel value ranges from 0 to 255. Images of this type, also known as grayscale, consist of shades of gray ranging from black at the weakest intensity to white at the strongest, where a pixel can be considered as an individual point in a binary image.

Working with binary images presents a challenge related to its brightness. This is because the color of an object changes due to the influence of light, and part of the information is irretrievably lost. For example, there is an issue with the incomplete recognition of the hand. During teaching gestures, it is necessary to transform the color space and resize the image. It is necessary to convert the color space of the image from red, green, blue (RGB) to hue, saturation, value (HSV), while the image size is transformed from 640 * 480 to 50x50. For training the gesture histograms of the Kazakh alphabet, we have data in the library for gestures for each of the 42 letters and additionally two gestures: space and delete. These data use 105,600 images for 44 different gestures. When training gestures, image rotation around the X-axis is applied. The obtained histograms enable the recognition of gestures from both hands. Figure 2 shows the resulting mirror-symmetric image of histograms during deep learning.

Figure 2. Obtained histograms during gesture learning

Input layer: the INPUT layer contains raw pixel values of the image, in this case, an image with a width of 50, a height of 50 pixels, and three HSV color channels. Passing through 3×3 masks results in an output of 49×49 pixels. Convolutional layer this layer is used to extract various features from input images. It performs the mathematical convolution operation between the input image and a filter of a specific size M×M. Moving the filter over the input image results in the scalar product between the filter and parts of the input image relative to the filter size.

Aggregation level: the main purpose of this layer is to reduce the size of the folded feature map to reduce the computational cost. This is achieved by reducing the number of connections between layers and with each feature map. Depending on the method used, there are several types of pooling operations. In max pooling, the largest element is taken from each feature map. In the second variant, average pooling calculates the average value of elements. The total sum of elements is computed in sum pooling. The pooling layer typically serves as a bridge between the convolutional layer and the fully connected (FC) layer. Classification in the fully connected layer. The FC layer is usually placed before the output. In this case, the input image from the previous layers is flattened and fed into the FC layer. The flattened vector then passes through several FC layers. At this stage, the classification process begins. Output layer: the overall architecture of hidden layers is shown in Figure 3.

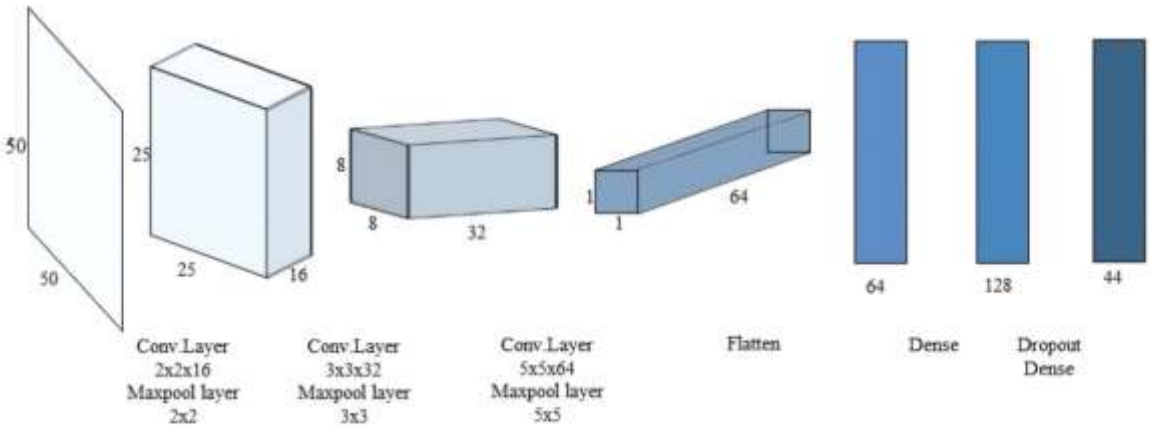


Figure 3. Architecture of hidden layers of CNN neural network

This CNN architecture improves hand movement recognition with its hidden layers. It uses a combination of convolution layers and association layers to extract relevant features from images, and then fully connected layers for prediction. This dataset is commonly used to train and evaluate the performance of CNN models in image classification tasks.

3. RESULTS AND DISCUSSION

3.1. Study performance evaluation

To evaluate the performance of the study, we performed a classification of images that depict sequences of letters, such as "*Ә, Ң, Қ, Ө, Ғ, Ү, Ұ*". In our studies, we limited the alphabet used by the system to a set of 44 letters, and the system only worked with these letters when classifying video materials. To account for overlap between training data, each video was classified using a limited training set that only included letters present in that particular word. For example, in Figure 4, when classifying the letters "*Ә, Ө, Ғ, Ү, Ұ*", the system successfully identified the classes "*ә, ө*". The thresholding method compares frame similarity with a specified threshold. If the similarity between frames is below the threshold, the frame boundary is considered found. The threshold can be global, adaptive, or mixed. The global threshold is experimentally chosen for the entire video. Local features are not taken into account, which negatively affects detection accuracy. Adaptive threshold algorithms based on a sliding window computer a local threshold. The sliding window primarily remembers previous points in the time series and returns them as a sequence of a specified size along with the current point.

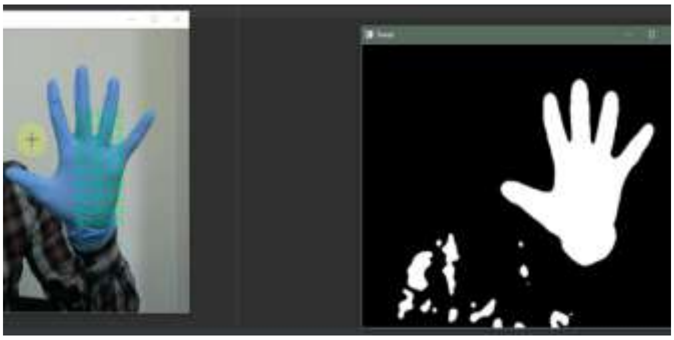


Figure 4. Image classification definition

To apply adaptive thresholding, prior knowledge of the video is necessary, such as the choice of window size. Combining adaptive and global thresholds allows creating local thresholds when common parameter values are present. Global thresholds are adjusted based on the requirements for accuracy and completeness. The use of the thresholding method during the training of similar letters in the Kazakh alphabet through a neural network contributed to satisfactory results. This indicates that our system correctly recognizes each letter during its identification. For example, the 0 symbol was correctly identified a total of 250 times out of 204 identifications, with 2 instances of confusion with the 8th symbol and the rest going unrecognized.

This task has limitations related to the visibility quality in the camera, lighting in the recognition zone, limited computational device resources, and other similar constraints. The advantage of this research is the high accuracy in recognizing hand gestures, which is critically important for use in human-computer interaction systems. Graphs were constructed based on the obtained results as shown in Figure 5.

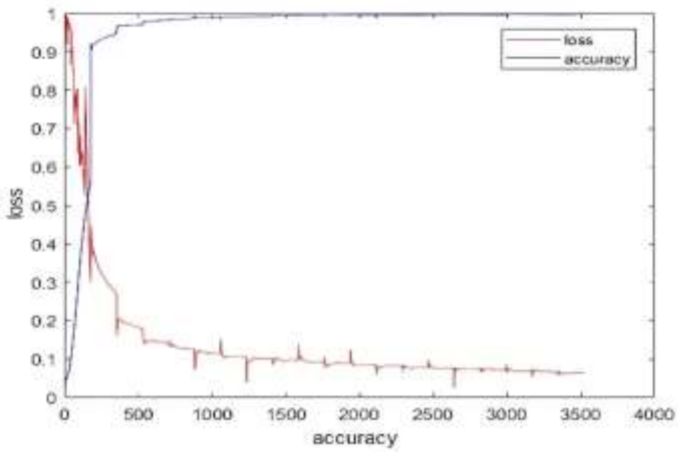


Figure 5. Epoch-test learning graph

There are *X* axes indicate the numbers of training epochs, and the *Y* axis indicates the accuracy of the model at the corresponding stages. The blue curve on the graph represents the model's accuracy on the training data, showing how well the model predicts results on this data as the number of training epochs increases. The orange curve on the graph represents the model's accuracy on validation data, i.e., on data that the model hasn't seen during training. It illustrates how well the model can generalize results to new data and helps researchers assess overfitting. The training accuracy graph will look like two curves that rise together at the beginning of training and then stabilize at a certain level of accuracy. However, various scenarios are possible, including overfitting and underfitting, which are easily identified by the shape, and position of the curves on the graph.

3.2. Virtual assistant of computer finger-spelling

Currently, with the developed program, a user proficient in Kazakh sign language can communicate with the computer using gestures. The computer perceives the gesture, processes it instantly, and displays the result on the screen. In this program, the accuracy of gesture recognition is around 97%. Based on this development, we have created a virtual assistant known as the Kazakh computer gesture typing language designed for the hearing impaired. Gesture typing is accompanied by articulation, and the deaf interlocutor sees themselves, the gestures, and the text on the monitor. Figure 6 shows words translated from gestures to text KAZAKH (ҚАЗАҚ).

Figure 6. Compiling words using Kazakh sign language

It should be taken into account that in sign language one gesture means one letter or one whole word, and dactylology is a unique form of speech where the dactyl alphabet is used, and each hand gesture illustrates a specific letter of this language. When using the Kazakh computer fingerprint sign language, we set ourselves the task quite accurately and in the simplest form. Based on this, rules for computer fingerprinting were developed. Rules for computer typing:

- The typing hand is bent at the elbow, slightly extended forward for webcam alignment, and placed in the green field to detect skin color. Typing can be done with either the right or left hand, but it is commonly done with the right hand as reading from the left hand can be less comfortable;
- Typing is accompanied by articulation, enhancing the perception and understanding of the text. The visually impaired interlocutor perceives themselves and the typed text on the monitor;
- During typing, the hand is facing the camera with the palm. The hand itself remains still, and only the wrist moves. Abrupt hand movements up and down or forward and backward significantly complicate reading, requiring camera refocusing;
- Typing follows the norms of orthography. Punctuation marks are not displayed, except when dictating text for recording;
- Doubled consonants in typing are represented by shifting the doubled consonant to the right or left;
- Typing is done smoothly and seamlessly. Words are separated by the “space” gesture. After completing the message, the hand should be lowered;

At this stage, the Kazakh sign language is divided into three categories based on the direction of the wrist. To determine the category of each letter, the system checks the pixel with the maximum value in each selected frame and checks the pixel with zero value. If neither of these situations occurs, the letter is classified into the category with the wrist image in the middle of the frame. This improves the accuracy of gesture recognition and reduces processing time by minimizing matching operations in subsequent stages.

The interface we developed for intelligent gesture control in human-computer interaction is applied in the field of education to assist inclusive individuals in learning sign language. In virtual learning, priority attention was given to the Kazakh sign language. The project utilized the Python language and libraries such as TensorFlow and OpenCV. Additionally, methods and procedures for image enhancement were considered as crucial stages in solving image processing problems. Procedures for image segmentation, dividing the image into specific objects, and informative image segmentation were also examined. For gesture recognition, a regular camera, webcam, or a standard camera with a distance of at least 1 meter is used. This project is employed in the education of individuals with hearing impairments. The general algorithm is divided into four stages:

- Gesture detection;
- Segmentation;
- Feature extraction;
- Classification.

For the Kazakh language, there are specific features related to orthographic gestures and other language traits. Additionally, one challenge in categorizing gestures is the absence of static gestures; there is no need for specific hand movements or positioning fingers in space. Throughout the considered period, the output is generated through hand movements. Most real-time gesture recognition systems typically employ only one of the proposed forms.

A distinctive feature of Kazakh sign language is the presence of similar symbols, such as "*Г, Ғ, К, Қ, Н, Ң, О, Ө, У, Ұ*". The gesture for these letters involves both upward and downward movements, along with specific degrees of rotation. The gestures for these symbols are depicted in Figure 7. Gestures of similar two letters to as shown in Figure 8. The Figures 8(a) and (b) shows the gestures of the above-mentioned letters of the Kazakh alphabet.

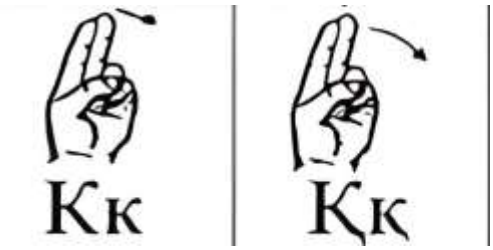
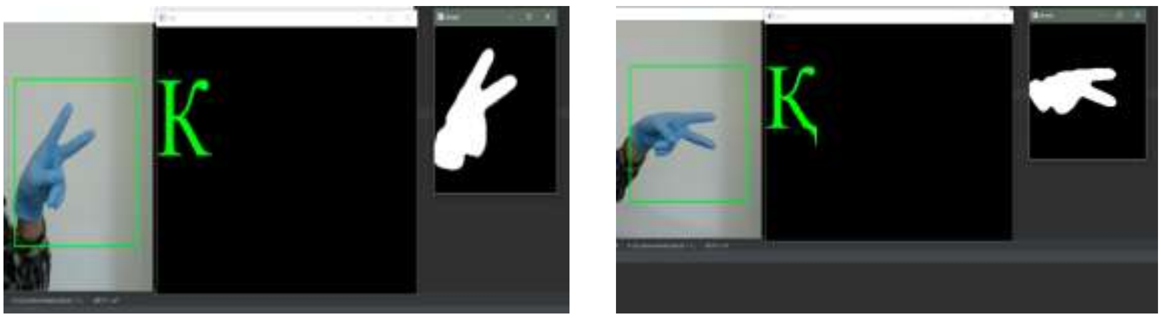


Figure 7. Similar two letters



(a) (b)

Figure 8. Gestures of similar two letters to (a) demonstration of statistical gesture letters and (b) demonstration of dynamic gesture letters

During the development of the Kazakh sign language, a database of 47 classes was compiled, with 1,500 images for each class represented in histogram form. In deep learning, a CNN neural network was applied, which can successfully recognize gestures anywhere within an image according to a possible scenario, for example, when the hand performing the gesture takes the place of a smaller image compared to a larger image of a frozen shape of an object, including the upper body. Presently, with the developed program, a user proficient in Kazakh sign language can communicate with a computer using gestures. The computer perceives the gesture, processes it instantly, and displays the result on the screen. In this program, the accuracy of gesture recognition is approximately 97%.

The created product can assist in virtual sign language learning. The trained artificial intelligence fully recognizes the 42 letters of the Kazakh alphabet. Additionally, other symbols like "space", "back", and others have been trained. These symbols are primarily used in constructing sentences through gestures. Training occurs through a CNN. The sign language is accompanied by articulation, allowing a non-hearing interlocutor to see not only themselves and the sign language but also the corresponding text on the monitor Figure 9.

It is important to note that in sign language, each gesture represents a whole word, whereas dactylology is a distinctive form of communication utilizing a tactile alphabet, where each hand gesture illustrates a specific letter of that language. Every natural language may have its own unique dactylology, which also differs from other languages. Accordingly, the Kazakh language has its own dactylology.

Developing methods and systems for recognizing Kazakh dactylological sign language poses several challenges, primarily linked to orthographic, gestural, and other language-specific features. Another issue involves categorizing gestures into static ones, where no hand movements are necessary, and dynamic ones, reproduced through hand movements. In most cases, real-time gesture recognition systems focus on either static or dynamic data in their databases, employing a single form for comparison.

Figure 9. Compiling words using Kazakh sign language

We applied a method that enhances images with special attention to dark areas. This method is applicable to both grayscale and color images. For color images, a color space transformation is necessary, as direct processing in the RGB color space may alter the original colors. Our goal was to develop more efficient and user-friendly interfaces for human-computer interaction, as gestures provide a natural and powerful means of communication. Additionally, it can be utilized for teleconferencing, as it doesn’t require special equipment, and can be applied in translation and sign language learning.

However, there are challenges in the development and usage of gesture control technology. For instance, detecting and interpreting complex or subtle gestures can be difficult, especially in real-world conditions with varying lighting and background noise. Creating standardized gestures that work across different applications and devices, as well as ensuring conditions for users to memorize and consistently perform gestures, can also be challenging. In general, gesture control technology has the potential to provide new and exciting ways to interact with computers and devices, but it is crucial to carefully design and test these interfaces to ensure their effectiveness and intuitiveness.

The program is launched in the PyCharm Community programming environment. To work with this program, the initial step involves setting the histogram of the hand image as shown in Figure 10. Here, we will check the required color for the proper functioning of the program by pressing the “*c*” key on the keyboard. If the accuracy of the hand image histogram meets your satisfaction, you can save it by pressing the “*s*” button. The hand histogram is implemented using the cv2.calcHist() method from the OpenCV library. After this, we launch and can check the operation of the general system and see the recognition process as shown in Figure 11.

Figure 10. Hand histogram adjustment

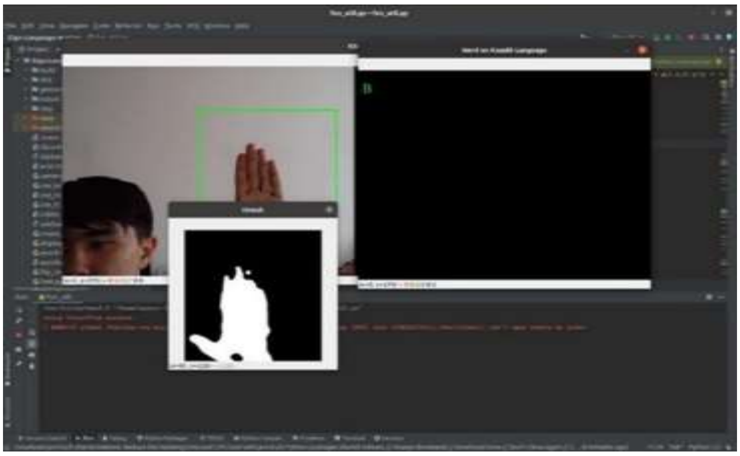


Figure 11. Gesture recognition process

The main network of the system is based on convolutional neural network (ConvNet). The peculiarity of ConvNet is that the color feature of the hand does not affect the system classifier when using a binary image. The images are pre-processed before being sent to the CNN classifier. That is, camera video is primarily processed by OpenCV, a library of computer vision, image processing, and general-purpose digital algorithms, including cv2.GaussianBlur(), a function that uses a Gaussian kernel to remove Gaussian noise from an image. The Gaussian filter coefficients used in noise removal are found using the following formula:

$$G_i = \alpha * e^{\frac{-(i-\frac{k-1}{2})^2}{2*\sigma^2}} \quad (1)$$

where, $i = 0 \ldots k - 1$, α – scaling factor, σ – Gaussian standard deviation, k – aperture size, must be an odd number and positive.

We utilize cv2.medianBlur(), a function that finds the median of all pixels in the image, to eliminate “salt and pepper” noise. Gesture recognition begins after the mentioned processing operations. Hands are placed inside a specific square area of the camera, considered the detector zone, where letter movements are demonstrated. The size of the detector zone matches the size of images in the training dataset used in the system. The image within the detector zone, obtained from the camera, is sent separately to the ConvNet classifier.

The result of hand movement recognition is displayed in the “Kazakh text” window. The hand histogram property is shown in the “thresh” window. The recognition result of sign language is displayed on the screen only if the recognition probability is above 80%.

To test the developed intelligent gesture control interface and video resources, remote testing was performed using our created flexible online platform and traditional training methods. We successfully coordinated and conducted an educational experiment to study the learning outcomes of the dactyl language developed by us, based on the interaction with computers, for hearing-impaired children at the College of Technology and Service in Almaty, Kazakhstan. The basis for choosing this college was that this is an inclusive specialized educational institution, first of all, this college is an “accessible environment”, which includes not only the architectural features of the educational institution, but also various special educational tools that allow you to adapt the education system taking into account special needs of students with disabilities and disabilities as shown in Figure 12. Additionally, it is an inclusive educational environment.

Thus, the developed computer dactyl sign language allows you to quickly and efficiently master the sign language of the Kazakh alphabet with the help of a virtual assistant. Knowledge of sign language expands communication skills and promotes wider social interaction. Gesture recognition technologies can support creative endeavors and improve human-computer interaction. However, less attention has been paid to recent research in HCI has made significant advances in hand gesture recognition, addressing the challenges in different environments. Negi *et al.* [35] presented vision-based real-time human-computer interaction through hand gesture recognition, addressing the practical aspects of this technology.

One of the main benefits of gesture control technology is that it provides a more natural and intuitive way to interact with computers and devices, as users can perform actions in a physical and tactile manner. Many researchers have developed various sign language programs from different countries that provide training using artificial intelligence. Zhang [36] explored the application of AI-based real-time gesture recognition and embedded systems in English major teaching, showcasing the diverse applications of gesture recognition technology. Gupta *et al.* [37] innovatively used edge computing systems with vision transformers and lightweight CNN for hand gesture recognition, addressing the challenges associated with real-time processing in HCI applications. The results of our study enable Kazakh sign language users to communicate with the computer.

Figure 12. Fragment of work with hearing-impaired students at the college of technology and service during a computer science class

4. CONCLUSION

Digitalization and technology for developing interfaces for sign language is an important research problem. Interface development technology is interdisciplinary in nature, including various approaches to the issues of gesture recognition, interface design, privacy and security, and accessibility. Gesture control interfaces contribute to more effective user interactions. Gesture recognition technology represents a new dimension in the human-computer interaction interface, with both advantages and limitations. Among the advantages are the naturalness and intuitiveness of interaction, improved usability, and enhanced functionality for individuals with limited abilities. We have analyzed various programs and applications designed for learning and translating sign language. However, gesture recognition technology may be less precise than other input methods and sometimes requires additional calibration to improve accuracy. Nevertheless, gesture recognition technology holds significant potential for further development and application in various fields such as gaming, education, healthcare, and industry. New machine learning and artificial intelligence methods can help enhance the accuracy and efficiency of gesture recognition. The presented development of an intelligent gesture control interface aims to enhance user experience and improve the efficiency of computer interaction. The proposed gesture recognition system based on artificial intelligence using CNN consists of three stages: pre-processing, selecting the best frame, and determining the gesture category. The designed dactyl language allows for a quick and high-quality mastery of the Kazakh sign language alphabet with the assistance of a virtual assistant. Proficiency in sign language expands communicative skills and promotes broader social interaction. The digital application of the dactyl language will significantly improve communication for individuals with hearing or speech impairments, making it more effective and impactful. Such a system eliminates the need for an interpreter.

In general, gesture recognition technology represents a significant technological innovation that can enhance human-computer interaction and has great potential for development and application in various fields. Each language requires the development of its own digital sign language translation. The presented interface can serve as an example for creating new programs and applications for the development of the Kazakh sign language.

ACKNOWLEDGEMENTS

This research has been funded by of the project number by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No AP19676457 and AP09259370).

REFERENCES

[1] S. Gupta, S. Bagga, and D. K. Sharma, "Hand gesture recognition for human computer interaction and its applications in virtual reality," in *Studies in Computational Intelligence*, vol. 875, Springer International Publishing, 2020, pp. 85–105.

[2] P. Gao, "Key technologies of human–computer interaction for immersive somatosensory interactive games using VR technology," *Soft Computing*, vol. 26, no. 20, pp. 10947–10956, Jun. 2022, doi: 10.1007/s00500-022-07240-3.

[3] M. A. Rahim, A. S. M. Miah, A. Sayeed, and J. Shin, "Hand gesture recognition based on optimal segmentation in human-computer interaction," in *Proceedings of the 3rd IEEE International Conference on Knowledge Innovation and Invention 2020, ICKII 2020*, Aug. 2020, pp. 163–166, doi: 10.1109/ICKII50300.2020.9318870.

[4] N. Kerimbayev, N. Beisov, A. Kovtun, N. Nurym, and A. Akramova, "Robotics in the international educational space: Integration and the experience," *Education and Information Technologies*, vol. 25, no. 6, pp. 5835–5851, 2020, doi: 10.1007/s10639-020-10257-6.

[5] J. John and S. P. Deshpande, "A comparative study on challenges and solutions on hand gesture recognition," in *Lecture Notes in Electrical Engineering*, vol. 984 LNEE, Springer Nature Singapore, 2023, pp. 229–240.

[6] T. H. Tsai, C. C. Huang, and K. L. Zhang, "Design of hand gesture recognition system for human-computer interaction," *Multimedia Tools and Applications*, vol. 79, no. 9–10, pp. 5989–6007, Dec. 2020, doi: 10.1007/s11042-019-08274-w.

[7] S. Paul, A. Bhattacharyya, A. F. Mollah, S. Basu, and M. Nasipuri, "Hand segmentation from complex background for gesture recognition," in *Advances in Intelligent Systems and Computing*, vol. 937, Springer Singapore, 2020, pp. 775–782.

[8] B. N. Subudhi, T. Veerakumar, S. R. Harathas, R. Prabhudesai, V. Kuppili, and V. Jakhetiya, "Deep learning in autoencoder framework and shape prior for hand gesture recognition," in *EAI/Springer Innovations in Communication and Computing*, Springer International Publishing, 2023, pp. 223–242.

[9] W. Zhang, Z. Shao, C. Wei, D. Wang, and J. Tu, "Design and implementation of simulation maintenance training system for a certain type of command and control system," in *Proceedings - 2010 International Conference on Optoelectronics and Image Processing, ICOIP 2010*, Nov. 2010, vol. 1, pp. 389–393, doi: 10.1109/ICOIP.2010.62.

[10] M. Panwar and P. Singh Mehra, "Hand gesture recognition for human computer interaction," Nov. 2011, doi: 10.1109/ICIIP.2011.6108940.

[11] A. Haria, A. Subramanian, N. Asokkumar, S. Poddar, and J. S. Nayak, "Hand gesture recognition for human computer interaction," *Procedia Computer Science*, vol. 115, pp. 367–374, 2017, doi: 10.1016/j.procs.2017.09.092.

[12] N. Yu et al., "Design and implementation of a dexterous human-robot interaction system based on haptic shared control (in Chinese)," *Chinese Journal of Scientific Instrument*, vol. 38, no. 3, pp. 602–611, 2017.

[13] D. Ionescu, B. Ionescu, C. Gadea, and S. Islam, "Gesture control: a new and intelligent man-machine interface," in *Topics in Intelligent Engineering and Informatics*, Springer Berlin Heidelberg, 2012, pp. 331–354.

[14] R. Elakkiya, K. Selvamani, S. Kanimozhi, A. Kannan, and V. Rao, "Intelligent system for human computer interface using hand gesture recognition," *Procedia Engineering*, vol. 38, pp. 3180–3191, 2012, doi: 10.1016/j.proeng.2012.06.369.

[15] J. Huang, P. Jaiswal, and R. Rai, "Gesture-based system for next generation natural and intuitive interfaces," *Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM*, vol. 33, no. 1, pp. 54–68, May 2019, doi: 10.1017/S0890060418000045.

[16] S. Zhao, "Research on human-computer interaction on the field of intelligent vehicles," in *Advances in Transdisciplinary Engineering*, vol. 33, IOS Press, 2022, pp. 240–245.

[17] W. Zhang, F. Zhu, P. Lu, P. Li, B. Sheng, and L. Mao, "3D geology scene exploring base on hand-track somatic interaction," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12221 LNCS, Springer International Publishing, 2020, pp. 364–373.

[18] B. Jiang, X. Wang, and Y. Wu, "Research on gesture interaction design for home control intelligent terminals," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12182 LNCS, Springer International Publishing, 2020, pp. 38–56.

[19] J. Qi, G. Jiang, G. Li, Y. Sun, and B. Tao, "Intelligent human-computer interaction based on surface EMG gesture recognition," *IEEE Access*, vol. 7, pp. 61378–61387, 2019, doi: 10.1109/ACCESS.2019.2914728.

[20] Y. Zou and L. Cheng, "A transfer learning model for gesture recognition based on the deep features extracted by CNN," *IEEE Transactions on Artificial Intelligence*, vol. 2, no. 5, pp. 447–458, Oct. 2021, doi: 10.1109/TAI.2021.3098253.

[21] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: a survey," *Artificial Intelligence Review*, vol. 43, no. 1, pp. 1–54, Nov. 2015, doi: 10.1007/s10462-012-9356-9.

[22] P. Trigueiros, F. Ribeiro, and L. P. Reis, "Generic system for human-computer gesture interaction: applications on sign language recognition and robotic soccer refereeing," *Journal of Intelligent and Robotic Systems: Theory and Applications*, vol. 80, no. 3–4, pp. 573–594, Feb. 2015, doi: 10.1007/s10846-015-0192-4.

[23] A. K. Dubey, "Enhanced hand-gesture recognition by improved beetle swarm optimized probabilistic neural network for human–computer interaction," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 9, pp. 12035–12048, Feb. 2023, doi: 10.1007/s12652-022-03753-9.

[24] S. Waldherr, R. Romero, and S. Thrun, "Gesture based interface for human-robot interaction," *Autonomous Robots*, vol. 9, no. 2, pp. 151–173, 2000, doi: 10.1023/A:1008918401478.

[25] F. Periverzov and H. T. Ilieş, "3D imaging for hand gesture recognition: Exploring the software-hardware interaction of current technologies," *3D Research*, vol. 3, no. 3, pp. 1–15, Jun. 2012, doi: 10.1007/3DRes.03(2012)1.

[26] S. M. Elatawy, D. M. Hawa, A. A. Ewees, and A. M. Saad, "Recognition system for alphabet Arabic sign language using neutrosophic and fuzzy c-means," *Education and Information Technologies*, vol. 25, no. 6, pp. 5601–5616, Jun. 2020, doi: 10.1007/s10639-020-10184-6.

[27] A. S. Al-Shamayleh, R. Ahmad, M. A. M. Abushariah, K. A. Alam, and N. Jomhari, "A systematic literature review on vision based gesture recognition techniques," *Multimedia Tools and Applications*, vol. 77, no. 21, pp. 28121–28184, 2018, doi: 10.1007/s11042-018-5971-z.

[28] N. Mirehi, M. Tahmasbi, and A. T. Targhi, "Hand gesture recognition using topological features," *Multimedia Tools and Applications*, vol. 78, no. 10, pp. 13361–13386, Mar. 2019, doi: 10.1007/s11042-019-7269-1.

[29] D. Sarma and M. K. Bhuyan, "Methods, databases and recent advancement of vision-based hand gesture recognition for HCI systems: a review," *SN Computer Science*, vol. 2, no. 6, Aug. 2021, doi: 10.1007/s42979-021-00827-x.

[30] H. Jiang, J. P. Wachs, and B. S. Duerstock, "An optimized real-time hands gesture recognition based interface for individuals with upper-level spinal cord injuries," *Journal of Real-Time Image Processing*, vol. 11, no. 2, pp. 301–314, Jun. 2016, doi: 10.1007/s11554-013-0352-3.

[31] B. Burger, I. Ferrané, F. Lerasle, and G. Infantes, "Two-handed gesture recognition and fusion with speech to command a robot," *Autonomous Robots*, vol. 32, no. 2, pp. 129–147, Dec. 2012, doi: 10.1007/s10514-011-9263-y.

[32] L. Pérez-Mayos, M. Farrús, and J. Adell, "Part-of-speech and prosody-based approaches for robot speech and gesture synchronization," *Journal of Intelligent and Robotic Systems: Theory and Applications*, vol. 99, no. 2, pp. 277–287, Nov. 2020, doi: 10.1007/s10846-019-01100-3.

[33] H. S. Badi and S. Hussein, "Hand posture and gesture recognition technology," *Neural Computing and Applications*, vol. 25, no. 3–4, pp. 871–878, Apr. 2014, doi: 10.1007/s00521-014-1574-4.

[34] M. Kowdiki and A. Khaparde, "Adaptive hough transform with optimized deep learning followed by dynamic time warping for hand gesture recognition," *Multimedia Tools and Applications*, vol. 81, no. 2, pp. 2095–2126, Oct. 2022, doi: 10.1007/s11042-021-11469-9.

[35] P. S. Negi, R. Pawar, and R. Lal, "Vision-based real-time human–computer interaction on hand gesture recognition," in *Lecture Notes in Networks and Systems*, vol. 106, Springer Singapore, 2020, pp. 499–507.

[36] T. Zhang, "Application of AI-based real-time gesture recognition and embedded system in the design of english major teaching," *Wireless Networks*, Jul. 2021, doi: 10.1007/s11276-021-02693-0.

[37] K. Gupta, A. Singh, S. R. Yeduri, M. B. Srinivas, and L. R. Cenkeramaddi, "Hand gestures recognition using edge computing system based on vision transformer and lightweight CNN," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 3, pp. 2601–2615, Dec. 2023, doi: 10.1007/s12652-022-04506-4.

BIOGRAPHIES OF AUTHORS

Nurbol Beisov is senior lecturer at the Department of Computer Science, al-Farabi Kazakh National university. He has more than 10 scientific papers, including 2 works in the ranking publications Robotics in the international educational space: Integration and the experience, The Use of Mobile Technologies in Education with an Emphasis on a Student-Centered Approach. h- index- 2. Was the executor of the following projects: AP09259370 "Development of a technological platform for virtual learning based on artificial intelligence approaches". AP19676457 "Modeling and feedback control in educational telematics" Research Institute of Mathematics and Mechanics KazNU named after. al-Farabi. Erasmus+ Professionalization of bachelor's and master's degrees in the field of strategic management and risk management in healthcare within the framework of open and distance learning in Russia, Kazakhstan and Azerbaijan - LMQS No. 618860-EPP-1-EL-EPPKA2-CBHE-JP. Research interests: information technology in education, artificial intelligence, semantic analysis. He can be contacted at email: beisov.nurbol@gmail.com.

Gulnar Madyarova Candidate of Pedagogical Sciences, Associate Professor of Al-Farabi Kazakh National University. She has more than 10 scientific papers, including 5 papers in the Web of Science and Scopus rating publications, h- index- 1. Research interests: informatization of education, computer science. Department of Computer Science, Faculty of Information Technology. Almaty, Kazakhstan. She can be contacted at email: madyarova-gulnar@mail.ru.

Nurassyl Kerimbayev Doctor of Pedagogical Sciences, Professor at the Department of Computer Science of Al-Farabi Kazakh National University. He is the leader of the following projects: AP09259370 "Development of a technological platform for virtual learning based on artificial intelligence approaches" and AP19676457 "Modeling and feedback control in educational telematics" of the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan. He are the head of the scientific school "Informatics and Informatization in Education". Research interests: informatization of education, computer science, artificial intelligence, robotics. He can be contacted at email: n_nurassyl@mail.ru.