

# Methodological Foundations of an Information System Construction for the Recognition of Ukrainian Sign Language

Taras Basyuk<sup>1</sup>, Andrii Vasyliuk<sup>1</sup>

<sup>1</sup>Lviv Polytechnic National University, Bandera str.12, Lviv, 79013, Ukraine

## Abstract

The article analyzes existing methods and known systems that provide means of recognizing Ukrainian sign language and describes the mechanisms of their implementation. Technologies and software tools for sign language recognition were analyzed, which made it possible to identify the main shortcomings of existing approaches and showed the relevance of the research. The diagram reflecting the main stages that must be implemented in the process of gesture recognition has been finalized. The structural design of the software system was carried out with the display of created diagrams in accordance with the IDEF0 standard. The article presents a context diagram and a decomposition diagram, which created the basis for the study of features and the formation of methodological foundations for the construction of an information system. The main stages of gesture recognition are highlighted and described, namely: transformation of the input image, its filtering and actual recognition. The justification of the choice of methods for displaying contours and recognizing gestures in the incoming information message was made and their analysis was carried out. The constructed prototype of the system for recognizing Ukrainian sign language consists of four main modules: HandGesturesRecognitionForm, NeuralNetwork, CsvManager, TrainingImageDataManager, which provide basic functionality. At the current stage, it can be useful as an additional communication tool for people with special needs. Further research will be aimed at testing and improving systems, eliminating conflicts and expanding functionality in accordance with the specified requirements.

## Keywords

Ukrainian sign language, pattern recognition, learning process, communication, information system

## 1. Introduction

Today, computer technologies are involved in almost all spheres of human life. With the help of various technical solutions, a person is able to solve daily tasks with greater simplicity and efficiency. If at the end of the 20th century computer technologies were primarily associated with the scientific and military spheres, then in the second decade of the 21st century they are associated with almost all spheres of human life. It is quite natural that various computer solutions are widely used in the field of communication between individuals with special needs. At the same time, one of the communication devices is sign language [1]. Sign language is a type of speech that makes it possible to express thoughts using facial expressions, emotions, and hand gestures that correspond to letters, words, or individual phrases. Despite the large number of people who suffer from hearing or speech impairments, sign language has received little attention from linguistics. In the world, the share of people with hearing problems is about 5% or 430 million [2] of the total population. Sign languages are not universal in all countries, as they arise and develop naturally in different territories and change over time with the emergence of new vocabulary. The debate about sign language has been going on for about half a century. Until recently, the attitude towards it in different countries ranged from introducing it for learning in Paragraph text. Paragraph text. Paragraph text. Paragraph text. Paragraph text. Paragraph text.

---

COLINS-2024: 8th International Conference on Computational Linguistics and Intelligent Systems, April 12–13, 2024, Lviv, Ukraine

✉ Taras.M.Basyuk@lpnu.ua (T. Basyuk); Andrii.S.Vasyliuk@lpnu.ua (A. Vasyliuk)

ORCID 0000-0003-0813-0785 (T. Basyuk); 0000-0002-3666-7232 (A. Vasyliuk)



© 2024 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

educational institutions for children with hearing impairments to ignoring the existence of the language and even completely banning it [3].

The hearing-impaired community is often embarrassed by their difficulties in communicating with the rest of the world. Although sign language is used as a means of conveying its message, there are still problems in communication because there are few people who are familiar with this type of language. In addition, the number of available translators is insufficient to solve the problem. This motivated scientists from different countries to study this problem and work on it. In general, this issue can be divided into two parts: the first involves the development of automatic sign language synthesizers that allow people with hearing impairments to understand messages transmitted by people who can hear; the second part - the opposite - concerns the development of sign language interpreters that allow the hearing community to understand sign language [4]. In view of the mentioned patterns, the urgent task is to develop an information system for recognizing Ukrainian sign language, which will provide additional means of overcoming the language barrier between communication subjects.

## **1.1. Analysis of recent researches and publications**

### **1.1.1. Analysis of known research**

Although sign language has been a subject of study for centuries, it was only at the end of the last century that the subject became the focus of linguistic research. This was facilitated by the publication of William K. Stoke "The Structure of Sign Language", which marked the beginning of the linguistics of sign language. The proposed structure consisted of 55 symbols, which formed three groups according to the parameters (place of execution of the gesture, nature of movement and shape of the hand), which Stoke considered relevant in determining the structure of the gesture. Stoke's notation formed the basis of the organizing principles of the first dictionary of American Sign Language [5].

As the analysis showed, the existing methods of gesture recognition in computer systems are divided into two types - recognition based on the creation of a 3D model and methods built on the principle of feature selection [6,7]. The first class of methods is based on the creation of a kinematic model. This model must take into account each of the possible degrees of freedom. When building such a model, hand gestures are evaluated using a comparison of hand coordinates on the image. Methods of this type make it possible to recognize a significant number of gestures, but when implementing them, you need to create a large-scale database with images. Images from the database will also be used to resolve conflicts during feature selection that arise due to various shapes and sizes of recognition objects. The second class of methods is based on the processing of details of the input data stream, which are designed to determine the coordinates of the object of recognition. This method can be applied only if it is possible to determine characteristic anchor points or features on the images of objects. Then the object itself can be defined as a combination of these points or planes that they form. In this case, instead of creating a complete object, a subset of its characteristic points or areas is created. This approach is resistant to deformations and changes in input sequences. In the presence of characteristic features, the object can always be unambiguously classified [8]. A separate approach to gesture recognition is a method based on artificial neural networks [9]. Convolutional neural networks can successfully identify individual dactyls, but this applies only to static gestures, the analysis of dynamic movements based on images is too cumbersome and resource-consuming [10].

In general, scientific research in this direction can be presented in the form of the following publications:

- "Sign language recognition using Microsoft Kinect" [11] authors developed a method for recognizing sign language using depth images from the Kinect sensor. The depth and motion profile are calculated from the generated images and used to construct a feature matrix for each gesture. Recognition is performed on the basis of a linear classifier based on the method of support vectors.

- In the work "Multi-sensor data fusion for sign language recognition based on dynamic Bayesian network and convolutional neural network" [12], a multi-sensor fusion structure based on convolutional neural network and dynamic bayesian network for sign language recognition is proposed. In this framework, Microsoft Kinect, which is an RGB-D sensor, is used as a human-computer interaction tool. In particular, in the proposed approach, data is first collected using Kinect, then all features of the image sequence are extracted using a convolutional neural network. Sequences of color and depth features are input to the DBN as observation data. The maximum level of recognition of dynamic isolated sign language is calculated based on the union of the graph model.
  - In the work "A Real-time Hand Gesture Recognition System for Human-Computer and Human-Robot Interaction" [13], the proposed gesture recognition system is designed to improve human-computer interaction and human-robot interaction. As the authors of the study assure, such interaction ensures natural and intuitive communication between people and technology using gestures.
  - The robot "3D Dynamic Hand Gesture Recognition with Fused RGB and Depth Images" [14] offers dynamic gesture recognition technology. In order to solve the existing technology problems, the authors propose to use a network model of three-dimensional dynamic gesture recognition, which uses CNN and LSTM networks and can combine information about RGB and image depth.
  - In the work "Hand gesture recognition using convolutional neural network and histogram of oriented gradients features" [15], the authors emphasize that gesture recognition is the main part of creating a sign language recognition system for people with hearing impairments and is widely used in human-computer interaction. The selected dataset for building the gesture recognition system model is based on American Sign Language using pre-trained AlexNet Convolutional Neural Network and Oriented Gradient Histogram.
  - In the work "Mid-air Gesture Recognition by Ultra-Wide Band Radar Echoes" [16], the authors propose the technology of using microwave radar sensors for human-computer interaction. The peculiarity is that the raw signals generated by such radars have a large dimension and are very difficult to process and interpret for gesture recognition. For these reasons, machine learning techniques are mainly used for gesture recognition, but require numerous gesture patterns for training and calibration, which are specific to each radar [17].
- The given list of studies that are used in the process of sign language recognition is not exhaustive. But the conducted analysis shows that the ideal method does not exist and is unlikely to exist. Therefore, it can be concluded that the mentioned approaches can be ensured by their further adaptation for Ukrainian-language content.

### **1.1.2. Analysis of the Ukrainian sign language development**

In many countries of the world, the possibility of creating and popularizing translators from audio language to sign language and vice versa is being investigated. However, the problem of translating sign language into Ukrainian-language audio content still remains unresolved. It is worth noting that Ukrainian sign language, like any other sign language, has its own rules and grammar, which in turn does not allow the use of existing dictionaries of foreign sign languages. On the territory of modern Ukraine, sign language began to develop in the 19th century - the time of the founding of the first communities, that is, Ukrainians have been creating their own sign language for about two centuries. In 1830, the Lviv school for hearing-impaired children was opened, and in 1843 - in Odesk - these are the approximate dates of the beginning of the development of Ukrainian sign language [18].

It was only recently that sign language was recognized and equated with verbal language. UN General Assembly Resolution 48/96 of December 20, 1993, "Standard Rules for Ensuring Equal Opportunities for Persons with Disabilities," stated that care should be taken to ensure that sign language is used in the education of deaf children, in their families and communities, and it was also recommended to provide sign language interpretation services to facilitate the

communication of sign language people with other people. Subsequently, the issue of using sign language became more active in Ukraine, but the use of Ukrainian sign language in education in independent Ukraine was not introduced until 2006 [19].

The study of sign language linguistics in Ukraine was started by R. Kraevskiyi. The speech-language pathologist worked on the study of sign language, thus carried out its linguistic description on the basis of Ukrainian studies material and created a unique sign dictionary in the form of a manual "Sign Language of the Deaf" [20]. For each gesture, the spatial position and way of movement of the hands are described. In the 21st century in Ukraine, N. Adamyuk, O. Drobot, S. Kulbida, O. Lozynska, M. Davydov are engaged in studying the peculiarities of the syntax of Ukrainian sign language.

Most of N. Adamyuk's scientific works are aimed at studying the peculiarities and linguistic didactic technologies of teaching Ukrainian sign language to deaf and hard-of-hearing children, studying the linguistic features of Ukrainian sign language, as well as studying the basic requirements for teachers of sign language in higher educational institutions and an innovative model their training and retraining [21]. The works of O. Drobot are devoted to the formation of communication skills and comprehensive development of preschool children with hearing impairment [22]. S. Kulbida's research is related to deaf pedagogy of the socio-cultural direction and the conceptual foundations of the development of Ukrainian sign language as a means of learning and a subject of study [23]. The researches of O. Lozynska and M. Davydov are related in their main emphasis on the translation of Ukrainian sign language based on ontology [24, 25].

The analysis of the completed work shows significant progress in popularizing the study of Ukrainian sign language, but the lack of problem-oriented software solutions makes its further research an urgent task.

## **1.2. The main tasks of the research and their significance**

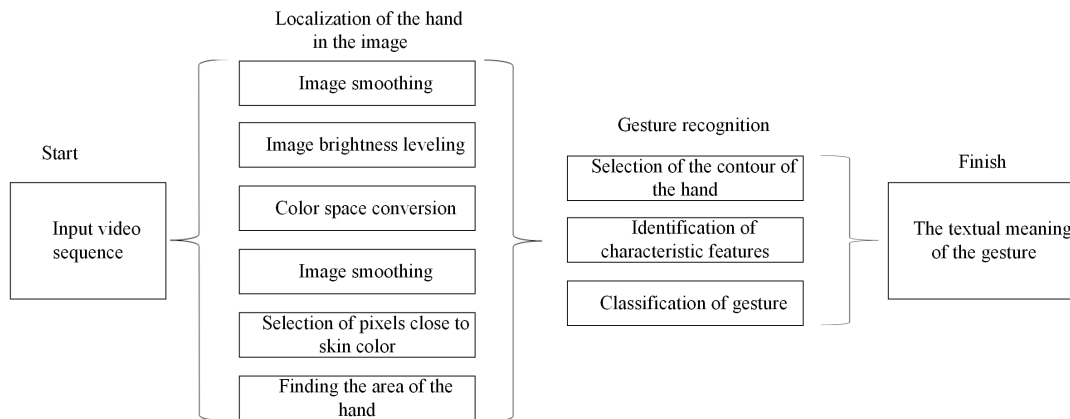
The purpose of the research is to develop an information system for the recognition of Ukrainian sign language. The conducted research will provide means for creating on its basis software for managing information and reference content, generating/transforming elements of sign language and forming an individual learning environment for people with special needs. To achieve the goal, the following tasks must be solved: analyze the existing approaches, methods and software tools used in the field of Ukrainian sign language recognition; to determine the main tasks that arise at the same time; analyze the methods and algorithms of sign language recognition that can be adapted during system development; implement a prototype system for recognizing Ukrainian sign language.

The results of the study solve the actual scientific and practical problem of recognizing Ukrainian sign language and will provide the means to open up additional opportunities for individualizing the educational process for people with special needs.

## **2. Major research results**

People are constantly faced with the task of object recognition. Namely, the human brain processes the information received from the senses, on the basis of which an appropriate decision is made. After that, thanks to the transmission of electrochemical impulses, certain organs carry out the decision made. The above process will occur every time there is a change in the environment. A key stage in this process is the recognition and classification of the surrounding environment, which will help to make the right decision. Given the current development of computer technology, pattern recognition tasks have become the beginning of an independent field and a multitude of tasks that can be solved using gesture recognition.

In order to present the main aspects of the studied subject area, a scheme was finalized that reflects the main stages that must be implemented in the gesture recognition system (Fig. 1).



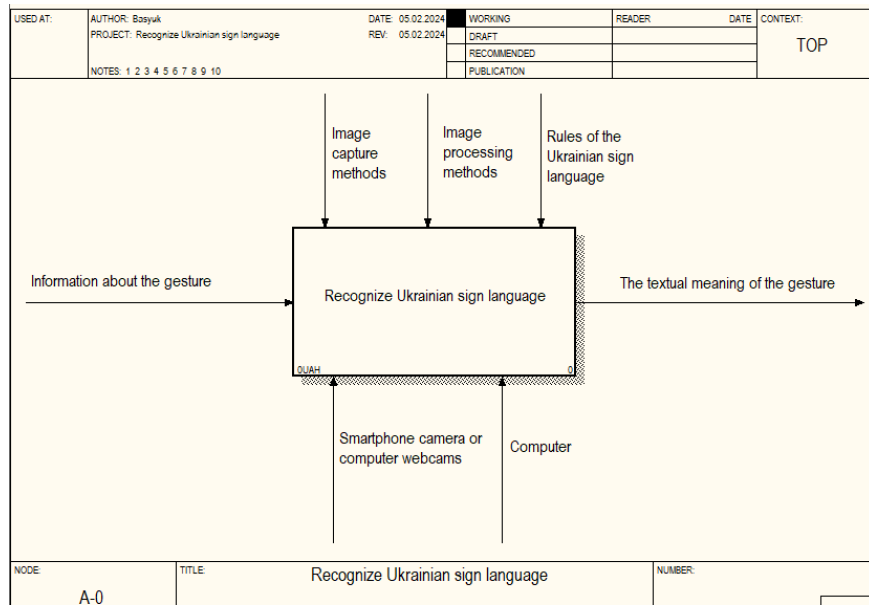
**Figure 1:** The sequence of stages in gesture recognition

As can be seen from Figure 1, the main tasks in the process of gesture recognition are:

- Obtaining an image - usually, this process is implemented using two or more synchronized infrared cameras or smartphone cameras, which continuously transmit a video stream to the system in real time (25-30 frames/sec.);
- Localization of the hand area in the image – on each frame (or series of frames) obtained from the video stream of the camera, the area on which the hand is located is determined. This procedure mainly consists of two stages. The first stage is segmentation (selection) and analysis from the received data of the hand area. This process is performed to remove artifacts from the image and separate the hand region in the image from the background region. As a result of this stage, a selected image of the hand suitable for further processing is formed in the system.
- Gesture recognition – at this stage, the contours of the hand and its characteristics are determined on the image obtained as a result of localization of the hand region. Based on the received data, the gesture is classified.

Let's consider in more detail the main tasks and stages of gesture recognition. The first step is to obtain images that are processed to separate the hand region from the background [26]. This phase is called object localization in the image. After collecting the information, it becomes possible to apply the primary information about the hand in order to filter the data and remove noise from the image. Noises can appear, for example, due to changes in lighting. Artifacts (such as the presence of tattoos, jewelry, etc. on the hand) are also removed. This procedure is very important, considering the set of gestures that need to be distinguished. At the phase of recognition of hand gestures, feature selection is performed [27]. This stage is an important part of the recognition process because hand movements have a significant number of shapes and textures. To recognize a static image of a hand, geometric features are used, among them: the location of the fingertips and their direction. The problem is that these features are not always available due to self-shading and lighting features. Next comes the stage of identifying specific gestures using methods of analyzing filtered data that carry information about hand movements. For this, the classification procedure is used. Before this stage, it is necessary to carry out a process of training the system to enable it to respond to gestures, and to carry out their adaptation for the correct detection of movements [28]. To create a comfortable environment for the user, all processes of capturing, classifying and transforming gestures into text instructions must be performed in real time with an update rate of 25-30 frames per second.

Further work was aimed at conducting a systematic analysis of the subject area using the methodology of functional modeling and graphic description of processes. For these purposes, a structural approach and the IDEF0 standard, which is intended for the formalization and description of business processes, were used. A context diagram showing the process of recognizing Ukrainian sign language is presented in Fig. 2.

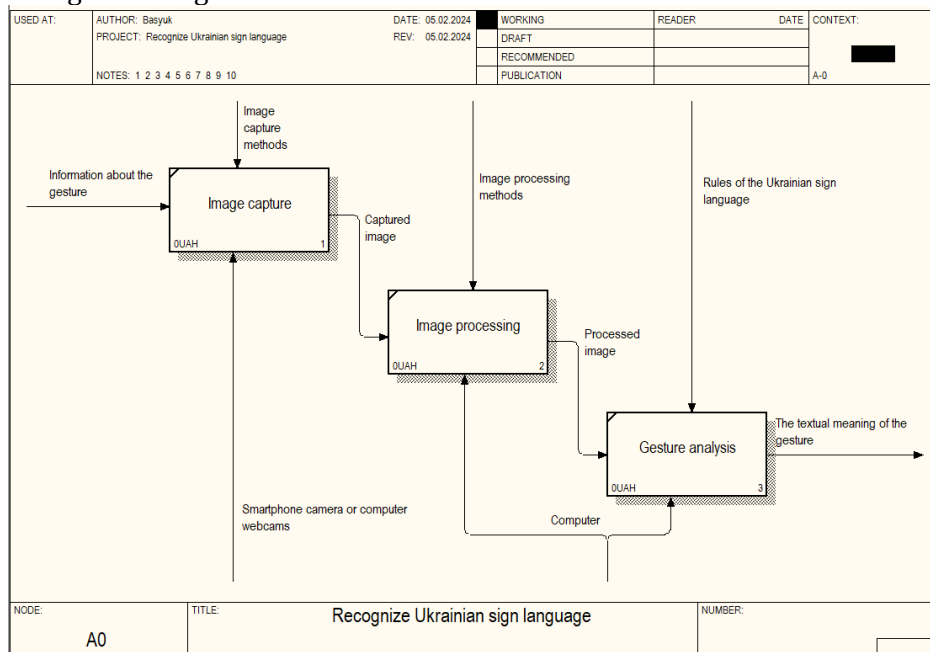


**Figure 2:** Context diagram of the designed system

In the specified model, the input receives information about the gesture, the value of which must be displayed on the screen. The gesture can be transferred in the form of a photo or a video sequence. The output data of the system is the recognized gesture and its text value of the gesture. The driving influences are: image capture methods (methods and algorithms needed to capture an image and localize a gesture on the captured image. These methods are based on the analysis of external features of the gesture); image processing methods (methods and algorithms needed to process the image and extract the outline of the gesture for further analysis. A contour is a curve of a function of two variables along which the function has a constant value. Contours are straight or curved lines that describe sharp changes in brightness in the image [ 29]. There is a high probability of obtaining more than one contour, which is formed in the image due to the presence of noise in the background. Methods for processing the image are necessary in order to remove excess noise from the image and select a clean contour of the gesture for further analysis); the rules of the Ukrainian sign language (information about the gestures of the Ukrainian language. As it is known [30,31], the Ukrainian sign language differs from other sign languages. The rules of the Ukrainian sign language are necessary to highlight the specific features of each gesture. This information is used during image analysis and the formation of the original result). Smartphone cameras or computer web cameras act as mechanisms.

For a more detailed understanding of the logic of the processes taking place in the gesture recognition system, the developed context diagram was decomposed into several sub-processes. The decomposition diagram is presented in Figure 3. As can be seen from the decomposition diagram, the entire process of gesture recognition has been broken down into several sub-processes for greater detail and understanding. Each of the sub-processes has its own input data, output data, control influences and mechanisms necessary for the operation of the process. The entire gesture recognition system is divided into the following three sub-processes: image capture process; image processing process; image analysis process. Image capture is the first sub-process of the entire system. At this stage, the input device converts the gesture into digital form and transfers it to the image processing unit. Image processing is the second sub-process in the entire system. At this stage, the image is processed in such a way that the outline of the hand is clearly visible. The result of the recognition of the gesture of the Ukrainian sign language depends on the result of this process. An incorrectly selected processing algorithm or an incorrectly set parameter of the selected algorithm (for example, the binarization threshold for the image binarization algorithm) will lead to a poor-quality selection of the hand contour, which in turn will not allow accurate identification of the gesture. After successful image processing, it is time for the final stage of gesture analysis. It is here that the image of the gesture is translated into text.

Considering the described stages, the most important stages of the work are contour selection and actual recognition of gestures.



**Figure 3:** Decomposition diagram of the system

## 2.1. Contour selection methods

To highlight the contour, you can use a number of methods: image binarization, wavelet transformation, Canny edge detecto algorithm. In order to choose the optimal one for this task, we will analyze them. The process of binarization is the conversion of a color image or a grayscale image into two-color black and white. The main parameter of this transformation is the threshold, with the value of which the brightness is then compared. After comparing a single image pixel, it is assigned one of two possible values: 0 - "object boundary" or 1 - "another area" [32]. The main goal of binarization is to reduce the amount of information you have to work with. Successful binarization greatly simplifies further work with the image. There are various methods of binarization, which can be conditionally divided into two groups: global (threshold); local (adaptive). Global binarization methods work with the entire image at once. Threshold methods of binarization include: binarization by the lower threshold; upper threshold binarization; double-constrained binarization; incomplete threshold processing; multilevel boundary transformation [33].

$$F'(m, n) = \begin{cases} 0, & F(m, n) \geq t \\ 1, & F(m, n) < t \end{cases} \quad (1)$$

If the first condition is fulfilled for the image point in the given formula 1, then such a point is an object point, if the second condition is fulfilled, then the point will be a background point. In some cases, you can use a variant of the binarization method with a lower threshold [34], which results in a negative of the original image. This method is called binarization with an upper threshold and is represented by the formula:

$$F'(m, n) = \begin{cases} 0, & F(m, n) \leq t \\ 1, & F(m, n) > t \end{cases} \quad (2)$$

If it is necessary to highlight certain areas in which the brightness values of pixels can vary in a certain range, then the binarization method with a double constraint is used [35]. This method is called binarization with an upper threshold and is represented by the formula:

$$F'(m, n) = \begin{cases} 0, & F(m, n) \geq t_1 \\ 1, & t_1 < F(m, n) \leq t_2 \\ 0, & F(m, n) > t_2 \end{cases} \quad (3)$$

If it is necessary to obtain the simplest image for further analysis, then it is worth applying the incomplete threshold processing algorithm, during which the image is deprived of the background with all its details that were in the original photo. Incomplete threshold binarization is represented by the formula:

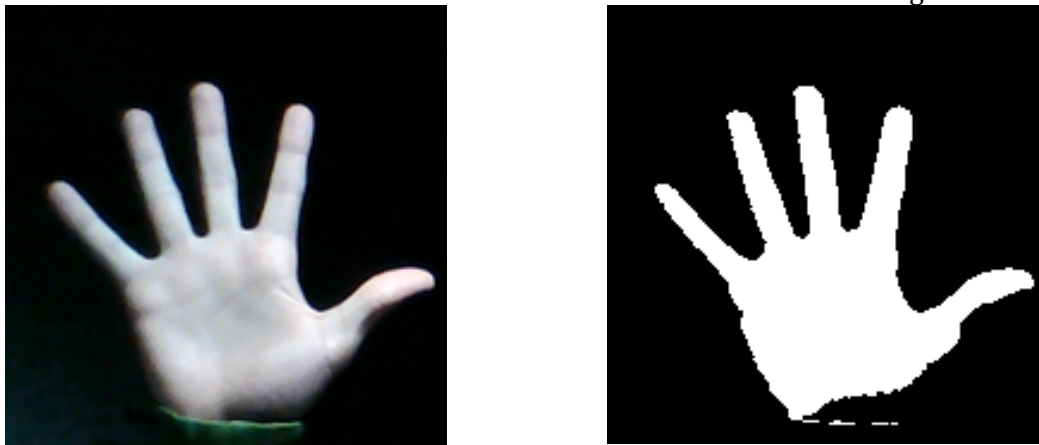
$$F'(m, n) = \begin{cases} F(m, n), & F(m, n) > t, \\ 0, & F(m, n) \leq t \end{cases} \quad (4)$$

If you need to get an image that contains segments with different brightness, you can apply the method of multi-level threshold transformation. However, at the same time, the image obtained during the transformation will no longer be binary [36].

The formula for this transformation is presented below:

$$F'(m, n) = \begin{cases} 1, & F(m, n) \in D1, \\ 2, & F(m, n) \in D2, \\ \dots & \\ n, & F(m, n) \in Dn \\ 0, & \text{в усіх інших випадках} \end{cases} \quad (5)$$

The conducted analysis showed that, taking into account the peculiarities of the input information, it is advisable to use a single binarization threshold, which is used to divide into black and white. The result of the threshold binarization method is shown in Figure 4.



**Figure 4:** An example of an image after conversion by the method of threshold binarization

*Wavelet* transforms are effectively used in signal compression and spectrum analysis [37]. Virtually all wavelets are traditionally defined as functions of a single real variable. Depending on the mathematical model (the structure of the domain of definition, the structure of the domain of possible values and the type of transformations), discrete and continuous wavelets are distinguished. Since the decomposition of wavelets is carried out using floating-point arithmetic, inaccuracies may occur, the magnitude of which is affected by the degree of approximation of the signal. Taking into account the specifics of the subject area, it is possible to use the Haar wavelet [38]. Its technical drawback is that it is not continuous, and therefore not differentiable. However, this property is an advantage when analyzing signals with sudden transitions (discrete signals) that are inherent in this area. In the traditional setting, the wavelet transformation in the Haar basis consists in the linear transformation of a vector of even dimension into another vector of the same dimension. Each pixel of the image can be represented in a binary number system. This decomposition determines the number of bits (N, usually N = 1, 8, 24) and their specific values for storing each pixel [39].

$$J = \sum_{k=0}^{N-1} J_k \cdot 2^k \quad (6)$$

To apply the wavelet transformation over the field GF (p), each pixel of the image must be represented in some number system. This decomposition determines the number of digits of the number system and their specific values that are used in the wavelet transform. The algorithm of order wavelet-transformation of the image is carried out according to the following stages: each pixel of the image is decomposed according to (6) into the digits of a certain numbering system p. A transformation is applied to all p digits with the same numbers. The digits of a certain



numbering system  $p$  of the transformation result are folded into one number according to (6). In fig. 5 presents the variants of the initial image and the image after wavelet transformation by rows in the Haar basis over the field GF (3) and GF (13).



**Figure 5:** Wavelet transform by rows in the Haar basis over the field GF (3) and GF (13)

The Canny edge detector algorithm was developed taking into account such criteria as fast detection and good contour localization. Based on these criteria, an objective function of the cost of errors was constructed, the minimization of which is the "optimal" linear operator for image convolution [40]. In general, Canny's algorithm consists of five stages.

1. Smoothing. At this stage, the image is blurred using a Gaussian filter for localization and noise removal [41].

$$f(x,y) = \frac{1}{2 \cdot \pi \cdot \sigma^2} \cdot \exp\left(-\frac{x^2+y^2}{2 \cdot \sigma^2}\right) \quad (7)$$

2. Search for gradients. Boundaries are searched - where the gradient reaches its maximum value, the boundaries are there:

$$T = \sqrt{G_x^2 + G_y^2} \quad (8)$$

$$\theta = \arctan\left(\frac{G_y}{G_x}\right) \quad (9)$$

The angle of the direction of the gradient vector is rounded and can take the following values: 0, 45, 90, 135. If the angle is from 1 to 20, then it refers to the value 0, and if it is greater than 20, then to the value 45, etc.

3. Muting the lows. Only local maxima are marked as limits.

4. Double threshold filtering. Potential limits are determined by thresholds.

5. Tracing the area of ambiguity. End boundaries are defined by muting all edges not connected to certain (strong) boundaries.

Before applying the detector, the image is usually converted to shades of gray to reduce computational losses. The contour detector algorithm is not limited to calculating the gradient of the smoothed image. Only the points of maximum gradient of the image remain in the contour of the border, and all others lying next to the border are removed. The inclusion of noise suppression in the Canny algorithm, on the one hand, increases the stability of the results, and on the other hand, increases the computational costs and leads to distortion and even loss of contour accuracy. The result of the Canny algorithm is shown in Figure 6.



**Figure 6:** Contour selection using Kenny's algorithm

Comparing the results of image processing by the mentioned algorithms showed that the image binarization method works faster than the wavelet transformation and the Kenny algorithm. However, it should be noted that clearer borders of objects in the image are obtained during processing based on the application of the Kenny algorithm. However, to detect a high-quality contour of the palm, the image binarization method is quite sufficient. In view of that, the image binarization method will be applied in further work.

## 2.2. Gesture recognition

There are many methods that can be used to recognize gestures, among the most common are methods based on the hidden Markov model and neural networks. The hidden Markov model [42] is a statistical model in which the system for which it is created is represented as a Markov process with invisible states. The model can also be represented as the simplest Bayesian network. The main application of hidden Markov models was in the field of recognition of images (gestures), speech, writing and bioinformatics. In addition, they are used in cryptanalysis, machine translation. The simplified structure of the hidden Markov model is represented by the following elements: ovals (these are variables that have random values, namely, the random variable  $x(t)$  is the value of the hidden variable at the time  $t$ , and the random variable  $y(t)$  is the value of the observed variable at the time  $t$ ); arrows (indicate conditional dependencies).

The probability of finding a sequence  $Y = y(0), y(1), \dots, y(L-1)$  of length  $L$  is determined by the dependence:

$$P(Y) = \sum_x P(Y/X)P(X) \quad (10)$$

This modeling technology gained considerable popularity as a result of its successful application and further development in the field of automatic recognition of speech and gestures. Research on hidden Markov models has outperformed all competing approaches, and is the dominant processing paradigm. Their ability to describe processes or signals has been successfully studied for a long time. The reason for this, in particular, is that the technology of building artificial neural networks is rarely used for gesture recognition and similar segmentation problems. However, there are a number of hybrid systems that consist of a combination of hidden Markov models and artificial neural networks, in which the advantages of both modeling methods are used [43].

In general, hidden Markov models describe a two-stage stochastic process. The first stage consists of a discrete stochastic process that is static, causal, and simple. The state space is considered finite. Thus, the process probabilistically describes the state of transition to discreteness, a finite space of states. It can be visualized as a finite automaton with transitions between any pairs of states that are denoted by the transition probability. The behavior of the process at the current moment of time  $t$  depends only on the immediate state of the previous element and can be determined by the dependence:

$$P(S_t | S_1, S_2, \dots, S_{t-1}) = P(S_t | S_{t-1}) \quad (11)$$

At the second stage, for each moment of time  $t$ , additionally, by derivation or on the basis of output data,  $O_t$  is generated. The associative probability distribution depends only on the current state  $S_t$ , not on any previous states or inputs.

$$P(O_t|O_1... O_{t-1}, S_1...S_t) = P(O_t|S_t) \quad (12)$$

The specified sequence of output data is the only thing that can be observed in the behavior of the model. On the other hand, the sequence state assumed during data generation cannot be examined. This is the so-called "hiddenness" from which the definition of hidden Markov models is derived. If you look at the model from the outside - that is, observe its behavior - quite often there are references to the sequence of initial states  $O_1, O_2 \dots O_t$ , as the reason for observing the sequence. Individual elements of this sequence are called observation results [44].

In the literature, behavior recognition patterns of the hidden Markov model are always considered at a certain time interval  $t$ . To initialize the model at the beginning of this period, additional probabilities are used to describe the probability distribution of states at time  $t = 1$ . An equivalent final state criterion is generally absent. Thus, the action of the model enters the final state as soon as an arbitrary state is reached at the time  $t$ . As for gesture recognition, in order to reliably determine the semantics of the movement, it is necessary to allocate it to one of the classes of gestures. Next comes the stage of calculating the probability of receiving a read gesture from models of available gestures. Then the received gesture is classified using the Bayesian classifier. Based on the classification, the gesture can be recognized as one of the available options.

The task of determining the end of a gesture is also not easy. For this, edge cases are considered [43]. When using this classification algorithm, it is highly undesirable to obtain unclear values (data about movements that cannot be clearly attributed to a certain class of gestures). To reduce the number of errors, in the algorithm described above, when situations arise that cannot be unambiguously attributed to a certain class, a weighted sum of the consequences of performing all the gestures classified by the algorithm should be used, or one of the gestures classified with the highest probability should be selected.

As for neural networks, the main research and scientific results obtained in the field of their application for gesture recognition include various methods and architectures that allow to perform this task effectively. In general, since an artificial neural network usually learns with a teacher, this means the presence of a training set (dataset). Ideally, this set contains examples with true values: tags, classes, metrics. An artificial neural network consists of three components: an input layer; hidden (computing) layers; source layer [45]. Neural network training takes place in two stages: direct error propagation; error back propagation. During direct error propagation, a response prediction is made. In backpropagation, the error between the actual response and the predicted one is minimized. Initial weights are randomly assigned. Next, the input data are multiplied by weights to form a hidden layer [46]:

$$h_1=(x_1 \cdot w_1)+(x_2 \cdot w_1) \quad (13)$$

$$h_2=(x_1 \cdot w_2)+(x_2 \cdot w_2) \quad (14)$$

$$h_3=(x_1 \cdot w_3)+(x_2 \cdot w_3) \quad (15)$$

The output data from the hidden layer is passed through a nonlinear function (activation function) to obtain the output of the network:

$$y=f(h_1,h_2,h_3) \quad (16)$$

During the backpropagation of the error, the total error is calculated as the difference between the expected value from the training set and the obtained value (calculated at the stage of forward error propagation), passing through the loss function. The derivative of the error is calculated for each weight (these differentials reflect the contribution of each weight to the total error). These differentials are then multiplied by the learning rate number. The obtained result is then subtracted from the corresponding weights. As a result, the following updated weights will be obtained:

$$w_1=w_1-(\eta \cdot \frac{\partial(err)}{\partial(w_1)}) \quad (17)$$

$$w_2=w_2-(\eta \cdot \frac{\partial(err)}{\partial(w_2)}) \quad (18)$$

$$w_3 = w_3 - \left( \eta \cdot \frac{\partial(\text{err})}{\partial(w_3)} \right) \quad (19)$$

Summarizing the conducted analysis, it can be determined that among the priority areas of neural networks application in the process of gesture recognition are:

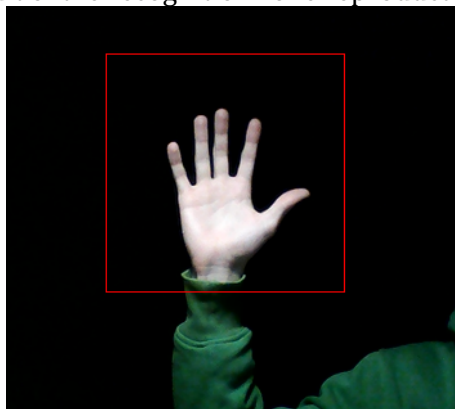
- Deep neural networks (DNN). They are used to interact with complex and variable gestures. The application of deep learning allows to automatically identify important features from a large amount of data [47].
- Recurrent Neural Networks (RNN). Can be used to analyze time dependencies in gestures. This is especially useful when interacting with a sequence of gestures, for example, in the case of sign language recognition [48].
- Convolutional Neural Networks (CNN). Effective in image processing and can be used to recognize spatial features in gestures, such as hand position or finger movement [49].
- Transfer learning. There are examples of using the transfer learning technique for gesture recognition, especially when the amount of annotated data is limited [50].

In summary, neural networks have been successfully used for gesture recognition in various contexts, including virtual reality, medical applications, and gaming industry. Considering that, the mechanism of neural networks will be used to implement the Ukrainian sign language recognition system.

### 2.3. System design

The next stage was the construction of the system, using modern software tools. To implement the software product, it was decided to use the C# programming language and the .NET cross-platform technology. We will use Visual Studio as a development environment. To work with a single-camera system and process the image for further analysis, the OpenCV library is used, or rather, the C# version of EmguCV. The Math.NET library is used to perform matrix operations. The constructed prototype system for recognizing Ukrainian sign language can be conditionally divided into several main independent parts: HandGesturesRecognitionForm, NeuralNetwork, CsvManager, TrainingImageDataManager.

HandGesturesRecognitionForm is the main class of the program, it contains methods for working with the form and analyzing and processing images. The constructor of the HandGesturesRecognitionForm class initializes all components located on the form: fields, buttons, menus, switches. Next, an object of the VideoCapture class is created, which is a class from the EmguCV library designed to capture an image from the device's camera. The Rectangle object is used to position and size the red rectangle in the video from which the gesture will be read for recognition. The result of the recognition zone reproduction is presented in Fig. 7.



**Figure 7:** Selection of the handGestureArea outline

NeuralNetwork – a class that represents a neural network, contains information about the number of nodes of the input, output and hidden layers, the learning coefficient, the matrix of weights between the input and hidden layers, the matrix of weights between the hidden and output layers. To create it, you need to set the following mandatory parameters:

- `inputLayerNodesCount` – the number of input layer nodes. In this case, a value of 4096 is passed, which corresponds to the value of each pixel of the 64 by 64 binary image.
- `hiddenLayerNodesCount` – the number of hidden layer nodes.
- `outputLayerNodesCount` – the number of output layer nodes.
- `learningRate` is a learning rate, a parameter of gradient learning methods of neural networks, which allows you to control the amount of weight correction at each iteration.
- `epochs` – the number of steps (epochs) required to find the optimal value.

`CsvManager` is a class responsible for saving the training set to a csv file. Contains private file name and save path information.

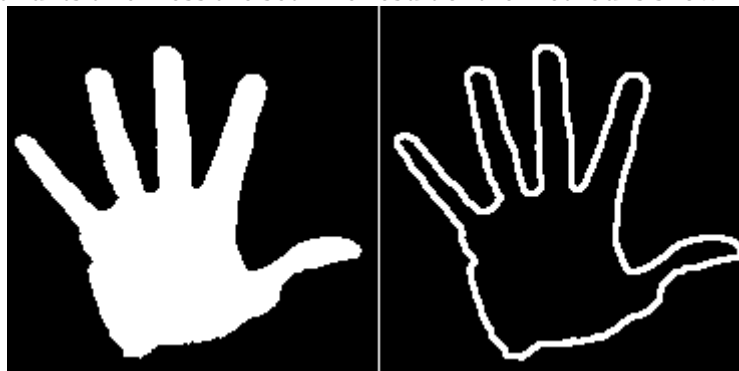
`TrainingImageDataManager` is a class that is responsible for saving pictures for neural network training. It contains private information about how photos are saved.

Let's take a closer look at the implementation of some key methods of the class. The `EmguCV Threshold` library method is used for binarization. The method accepts a binarization threshold that is obtained from a control of the form `binaryImageThresholdTrackBar`. The result of image binarization is shown in Figure 8.



**Figure 8:** Image binarization

The `DrawContours` library method is used to select the contour. When the method is called, the contour color and its thickness are set. The result of the method is shown in Figure 9.



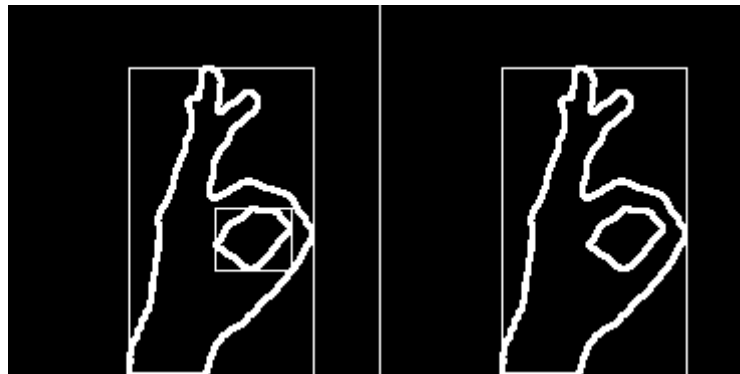
**Figure 9:** Selection of the contour of the palm

`DrawBiggestContourBoundingRectangleImage` is the method, responsible for rendering the image of the largest contour. Since the gesture is the largest contour in the image, it was decided to display it in a separate window to increase visibility. Method code:

```
private void DrawBiggestContourBoundingRectangleImage(
    Image<Gray, byte> biggestBoundingRectangleImage,
    Rectangle biggestBoundingRectangle)
{
    biggestBoundingRectangleImage.Draw(
        biggestBoundingRectangle, new Gray(255));
    biggestContourBoundingRectanglePictureBox.Image =
        biggestBoundingRectangleImage.Bitmap;}

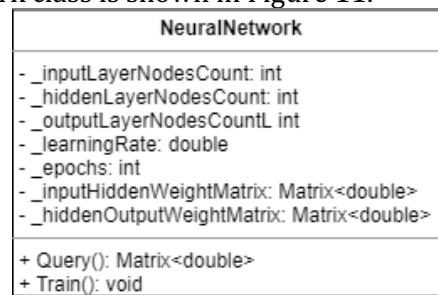
```

Figure 10 shows a gesture with all contours closed and a gesture with the largest contour selected.



**Figure 10:** Selection of the largest contour

The NeuralNetwork class is designed to create a neural network object, train it, and poll it. The class contains information about the nodes of the input, output, and hidden layers, the values of certain learning coefficients, and methods for training and polling a neural network. The structure of the NeuralNetwork class is shown in Figure 11.



**Figure 11:** Structure of the NeuralNetwork class

When creating an object of the neural network class, the weight matrices are initialized. Large values should be avoided when initializing the initial values of the weights, as using the activation function in this range of values may reduce the ability of the network to learn to better values. Therefore, the weights are selected from a normal distribution centered at zero and with a standard deviation whose value is inversely proportional to the square root of the number of input nodes. The Query method accepts the input data of the neural network as an argument and returns its output data. To do this, signals from the input layer nodes must be passed through the hidden layer to the output layer nodes to receive the output data. At the same time, as the signals spread, it is necessary to smooth them using the weighting coefficients of the connections between the relevant nodes, and also to apply the sigmoid to reduce the output signals of the nodes. To obtain the output signals of the hidden layer, it is necessary to apply them to each sigmoid value. Training includes two phases: the first is the calculation of the output signal, which is what the Query function does, and the second is the backpropagation of errors, which informs what the corrections to the weighting factors should be.

The first part is the calculation of output signals for a given training example. The second part is a comparison of the calculated output signals with the desired response and updating the weighting coefficients of connections between nodes based on the differences found.

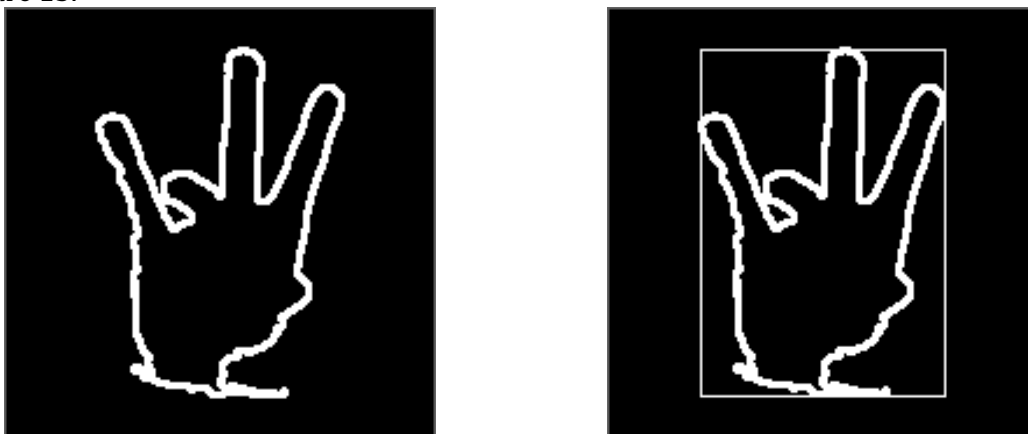
As a result of the work, a prototype of the application was developed, which is able to recognize the gesture of the alphabet of the Ukrainian sign language. For clarity, the program outputs the result at each iteration, starting with the raw video and ending with the recognition result in the form of a gesture value. It all starts with video capture. An example of a frame from the original, unprocessed video stream and its binarized version is shown in Fig. 12.





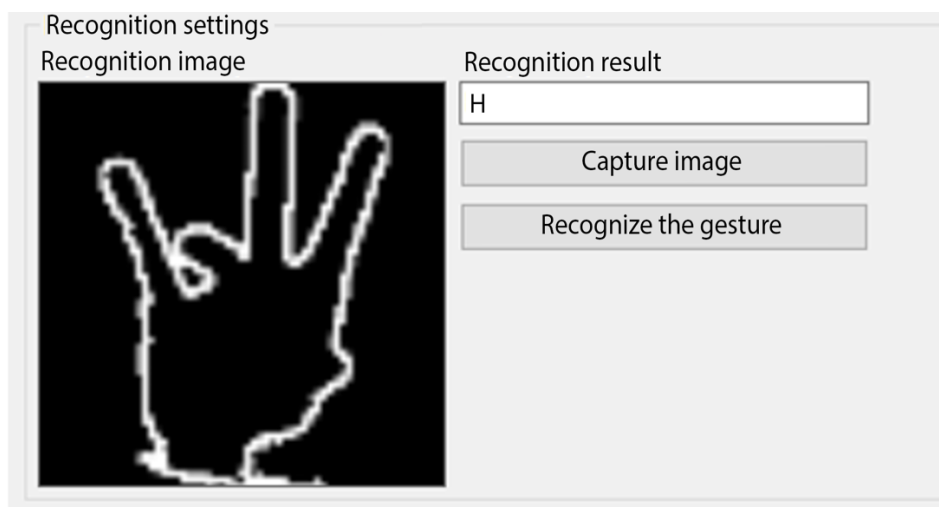
**Figure 12:** The original frame and its binarized version

By capturing a binarized image, you can shape the outline of the palm. The output contour together with the selected region of the largest closed contour highlighted by a rectangle is shown in Figure 13.



**Figure 13:** Contour of the palm and selection of the largest contour

After selecting the working surface, it is necessary to reduce it to a square image, since the selected largest contour is not always a square, as shown in Figure 13. Since the neural network contains 4096 input layers, the final image is reduced to a size of 64 by 64 pixels. After capturing the image and clicking on the "Recognize the gesture" button, the recognition settings panel looks like this:



**Figure 14:** A fragment of the recognition panel

The text value of the gesture is displayed in the "Recognition result" text field. The program successfully recognized the demonstrated gesture and displayed its explanation on the screen.

## Conclusion

As a result of the conducted research, the existing methods and known systems that provide means of recognizing Ukrainian sign language and describe the mechanisms of their implementation were analyzed. Technologies and software tools for sign language recognition were analyzed, which made it possible to identify the features of existing approaches. As the analysis showed, today there are many software systems, but all of them are characterized by certain shortcomings, from the commerciality of the application to the impossibility of the application for the recognition of Ukrainian-language content, which makes the task of constructing an information system for the recognition of Ukrainian sign language urgent. In order to present the main aspects of the studied subject area, a scheme was finalized that reflects the main stages that must be implemented in the gesture recognition system. The next stage was the design of the software system using a structural approach and displaying the created diagrams in accordance with the IDEF0 standard. The study presents a context diagram and its decomposition, which created the basis for the study of features and the formation of methodological foundations for the construction of an information system. The analysis and justification of the choice of methods for the selection of contours and recognition of gestures in the incoming information message was carried out. The developed prototype is characterized by modular construction, the ability to recognize gestures of the Ukrainian alphabet and can be useful as an additional communication tool. The conducted research provides methodological and algorithmic foundations for building a communication environment for people with special needs.

Further research will be directed to testing and improving the system, eliminating conflicts and expanding functionality in accordance with the specified requirements.

## References

- [1] M. Coster, D. Shterionov, M. Herreweghe, J. Dambre, Machine translation from signed to spoken languages: state of the art and challenges. *Universal Access in the Information Society*. 2023, pp. 1-27.
- [2] World Health Organization. Deafness and hearing loss, URL. <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss>
- [3] A. Núñez-Marcos, O. Perez-de-Viñaspre, G. Labaka, A survey on Sign Language machine translation. *Expert Systems with Applications*. 2023, Vol.213: pp.1-28
- [4] N.Adaloglou, T.Chatzis, I.Papastratis, A.Stergioulas, G. Papadopoulos, V. Zacharopoulou, A comprehensive study on sign language recognition methods, arXiv:2007.12530, 2020.
- [5] S. McBurney, Sign Language: History of Research. *Encyclopedia of Language & Linguistics*, 2006, pp.310-318.
- [6] F. Quek, D. McNeill, B. Bryll, S. Duncan, X. Ma, C. Kirbas, K. McCullough, R. Ansari. Multimodal Human Discourse: Gesture and Speech, *ACM Transactions on Computer-Human Interaction*, vol. 9, no. 3, 2002. pp.171-193.
- [7] S. Jiang, B. Sun, L. Wang, Y. Bai, K. Li, Y. Fu, Skeleton aware multi-modal sign language recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021. pp. 3413–3423
- [8] R. Minu, A Extensive Survey on Sign Language Recognition Methods. In *Proceedings of the 2023 7th International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India, 23–25 February 2023; pp. 613–619.
- [9] O. Mediakov , T. Basyuk, Specifics of Designing and Construction of the System for Deep Neural Networks Generation // *CEUR Workshop Proceedings*. – 2022. – Vol. 3171 : *Computational Linguistics and Intelligent Systems 2022 : Proceedings of the 6th*



- International conference on computational linguistics and intelligent systems (COLINS 2022). Vol. 1 : Main conference, Gliwice, Poland, May 12-13, 2022, pp. 1282–1296.
- [10] D. Zhu, V. Czehmann, E. Avramidis, Neural Machine Translation Methods for Translating Text to Sign Language Glosses. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. Vol.1: Long Papers, 2023, Toronto, Canada, pp. 12523–12541.
- [11] A. Agarwal, M. Thakur, Sign language recognition using Microsoft Kinect, Proceedings of the sixth International Conference on Contemporary Computing (IC3), Noida, India, 2013, pp. 181-185.
- [12] Q. Xiao, Y. Zhao, W. Huan, Multi-sensor data fusion for sign language recognition based on dynamic Bayesian network and convolutional neural network. *Multimed Tools Appl.* 2019. Vol. 78, pp. 15335–15352.
- [13] V. Ponzi, E. Iacobelli, C. Napoli, J. Starczewski, A Real-time Hand Gesture Recognition System for Human-Computer and Human-Robot Interaction. Proceedings of the International Conference of Yearly Reports on Informatics, Mathematics, and Engineering, Catania, Italy, August 26-29, 2022, pp.52-58.
- [14] Y. Qingshan, B. Yong, C. Lu, J. Wenjie, 3D Dynamic Hand Gesture Recognition with Fused RGB and Depth Images. Proceedings of the 2022 3rd International Conference on Big Data & Artificial Intelligence & Software Engineering, Virtual Event, Guangzhou, China, October 21-23, 2022, pp.38-44.
- [15] A. Kika, A. Koni, Hand gesture recognition using convolutional neural network and histogram of oriented gradients features. Proceedings of the 3rd International Conference on Recent Trends and Applications in Computer Science and Information Technology Tirana, Albania, November 23rd to 24th, 2018, pp.75-79.
- [16] A. Sluÿters, Mid-air Gesture Recognition by Ultra-Wide Band Radar Echoes. Proceedings of the Workshops on Engineering Interactive Computing Systems (EICS-WS 2022) co-located with the 14th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (SIGCHI 2022). Sophia Antipolis, France, June 21, 2022, pp.28-39.
- [17] A. Vasyliuk, T. Basyuk, V. Lytvyn. Specialized interactive methods for using data on radar application models, Proceedings of the 2nd International workshop on modern machine learning technologies and data science (MoMLeT+DS 2020). Vol. I: Main conference, Lviv-Shatsk, Ukraine, June 2-3, 2020, Vol. 2631: pp. 1-11.
- [18] Association of deaf teachers. Educational institutions for the deaf in the pre-revolutionary period. URL. [https://onp.ucoz.ua/news/navchalni\\_zaklady\\_dlja\\_gluxyx\\_v\\_dorevoljuciyniy\\_period/2013-06-13-51](https://onp.ucoz.ua/news/navchalni_zaklady_dlja_gluxyx_v_dorevoljuciyniy_period/2013-06-13-51). (In Ukrainian)
- [19] S. Kulbida, Gesture bilingual approach in the practice of special institutions of Ukraine. Special child: training and education. - 2022. - N 3. - P. 7-18. (in Ukrainian).
- [20] R. Kraevskiy, Sign Language of the Deaf. – K. 1964. P.220. (in Ukrainian).
- [21] N. Adamyuk, Features of socio-cultural communication of sign language people in the educational process. Abstracts of XXX International Scientific and Practical Conference Interaction Of Society And Science: Problems And Prospects. London, England June 15 – 18, 2021, pp. 307-311. (in Ukrainian).
- [22] O. Drobot, Psychophysiological features of the formation of the lexical competence of national verbal languages among students with hearing impairments. Education of persons with special needs: ways of development, 2019, № 15. pp.67-77. (in Ukrainian).
- [23] S. Kulbida, A competent approach in the training of deaf-pedagogical personnel. Modern technologies for the development of professional skills of future teachers: coll. of science Proceedings of the First International Internet Conference, October 26, 2017. Uman: FOP, 2017, pp. 122-124. (in Ukrainian).
- [24] O. Lozynska, M. Davydov, Information technology for Ukrainian Sign Language translation based on ontologies. *Econtechmod. An International Quarterly Journal.* 2015, vol. 04, No. 2, pp.13–18.
- [25] T. Basyuk, A. Vasyliuk, Approach to a Subject Area Ontology Visualization System Creating, Proceedings of the 5rd International Conference on Computational Linguistics and

- Intelligent Systems (COLINS-2021). Volume I: Main Conference, Kharkiv, Ukraine, April 22-23, 2021, Vol-2870, pp. 528-540.
- [26] D. Khurana, A. Koli, K. Khatte, Natural language processing: state of the art, current trends and challenges. *Multimed Tools Appl* 82, 2023, pp.3713-3744.
- [27] D. Bragg, O. Koller, M. Bellard, L. Berke, P. Boudreault, A. Braffort, N. Caselli, M. Huenerfauth, H. Kacorri, T. Verhoef et al., Sign language recognition, generation, and translation: An interdisciplinary perspective, arXiv preprint arXiv:1908.08597, 2019, pp16-31.
- [28] T. Adugna, A. Ramu, A. Haldorai, A Review of Pattern Recognition and Machine Learning. *Journal of Machine and Computing*. 2024, pp.210-220.
- [29] M. Baker, U. Solanki, Artificial Intelligence Models in Pattern Recognition. In: *Handbook of Artificial Intelligence Applications for Industrial Sustainability*. CRC Press. 2024, pp. 18-36.
- [30] A. Zamsha, The category of quantity in signs of Ukrainian Sign Language. *Proceedings of the International scientific conference "Current trends and fields of philological studies in the challenging reality"*. Riga, the Republic of Latvia, July 29-30, 2022, pp.268-270.
- [31] T. Basyuk, A. Vasyliuk, Peculiarities of an Information System Development for Studying Ukrainian Language and Carrying out an Emotional and Content Analysis // *CEUR Workshop Proceedings*. – 2023. – Vol. 3396: Computational Linguistics and Intelligent Systems 2023: Proceedings of the 7th International Conference on Computational Linguistics and Intelligent Systems. Volume II: Computational Linguistics Workshop, Kharkiv, Ukraine, April 20-21, 2023, pp. 279-294.
- [32] M. Prodan, C.-A. Boiangiu, Document Image Binarization Process. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, 2023. Vol. 14(2), pp.93-114.
- [33] F. Kasmin, A. Abdullah, A. Prabuwo, Ensemble of Steerable Local Neighbourhood Grey-level Information for Binarization. *Pattern Recognition Letters*. 2017. Vol. 98, pp.8-15.
- [34] S. Abdullah, S. Ismail, M. Hasan, P. Shivakumara, Novel Adaptive Binarization Method for Degraded Document Images. *Computers, Materials & Continua* 2021, Vol. 67(3), pp.3815-3832.
- [35] J. Wu, Z. Li, Y. Liu, Double-Constraint Inpainting Model of a Single-Depth Image, *Data, Signal and Image Processing and Applications in Sensors*. 2020, Vol. 20(6), pp. 345-364.
- [36] A. Abubakar, Multilevel Thresholding for Image Segmentation Using Mean Gradient, *Journal of Electrical and Computer Engineering*. Vol. 2022 (1), pp.1-9.
- [37] A. Osadchiy, A. Kamenev, V. Saharov, S. Chernyi, Signal Processing Algorithm Based on Discrete Wavelet Transform. *Designs* 2021, Vol 5(41), pp. 1-13.
- [38] X. Guoping, L. Wentao, Z. Xuan, L. Chang, H. Xinwei, W. Xinglong. Haar Wavelet Downsampling: A Simple but Effective Downsampling Module for Semantic Segmentation. *Pattern Recognition*. 2023. Vol. 143, pp.678-689
- [39] P. Fleet, The Haar Wavelet Transformation. *Journal of Computer Engineering*. 2019, Vol. 34, pp.125-181.
- [40] X. Qin, A modified Canny edge detector based on weighted least squares. *Computational Statistics*. Issue 1. 2021. pp 641-659.
- [41] J. Patel, J. Patwardhan, K. Sankhe, R. Kumbhare, Fuzzy inference based edge detection system using Sobel and Laplacian of Gaussian operators. *ICWET '11: Proceedings of the International Conference & Workshop on Emerging Trends in Technology* February. 2011, pp.694-697.
- [42] M. Franzese, A. Iuliano, Hidden Markov Models. *Encyclopedia of Bioinformatics and Computational Biology*. Vol. 1, 2019, pp.753-762.
- [43] T. Hiraoka, S. Takase, K. Uchiumi, A. Keyaki, N. Okazaki, Recurrent Neural Hidden Markov Model for High-order Transition. *ACM Transactions on Asian and Low-Resource Language Information Processing*. 2021. Vol. 21, Issue 2. pp 1-15
- [44] A. Tur, H. Keles, Evaluation of hidden Markov models using deep CNN features in isolated sign recognition. *Multimed Tools Appl*. 2021. Vol. 80, pp 19137-19155.
- [45] H. Ahn, J. Kim, J. Shim, J. Kim, Hand Gesture Recognition for Doors with Neural Network. *Proceedings of the International Conference on Research in Adaptive and Convergent Systems (RACS '17)*. 2017, pp.15-18.

- [46] G. Murthy, R. Jadon, Hand Gesture Recognition using Neural Networks. Advance Computing Conference (IACC). 2019 IEEE 2nd International, pp.134-138.
- [47] Z. Bao, T. Liu, Radar micro moving gesture recognition method based on multi-scale fusion deep network. Proceedings of the 2022 5th International Conference on Artificial Intelligence and Pattern Recognition (AIPR '22). 2022, pp. 657–663.
- [48] H. Alimam, W. Mohamed, A. Selmy, Deep Recurrent Neural Network Approach with LSTM Structure for Hand Movement Recognition Using EMG Signals. Proceedings of the 2023 12th International Conference on Software and Information Engineering (ICSIE '23). 2023, pp 58–65.
- [49] L. Li, W. Wei, D. Chen, W. Yang, H. Jiang, Gesture Recognition with Complex Background Based on Improved Convolutional Neural Network. Proceedings of the 2021 5th International Conference on Electronic Information Technology and Computer Engineering (EITCE '21). 2021, pp.1345–1349.
- [50] L. Gao, L. Zhu, S. Xue, L. Wan, P. Li, W. Feng, Multi-View Fusion for Sign Language Recognition through Knowledge Transfer Learning. Proceedings of the 18th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry. (VRCAI '22). 2022, pp 1–9.