Kazakh sign language recognition system based on the Bernsen method and morphological structuring

Saule Kudubayeva[1], Nurzada Amangeldy[1], Zakirova Alma[1]

[1]L.N.Gumilyov Eurasian National University of Nur-Sultan, 2, Satpayev str., Nur Sultan, 101008, Kazahstan
saule.kudubayeva@gmail.com, nurzadaamangeldy@gmail.com, alma_zakirova@mail.ru

Abstract

This paper explores an approach to constructing classes of various movements of a person's hands while showing gestures and methods for recognizing them. The localization of the hand relative to the body, the direction of movement of the hands and the orientation of the palm are taken as the main properties while showing gestures. To build classes, the use of an ontological model of the subject area, focused on the problems of recognizing sign language is proposed. To analyze the data and build the ontological model, about a thousand gestures that characterize the possible variations of gestures in the form of their demonstration were selected. As a result of the research, more than two hundred different classes have been identified for which various methods and recognition algorithms have been developed taking into account the specific features of the classes. Approaches to the detection and recognition of gestures during the implementation of intelligent human-machine interface technology are considered. A new algorithm based on the Bernsen method and morphological structuring and correlation analysis is proposed. Based on the algorithm, a system was created and an experiment was carried out. The experimental results showed the effectiveness of the proposed algorithm and can be used to recognize other types of classes proposed in the classification by modifying the proposed algorithm, that is, using various methods for processing matrices.

Keywords: gestures recognition, human-machine interface, Viola-Jones detector, correlation analysis, sign language translation, the Bernsen method.

INTRODUCTION

Intellectual technologies for sign languages (including Kazakh) and their features have not been considered properly in world science y et, therefore the development of such technologies is more relevant than ever. Also, the Kazakh sign language is subject to partial extinction, and this requires special attention from linguists, sign language interpreters and software developers.

The issue of lack of sign language interpreters of the Kazakh Sign Language (KSL) in Kazakhstan is very serious. First of all, this is due to the lack of schools and an approved state standard, according to which it is necessary to train future specialists. In addition, three main dialects are distinguished in the KSL: northeastern, southern and western. This causes presents in the study of elements of the KSL.

Kazakhni scientists have developed a website www.surdo.kz, in which there is a fingerprint alphabet developed by the Kazakh Society of the Deaf at the end of the 20th century based on the Russian sign language, a sign fingerprinting dictionary, tabs: proverbs — sayings, songs, fairy tales with sign language [2]. In Kazakhstan about half a million people use the sign language, and the number of translators willing to help them is hundreds of times smaller. Seeing a doctor or notary, study at the university — everything requires translation. In Europe and America, paramedics, police officers and a number of other professions are required to know and understand the sign language. In Kazakhstan, it presents some difficulties. You can get the help of a translator by submitting an application or paying money for the service, so these services are used infrequently.

In Kazakhstan there also exists «Surdo-Online» social project, which provides a social rehabilitation and adaptation of deaf-mute and hearing-impaired people in society, offering high-quality services in all organizations and enterprises of various forms of ownership. Besides, the project provides jobs for more than three hundred people, which will further enable the provision of the Surdo-Online services throughout the Republic of Kazakhstan.

«Surdo-Online» service is a conference between a person with impaired hearing who needs information and a service provider on the one hand, and an operator of the sign language specialist, on the other. The service provides certified specialists who have passed the certification of sign language specialists in Kazakhstan and beyond its borders, regardless of the location and time of the disabled person's access; the service is available around the clock; the only condition of getting it is access to the Internet. With the Purpose of Equal Opportunities provision, this service can be used both from computers and from tablets and mobile phones [3].

1.   THEORETICAL PART

Gesture (from lat gest) is the movement of the human body or its parts, which has a certain meaning, that is, a symbol or emblem.

Gesture speech is a form of interpersonal communication, which is characterized by special lexical and grammatical patterns that support people with hearing impairment.

Sign language (SL) is a non-verbal communication system between hearing people and people with hearing problems, and for the latter it is used as the main method of communication in which a gesture can be found to express almost any word [3]. The main unit of a sign language is

a gesture that is characterized by iconicity, i.e. the ability to designate an object by demonstrating with the help of hand movements, facial expressions and articulation of the face, turning the head, etc. visual parameters of the object.

It is often impossible to show first and last names, foreign, technical and medical terms using a sign language; therefore, along with a sign language, the hearing impaired widely use the Dactyl alphabet. Dactyl grammar is similar to a deaf person's native language grammar. Fingerprinting is often referred to as finger writing in the air: visually acceptable and based on all spelling rules. But it does not concern punctuation, for the exclamation and question marks are conveyed by the corresponding facial expressions; point and multipoint — by a pause; hyphens, colons and other punctuation marks, despite the fact that there are display forms, are also not displayed in the fingerprint.

To parameterize the display of gestures, five components of the gesture are distinguished: configuration (shape of the arm / hands), place of performing (localization), direction of movement, nature of movement and non-manual component (facial expression and articulation) [3,4].

Examples of KSL gestures characterized by L. S. Dimskis notation components are shown in Table 1 and show how each gesture can be represented using the characteristics of its components [5].

Table 1

The example of the notation of gestures of the KSL by L. S. Dimskis [5]

Components	Characteristics of components and notations
Configuration (finger position)	A-configuration Ә-configuration Б-configuration …1-configuration 2- configuration …
Palm direction	down up to yourself ………
Localization (location of gesture)	at forehead level with riding both cheeks at shoulder level without touching at shoulder level with touching of right shoulder ………
Direction and nature of the movement	down to yourself from yourself

As the main properties when showing gestures let us take localization (fig. 1), direction of the movement (fig 2.) and orientation of the palm, and introduce the following concepts and designations for building the model (Table 2.)

5

Main parameters for gestures showing

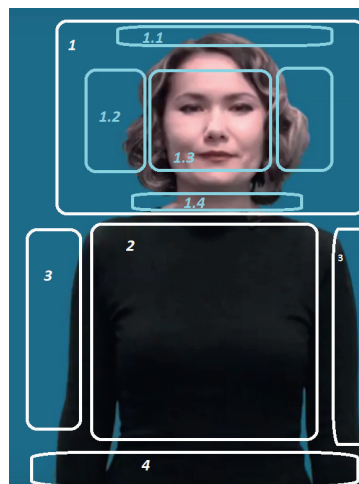| Localization		Ω
1	in the area of the head	HA
1.1	over the head	HA/OH
1.2	to the right or to the left of the head	HA/RLH
1.3	touching the face	HA/TF
1.4	touching the neck	HA/TN
2	neutral zone	NZ
3	by the right or left shoulder	NRLSH
4	in the area of the waist	W
Palm, hand orientation		**ψ**
1	the palm is directed to the right, left, up, down	PLRLUD
24	the palm is directed to the speaker or from the speaker	PTFS
Movement direction		**↔**
1	from the speaker or to the speaker	TFS
2	up, down, to the right, to the left	DULRS
3	circular movements	CM
4	motionless, static	ML
Classes		**K**
1	1 handed	1H
2	2 handed	2H
2.1	hands do not cross	NTINT
2.2	hands cross	INT

Figure 1. Localization (lace of a gesture)

The ontological domain model is implemented and tested in the Protégé 5.5.0 beta system. As a result of a detailed study of the sign language vocabulary [2] in the form of presentation of gestures, in the position or direction of the structural elements of the gesture, the ontological model was obtained in the form of 204 classes of forms of gesture performance. This does not include gestures when showing which two hands cross, since such representations require further detailed analysis.

Figure 2. Directions of the movement (movements can be direct, intermittent, spasmodic, repeating)

Ontology in the context of information technology is represented by a hierarchical system of concepts and terms (structure, model) of a certain subject area [6]. To create an ontological model of the subject area, the Kazakh sign language was chosen; about a thousand gestures (See on the site) were taken as a vocabulary of the language [3].

The main components of an ontology are classes or concepts. Classes are abstract groups that can include instances, other classes, or a combination of both. Classes in ontologies are usually organized into a taxonomy — a hierarchical classification of concepts in relation to inclusion. To construct an ontological model of a sign language two classes of gestures of the first level were distinguished — one-handed (a gesture is demonstrated with one hand) and two-handed, when information is transmitted using the movements of two hands. A class can be subdivided into subclasses that represent more specific concepts than a superclass.

2. EXPERIMENTAL PART

2.1. Description of some classes and experiments on them

Gestures can be divided into static and dynamic. Static gestures of a fixed-arm state (Fig. 4,5) in space are spatial, and dynamic signatures are characterized by continuous

movement of the hand from the starting point to the finishing point over a certain period of time. There will be 42 letters on the Kazakh fingerprint, as it is in the Kazakh alphabet.
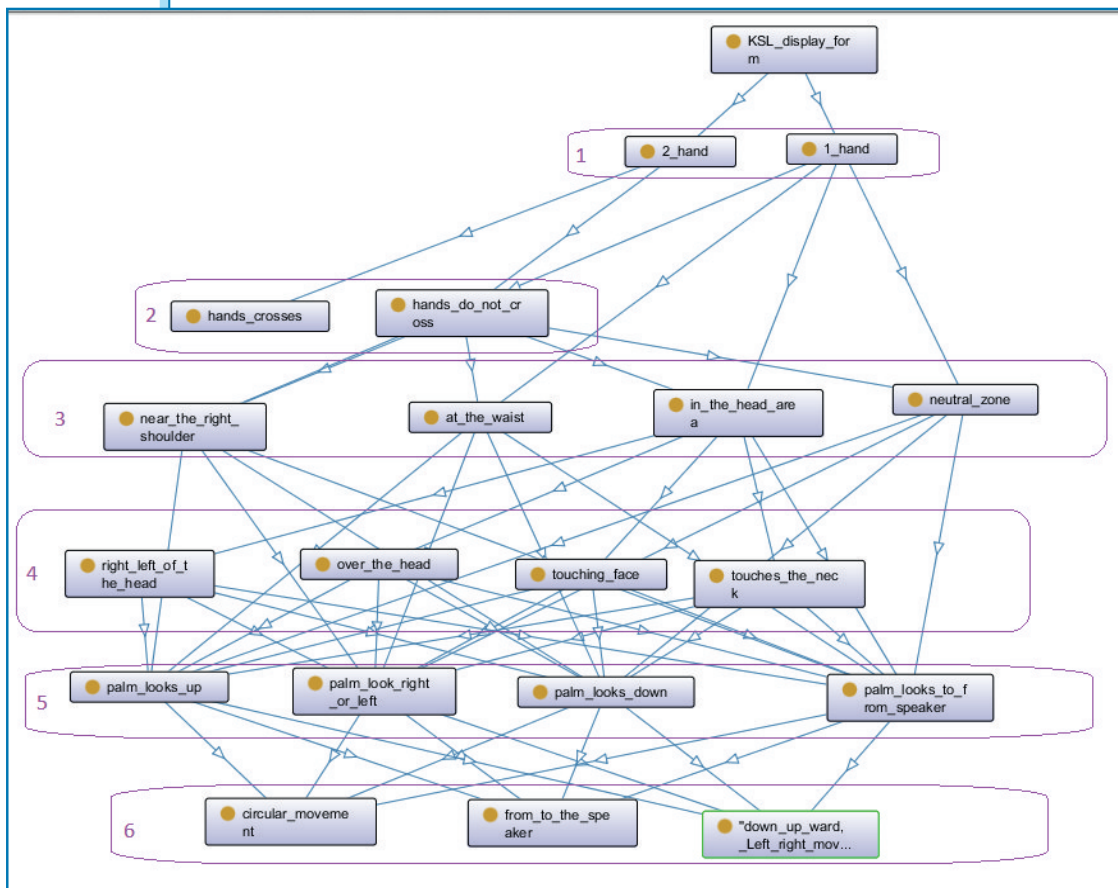

Figure 3. Ontological model

Figure 4. One-handed static gesture classes

Figure 5. Two-handed static gesture classes

We call gestures with two hands symmetrical if both the wrists and the direction of movement are the same, or they repeat the movement of each other, as if reflected in a mirror.

About 350 thousand gestures studied in qualifications are displayed with one hand, and about — 650 with two hands.

Figure 6. One-handed, neutral zone, the palm is directed to the speaker or from the speaker, movement direction up, down, to the right, to the left gesture classes
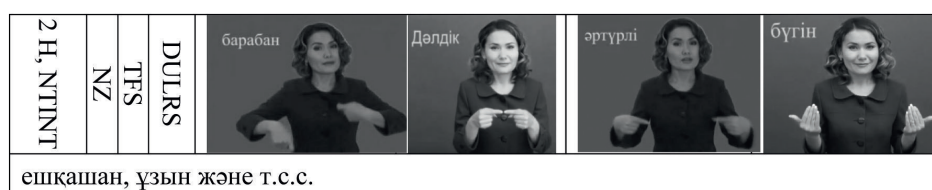

Figure 7. Two-handed, neutral zone, the palm is directed to the speaker or from the speaker, movement direction up, down, to the right, to the left gesture classes

Figure 8. Two-handed, by the right or left shoulder, the palm is directed to the speaker or from the speaker, movement direction up, down, to the right, to the left gesture classes

In asymmetric two-handed gestures, one hand is often motionless — this is called a passive hand, and the second hand can make complex movements — this hand is called active; often the shape and movement of the gesture are determined by the active hand.

2.2. Experiment description

Below we proposed an algorithm (fig.9) for recognizing such signatures for classes of one-handed and two-handed signatures that satisfy the following conditions (Fig. 6,7,8):

$$- \Omega \text{ NZ } \psi \text{ PTFS} \leftrightarrow \text{DULRS};$$

$$- \Omega \text{ NSH } \psi \text{ PTFS} \leftrightarrow \text{DULRS};$$

$$- \Omega \text{ W } \psi \text{ PTFS} \leftrightarrow \text{DULRS};$$

The use of the correlation method and the method of direct comparison with the standard imposes general requirements on the process of image preprocessing. They consist

in the fact that the image and the standard must be equally oriented, have an equal scale and be not shifted relative to each other in the image field. The experiment was conducted using Math Works MATLAB R2018a — a specialized package for solving engineering, scientific, technical and economic problems [10, 11].
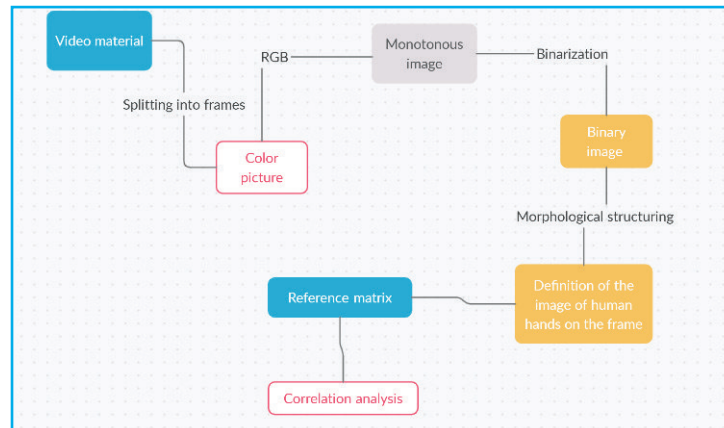

Figure 9 The structural diagram of the recognition system of the Kazakh sign language

We can offer standard color images in two ways: an RGB image and an indexed image. An RGB image consists of three matrices (red, green, blue), corresponding to three color components with dimensions WxHx3.

$$R=\begin{pmatrix} r_{00} & r_{01} & \dots & r_{0w} \\ r_{10} & r_{11} & \dots & r_{1w} \\ \dots & \dots & \dots & \dots \\ r_{h0} & r_{h1} & \dots & r_{hw} \end{pmatrix} \quad G=\begin{pmatrix} r_{00} & r_{01} & \dots & r_{0w} \\ r_{10} & r_{11} & \dots & r_{1w} \\ \dots & \dots & \dots & \dots \\ r_{h0} & r_{h1} & \dots & r_{hw} \end{pmatrix} \quad B=\begin{pmatrix} r_{00} & r_{01} & \dots & r_{0w} \\ r_{10} & r_{11} & \dots & r_{1w} \\ \dots & \dots & \dots & \dots \\ r_{h0} & r_{h1} & \dots & r_{hw} \end{pmatrix}$$

The RGB section can be thought of as a combination of three monochrome grayscale images. All three correspond to the red, green, blue input of the monitor, and as a result we can see a color image on the screen. Three monochrome shapes form an RGB composite image, and matrices can be called red, green, blue or image component. An image component class defines the range of its values. Accordingly, if the class is double, the range of values is the interval [0,1]. Likewise, uint8 is [0.255], uintl6 is [0.65535]. The bit that an RGB image uses to represent the pixel value of all of its components is called image depth. In most cases, the bits of the components are equal to each other, so if we take an 8-bit image, the image depth is 24 bits, and each component has 16,777,216 colors.

To implement this system, we create a matrix of 0, the width of which is equal to the length of the frame.

$$zeros(frameH, frsmeW, 3, 'uint8')=\begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

Frames are written into the prepared structure Zeros, as in the matrix set in the figure below.
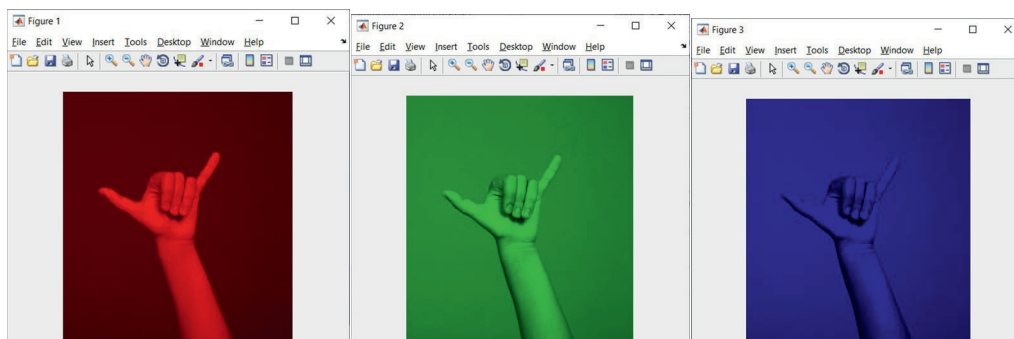
$$frame_0 = \begin{pmatrix} f_{00} & f_{01} & \cdots f_{0w} \\ f_{10} & f_{11} & \cdots f_{1w} \\ \cdots\cdots\cdots\cdots \\ f_{h0} & f_{h1} & \cdots f_{hw} \end{pmatrix} \quad frame_1 = \begin{pmatrix} f_{00} & f_{01} & \cdots f_{0w} \\ f_{10} & f_{11} & \cdots f_{1w} \\ \cdots\cdots\cdots\cdots \\ f_{h0} & f_{h1} & \cdots f_{hw} \end{pmatrix} \quad \vdots \quad frame_n = \begin{pmatrix} f_{00} & f_{01} & \cdots f_{0w} \\ f_{10} & f_{11} & \cdots f_{1w} \\ \cdots\cdots\cdots\cdots \\ f_{h0} & f_{h1} & \cdots f_{hw} \end{pmatrix}$$

With the following notation, we can separate three components: R, G, B frame.

$R = I; R(:,:,2) = 0; R(:,:,3) = 0;$

$G = I; G(:,:,1) = 0; G(:,:,3) = 0;$

$B = I; B(:,:,1) = 0; B(:,:,2) = 0;$

Next, to get a halftone frame, use the following statement

rgb image = $cat(c_{i, ir, ig, ib})$

It is important for us that the color components are in the same order in the formula. For example, if ci = 1, then the array is vertical, if ci = 2, then horizontal, and if ci = 3, then the third dimension is combined. That is, dim = 3 for an RGB image with three components. When the components match, we get a gray image.
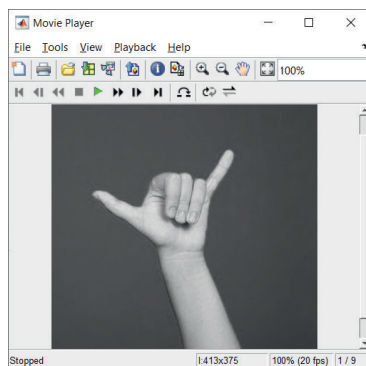
One of the most popular methods for obtaining a binary image is the Bernsen method. The Bernsen method consists of the following steps:

1. A regular square hole with an odd number of pixels goes through all pixels of the original image in a loop. Each step is min. And max.

2.   Found the average value Avg = (Min + Max) / 2.

3.   If the current pixel is greater than Avg <E, it will be white, otherwise it will be black. E is a user-defined constant.

4.   If the average value is less than the contrast threshold, the current pixel will be converted to the color specified in the «questionable pixel color» setting.

For each (x; y) pixel, the brightness limit is selected B (x; y) = $(B_{min}-B_{max})$ / 2, where $B_{min}B_{max}$ is the maximum or minimum pixel brightness of the pixel square (x; y), respectively. If the contrast level (the maximum difference between the minimum level) exceeds a certain limit, then the pixel is called either white or black. For all frames, this contrast threshold should be selected consistently and interactively.

$$\text{frame}(x; y) = (f_{min} - f_{max})/2, \quad f = \begin{cases} \text{frame}(x; y) < \varepsilon, & 0 \\ \text{frame}(x; y) > \varepsilon, & 1 \end{cases}$$

Strel is a flat morphological structural element that is an integral part of morphological expansion and erosion [12]. A planar feature is a two-dimensional area, two-dimensional or multi-dimensional, containing morphological calculations for true pixels, not false pixels. The center pixel of the structure element is defined as the source of pixels by the pixel of the frame being processed. We use the strel function to create the structure element.

1	1	1	1	0
1	1	1	1	0
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
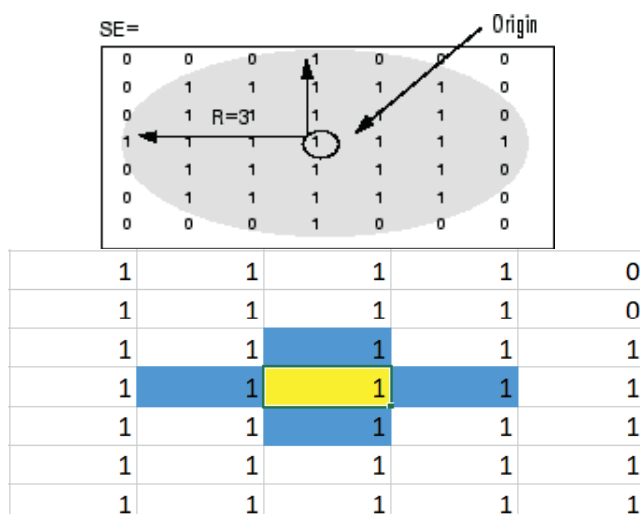
Take the center point and examine the top, bottom, left and right adjoining parts of the point.

[L,n] = bwlabel(____)

Determination of the number of one-component elements. SE = strel('octagon',r) creates an octagonal feature, measured along the horizontal and vertical

axes, that defines the distance from the origin of the feature to the sides of the octagon. must be a product not equal to r3.

[L,n] = bwlabel(____)

Also returning n, the number of BW added objects found. [label, numObj] = bwlabel (IopenedDisk, 4) Computes a set of characteristics for each area.

iRegion = regionprops (IopenedDisk, 'centroid')

The center of mass of the region is returned as a vector 1 Q. The first element is the horizontal coordinate (or x coordinate) of the center of mass of the Centroid. The second element is the vertical coordinate (or y coordinate). All other elements have Centroid dimensions. This figure shows the centroid and rectangle for the uncoordinated area. The area is composed of white pixels; the green frame is the junction box and the red dot is the centroid.
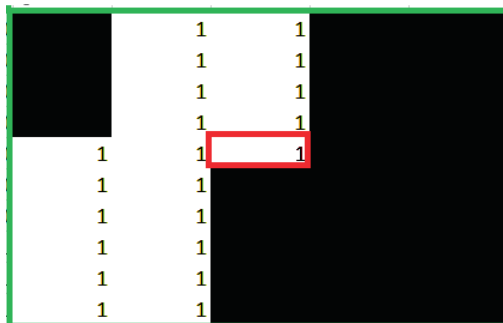
Another property of these methods that should be considered is the need to use a large number of standards. This is especially important in cases where dynamic gestures are recognized.

In case of directions of movements — left, right, up, down; palm orientation — left, right, up, down (Fig. 12, 13), we can solve this problem by installing two cameras to correctly read the configuration of the hand from the computer.

A palm isolated from a conventional camera directed to or from the speaker will look like the images in Fig. 10.

Figure 10. Configuration of the hand

When the palm is oriented left, right, up and down, a conventional camera will not be able to read the configuration of the hand correctly (fig. 11).

13

Figure 11. Configuration of the hand

1 H	ZN	PLRLUD	DULRS	Шыдау	спорт жейдесі	кіші	қанша?
Не үшін?Неге, Қабат,Аз ,Минут, Қаңтар, сәуір, тамыз, Сағат, бөтелке, бағдаршам							

Figure 12. The palm is directed to the right, left, up, down, one-handed static gesture classes

2 H, NTINT	ZN	PLRLUD	DULRS	не?	альбом	Жақсы	қайта-қайта
Жеңіл, Жанында, Қысқа, Өтті, Салмақ, Үлкен, құты,көпір							

Figure 13. The palm is directed to the right, left, up, down, two-handed static gesture classes

CONCLUSION

Thus, the constructed ontological model allows us to describe about -70% of all possible variations of gestures, for which effective recognition procedures can be proposed [8,9]. Recognizing the rest of the gestures requires different approaches, taking into account the specifics of the of gestures performance. We should also mention that in order to recognize gestures, certain rules to record gestures must be observed, in particular, there should be a monochromatic background, sign language speakers should be dressed in monochromatic clothes, contrasting with the color of the hands, etc.

Correlation methods of image recognition and direct comparison with the standard are likely to be used in computer vision due to their high resistance to local and background distortions of objects. The undoubted advantage of the approach is its universality in terms of taking into account the diversity of possible conditions arising while recognition of visual objects. Based on the Bernsen method, a system was created and an experiment was carried out — the recognition of most gestures varies in the range of 80–96%.

Saule Kudubayeva, Nurzada Amangeldy, Zakirova Alma, **Kazakh sign language recognition system based on the Bernsen method and morphological structuring**

References:

1. www.parlam.kz/ru/senate/press-center/article/35658. 2018/ Response of the RK Prime Minister to # 15–13-32 r. from February, 16,.
2. www.surdo.kz/
3. www.surdo-online.kz/
   Dimskis L. S. Let Us Study Sign Language. M.: «Academy» publishing center, 2002. 128 p.
4. *Zaitseva G. L.* Gesture speech. Dactylology. M.: Vlados, 2000. 192 p.
5. *Kudubayeva, S., Zhussupova, B., Aliyeva, G.* Features of the representation of the Kazakh sign language with the use of gestural notation: Compiling the dictionary of gestures of the KSL based on the notation of L. S. Dimskis. ISEMIS 2019, Nur-Sultan.
6. *Ya. A. Smekhun,* Ontologies in systems based on knowledge: Possibilities of their applications, May, 2016.
7. *Kudubayeva, S., Amangeldy N., Sundetpaeva A., Sarinova A.* The use of correlation analysis in the algorithm of dynamic gestures recognition in video sequence. ISEMIS 2019, Nur-Sultan.
8. *Kudubayeva, S., Ryumin, D., Kalghanov, M., Assem, A.*2016. Automated recognition system of statistic gestures via Kinect sensor. Application of information and Communication Technologies. Baku. Azerbaijan.
9. *Kudubayeva, S., Ryumin, D., Kalghanov, M.* The influence of the Kazakh language semantic peculiarities on computer sign language Proceedings of the International Conferences on ICT, Society, and Human Beings 2016, Web Based Communities and Social Media 2016, — Part of the Multi Conference on Computer Science and Information Systems 2016, 2016, c. 221–226
10. *Gonzalez R., Woods R.* 2006. Digital Image Processing
11. http://matlab.exponenta.ru/imageprocess/index.php
12. https://www.mathworks.com/help/images/ref/strel.html