Computer Technologies for Gestures Communication Systems Construction

Iu. Krak[✉]

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine
`krak@univ.kiev.ua, Yuri.krak@gmail.com`

Abstract. The approach to the development of informational computer technologies for the interactive learning of sign language which use 3D model human expart of sign language is proposed. Methods for modeling fingerspelling information and it's recognizing uses convolution neural networks are created. The implementation of these methods under the operating system Microsoft Windows and on the cross-platform technology have been developed. Methods and software for gestures modeling base on 3D human model and using Motion Capture technology are created.

Keywords: Sing language · Modeling · Recognition · Deaf people · 3D design

1 Introduction

Currently, the society tries to solve the problem of integration of disabled people. The same is true in the relations to the people with hearing impairment and the deaf-mutes. As a rule, sign language (SL) remains the main way of communication among those people [1]. In terms of SL, the movements of body and hands, face, eyes code the information, which is perceived visually.

Therefore the fundamental peculiarities of the SL are that the leading role in the communication belongs to the settings and the elements of gestures are performed and perceived simultaneously.

Gesture is the main unit of the SL. There are three parameters necessary for the description of gesture structure [2]: relations between the place of gesture performance and the body of a communicator, design of a hand demonstrating the gesture; trajectory of the hand. Today, existing systems for learning the sign language are oriented, mainly, to video recording of gestures. The efficient solution of the problem of SL investigation requires the creation of the technologies responsible for the computerization of this process. In fact, it is quite possible to perform such tasks due to the accessibility and modernity of the approach, considering pace of the development of graphic processors and technologies for the 3D graphics. As noted in the paper [3], development of the human-like avatars for gestures modeling with applicatications to creation of the SL teaching systems are very important and perspective. From the other hand avatar-oriented

approach to SL modeling has general character and can be use to realization any sign language. In this article it is proposed to develop the approaches given in the papers [4–6], and expand them for modeling based on 3D model of fingers of the human hand and the recognition of signs of the alphabet. Develop a system of sign communication based on 3D human model and implement it in the form of the cross-platform information technology.

2   Modeling and Recognition of the Sign Language Fingerspelling Alphabet

Dactyl (fingerspelling) SL is an integral part of sign communication systems and to show individual letters (dactylemes) of the alphabet is designed. Note, that for the main languages of the world the dactylemes are demonstration with the fingers of one human hand (right hand as rule). Necessary create mathematical methods and computer technologies for modeling and demonstration as separative dactylemes as well as continuous sequence of dactylemes for words showing. For given problems solving, investigations were conducted and base on it a new methods, algorithms and models of computer synthesis of fingerspelling alphabet were and effective broadcast of gestures animation via the Internet were created.

To create a software tools, a mathematical model of a simplified human skeleton was used. The hand is represented as a hierarchical structure of bones forming an acyclic directed graph. As mathematical model for fingerspelling alphabet modeling was used of a simplified human hand skeleton, were the hand is represented as a hierarchical structure of bones forming an acyclic directed graph. Based on the proposed mathematical model, an information model that contains a set of fingers vertices in the initial state, a sets of indices for representing triangles, a set of normals for each vertex, a hand texture coordinates describing the surface of the hand was created [4]. Hand visualization occurs base on through the spinning procedure, which it the most effective method of animation [4].

The developed methods and algorithms were implemented in the form of information technology (see Fig. 1), were indicated numbers mean next activities: 1 – area of displaying dactylemes alphabet; 2 – panel of displaying playback progress of dactylemes or words; 3 – input panel for words; 4 – list of letters of the fingerspelling alphabets; 5 – button «spell», the process of fingerspelling of input word begins when the button is clicked; 6 – panel to demonstrate the verbal description of a hand configuration that correspond to the current displayed dactylemes; 7 – panel to display written letter and a picture that correspond to the current displayed dactylemes; 8 – indicator of a location of a hand rotation; 9 – define the pace of fingerspelling [5].

It should be noted that on the basis of the created information technology software tools can be developed for the simulation of any dactyal sign language from a one-handed finger alphabet. For example, on a Fig. 2 shows the work of the created software for fingerspelling alphabet modeling of the Kazakh sign language.
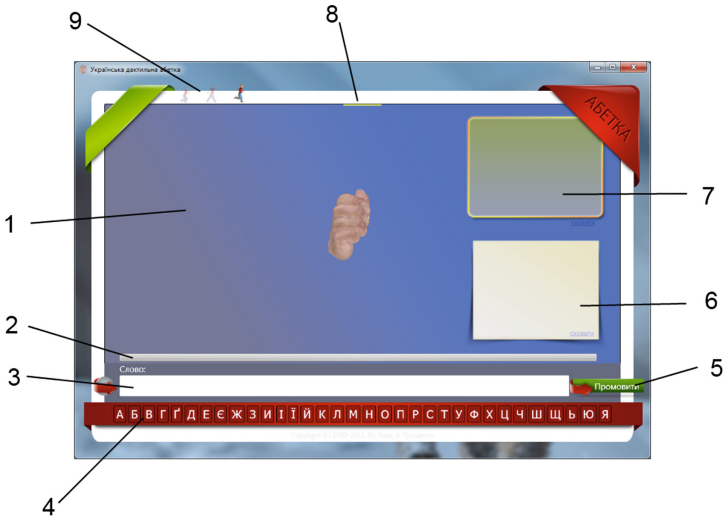
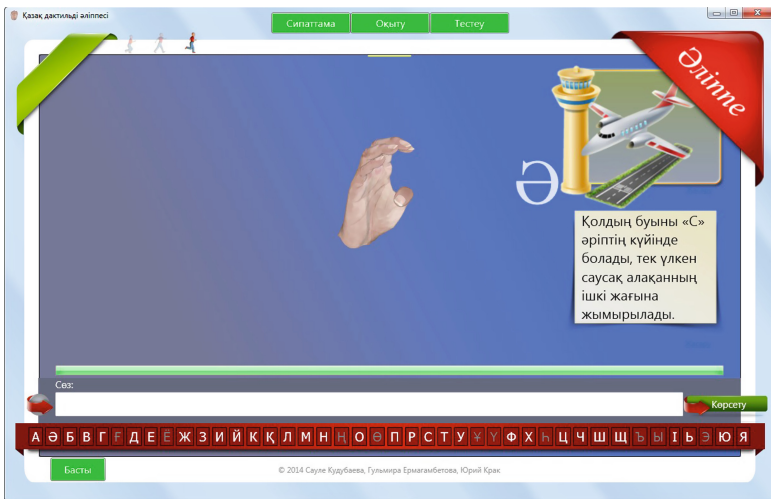**Fig. 1.** The main activities of a software for fingerspelling alphabet modelling

Fig. 2. Fingerspelling alphabet modeling of Kazakh sign language

The created software package runs in the OS Windows environment and over the Internet. It is important to note that the parallelization of frame preparation algorithms allows you to play animation with an increased frame rate when using multi-core processors, thereby increasing the readability of the source files of applications. As advantage of created approach is it transfer a controlled media stream through the Internet with the adaptation to the width of the data

transmission channel is created. This allows to create software media applications that can be accessed by all Internet users.

Note that created software was oriented on OS Windows and for running proposed information technology in any operation systems the cross-platform software dactylemes alphabet modeling and recognition was created [6]. It very important that created software should solve the problem of running on existing platforms using cross-platform development without implementing the functionality for each platform separately.

Infological model of the cross-platform technology which is composition of three cross-platform modules is demonstrate in Fig. 3: 3D hand model and user interface (which are implemented with cross-platform framework Unity3D [7]), and software for gesture recognition (implemented with cross platform framework Tensorflow [8]). The main functionality is implemented with C# and Python and runs on desktop OS (MacOS, Linux, Windows) and on mobile OS (Android, iOS).
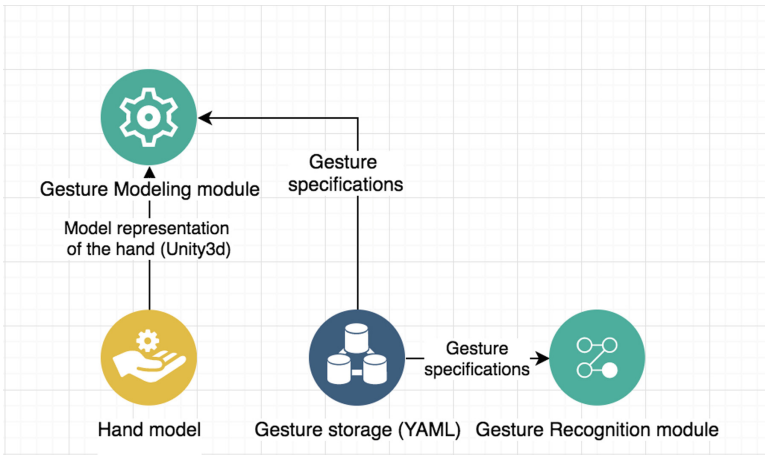

Fig. 3. Infologic model of cross-platform gesture communication technology

3D human hand model module is cross-platform and provides hand model representation for gesture recognition module. Function renderer for human hand receives hand model representation and gesture specifications from gesture storage module, and provides a high-polygon rendered hand model. Module for gesture learning and module for gesture modification are implemented with cross-platform Unity3D, both taking as input results of hand model renderer. Module for gesture modification provides updated gesture specifications and transmits them to gesture storage. Note that for the storage of a gesture, the BVH file format was used. For gesture recognition is proposed to be implemented with Tensorflow framework and receives as input data: model of the 3D human hand, gesture specifications and input from usual camera [6].

Note that for Unity3D framework is able to effectively reproduce a realistic human hand 3D model which consists of more than 70,000 polygons (see Fig. 4). Based on the anatomy of the hand within Unity3D hand model was developed with 25 degrees of freedom, four of them located in the metacarpal-carpal joint, to the little finger and thumb to provide movement palm. The thumb has 5 degrees of mobility, middle and index fingers have 4 degrees of mobility (metatarsophalangeal joint with two degrees of mobility, and the distal and proximal interphalangeal joints each have one) [9].

Fig. 4. 3D model for gestures demonstration under iOS platform

Modules for gesture modeling and recognition are developed with cross-platform tools (frameworks based on Python, C++) can be embedded into information and gesture communication cross-platform technology. Multiple approaches were considered as an approach for gesture recognition [10,11].

Thus SL recognition tools are consists of taking an input of video stream, extracting motion features that reflect SL linguistic terms, and then using pattern mining techniques or machine learning approaches on the training data. For example, in the paper [9] propose a novel method called Sequential Pattern Mining (SPM) that utilizes tree structures to classify characteristic features.

Convolution neural networks have such advantages [6]:

(1) no need in hand crafted features of gestures on images;
(2) predictive model is able to generalize on users and surrounding not occurring during training;
(3) robustness to different scales, lighting conditions and environment.

As a result of experimental studies F1-score of gesture recognition on test dataset of 0.2 fraction of whole dataset for 100 image samples is equal 0,6, for 200 image samples is equal 0,74, for 500 image samples is equal 0,8 and for 1000 image samples is equal 0,82.

Usage of cross-platform neural network framework such as Tensorflow allows to implement gesture recognition as a cross-platform module of proposed technology and serve trained recognition model on server or transfer it to the device [6].

Created software [6] offered and used in the implementation of information technology is cross-platform and operates unchanged regardless of operating system (Windows, Linux, Android, iOS), CPU type (x86, arm), and the type of hardware (mobile or stationary device).

With its cross-platform build system Unity3D it is possible to create applications for each platform without porting or changing the original code.

As there are no specific hardware requirements for information technology for modeling SL, there are objective obstacles for performance speed of older generations devices. To overcome this problem, the following adaptive approach to information technology was proposed as shown on Fig. 5.

Further modules implementation will leverage from existing cross-platform technology. Modules for gesture learning and recognition, developed with cross-platform technologies (Python, Tensorflow) will be embedded into information and gesture communication cross-platform technology. In case of the mobile app (iOS, Android) or application on the device with a stationary operating system (Windows, Linux), during installation on the device information technology analyses the existing hardware and, depending on its capacity, conducts a series of adjustments.
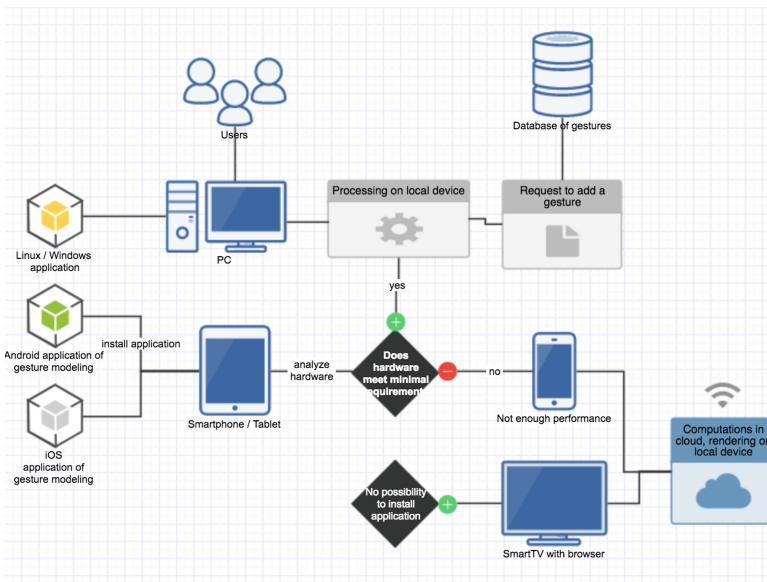

Fig. 5. Cross-platform and adaptive execution of information technology scheme

The effectiveness of the proposed approach is shown in building cross-platform technology for modeling and recognition of Ukrainian fingerspelling (dactyl) alphabet [4,6,11]. Based on this technology, training programs for any one-handed fingerspelling alphabet can be created.

3   Information Technology for Simulation of Sign Language

The general scheme for SL uses for communication with deaf people is shown on a Fig. 6 and will provide the following features [5]:
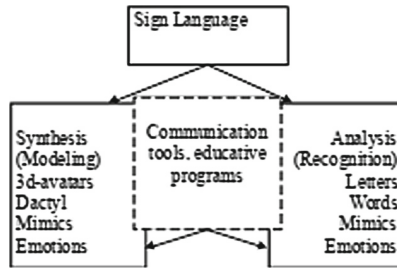

Fig. 6. Sign language communication technology

– a module for translation of the usual text into the SL (text-to-gesture); the module will provide demonstration animation of a common and official SL by presenting the output on a 3D human model;
– mimics and animation (with regard to emotional components) during the pronunciation process;
– lips reading module for recognition of the text being pronounced.

For the implementation of the suggested concept of computer-aided non-verbal communication, a series of research works have been made and the appropriate software has been developed (see, for example papers [2,5,10]). For the 3D model for SL animation synthesis, the geometrical classes of vector-based gestures are described. These classes were formed using Motion Capture (MoCap) technology [12]. MoCap is a technology for retrieving real-world 3D coordinates using multiple video streams recorded from different viewpoints. Then the coordinates are used to determine values in the 3D mathematical model. The key frames are determined by using tracking technology [5,13].

BVH file format for the storage of a gesture was used [5].

Method of the 3D model creation of real human using Motion Capture technology is show on the Fig. 7. Note that the sign language interpreter was used for gestures demonstration (on the Fig. 7 it a left hand picture) and his movements were transferred to the 3D model (on the Fig. 7 it a right hand picture).

Fig. 7. 3D model creation for gestures demonstration

The main advantage of proposed approach is technology for 3D human model creation which maximal similar on real human (on the Fig. 7 it a central picture).

Note that gestures obtained in this way (from sign language interpreter demonstration) can be used as a standard for displaying gestures of a particular sign language.

For the input text preprocessing, the appropriate informational technology was created, which considers the stress location for each word, specifies its normalized word form; contains synonyms and idioms. The model is represented as a set of tables in a relational database along with a set of stored procedures which implement all the required functionality.

For the implementation of visualization and pronunciation feature of a custom text, the appropriate synthesizer has been created. It allows creating the voice equivalent of a custom text using different voices and voice characteristics (volume, distance).

For the complex verification of the suggested technology the appropriate software has been created (Fig. 8). It is used for translation of a custom text into the SL.

The created approach which base on uses the following algorithm for SL synthesis:

(1) a speech equivalent is synthesized for the input text;
(2) the input text is parsed as separative words;
(3) create speech synthesis for the input text;
(4) for each word its normalized form (infinitive) is found by performing a look up in the database;
(5) for each normalized word form a gesture is looked up (represented as a 3D model human movements);
(6) in case the gesture is not found, the word will be shown using fingerspelling alphabet.

The created information technologies was implemented in the specialized schools for deaf people for sign language learning was implemented and effectiveness ones was demonstrated.

Fig. 8. Computer system for SL modelling and learning

4   Conclusion

In the paper suggests a complex approach to creation of the informational technologies for communication with deaf people using sign language: gestures creation, modeling, animation and recognition. Convolution neural networks are used for recognition dactylemes and experiments good recognition results are demonstrated.

As novel approach to the 3D human model creation for sign language demonstration was create model maximal similar on real human. The technology uses 3D models of human body, hands and fingers. Cross-platform technology proposed for solves the problem of execution on the existing multiple platforms without implementing functionality under each platform separately is created. This is one of the important advantages of the proposed approach to simulation and recognition of sign languages.

Thus, in this paper it was shown the effectiveness of the technologies 3D modeling sign language for creation using cross-platform common tools for modeling and recognition sign language. Effectiveness created technologies was demonstrated for modeling and recognition of Ukrainian sign language. Information and gesture communication technology was developed with further scaling capabilities in mind for gestures of other languages such as Polish sing language, Kazakh sign language, English sign language etc.

References

1. Brentari, D. (ed.): Sing Languages. Cambridge University Press, Cambridge (2010)
2. Stokoe Jr., W.C.: Sing language structure: an outline of the visual communication systems of the American deaf. University of Buffalo (1960)
3. Smith, R., Morrissey, S., Somers, H.: HCI for the deaf community: developing human-like avatars for sign language synthesis. In: Proceedings of the 4th Irish Human Computer Interaction Conference, Dublin, pp. 129–136 (2010)
4. Kryvonos, Iu.G., Krak, Yu.V., Barchukova, Yu.V., Trocenko, B.A.: Human hand motion parametrization for dactylemes modeling. J. Autom. Inf. Sci. **43**(12), 1–11 (2011)
5. Krak, Iu., Kryvonos, Iu., Wojcik, W.: Interactive sytems for sign language learning. In: 6th International Conference on Application of Information and Communication Technologies (AICT), pp. 114–116 (2012)
6. Kondratiuk, S., Krak, Iu.: Dactyl alphabet modeling and recognition using cross platform software. In: IEEE Second International Conference on Data Stream Minning & Processing (DSMP), pp. 420–423 (2018)
7. Unity3D framework. [Electronic resource] www.unity3d.com. Accessed 20 Apr 2018
8. Tensorflow framework documentation. [Electronic resource] www.tensorflow.org/api/. Accessed 20 Apr 2018
9. Ong, E.-J., et al.: Sign language recognition using sequential pattern trees. In: IEEE Conference on IEEE Computer Vision and Pattern Recognition (CVPR), pp. 2200–2207 (2012)
10. Kryvonos, I.G., Krak, I.V., Barmak, O.V., Shkilniuk, D.V.: Construction and identification of elements of sign communication. Cybern. Syst. Anal. **49**(2), 163–172 (2013)
11. Krak, Yu.V., Golik, A.A., Kasianiuk, V.S.: Recognition of dactylemes of Ukrainian sign language based on the geometric characteristics of hand contours defects. J. Autom. Inf. Sci. **48**(12), 1–11 (2011)
12. Menache, A.: Understanding Motion Capture for Computer Animation and Video Games. Morgan Kaufmann, Burlington (2000)
13. Avidan, S.: Support vector tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition 1, Kauai, Hawaii, pp. 84–191 (2001)