IEEE Second International Conference on Data Stream Mining & Processing
August 21-25, 2018, Lviv, Ukraine

Dactyl Alphabet Modeling and Recognition Using Cross Platform Software

Sergii Kondratiuk
Dept. of Theoretical Cybernetics, faculty of CS and Cybernetics
Taras Shevchenko National University
Kyiv, Ukraine
kondratiuk@univ.net.ua

Iurii Krak
Dept. of Theoretical Cybernetics, faculty of CS and Cybernetics
Taras Shevchenko National University
Kyiv, Ukraine
kondratiuk@univ.net.ua, krak@univ.net.ua

***Abstract*—The technology, which is implemented with cross platform tools, is proposed for modeling of gesture units of sign language, animation between states of gesture units with a combination of gestures (words). Implemented technology simulates sequence of gestures using virtual spatial hand model and performs recognition of dactyl items from camera input. With the cross platform means technology achieves the ability to run on multiple platforms without re-implementing for each platform.**

Keywords— cross platform; sing language; dactyl modeling; gesture recognition

I. Introduction

Communication via gestures is one of the three main means of transmission of information between people, among character (text) and voice (speech) communication. Sign language is usually used by people with hearing disabilities to communicate with each other and with their environment, increasing the number of people who need to know sing language. Note that sign language is universal in the sense that you can send text information via gestures and in case if certain words don't have corresponding gestures (names, cities, areas, etc.), they may be shown via letters one-by-one using dactyl alphabet. Modern hardware is able to collect information fast and almost without restriction, process data both in cloud computing (model, which provides a universal, easy access on demand through the network to the virtual cluster computing resources) [1] and locally on the device, and through data channel processing results are returned to the user. All this is also true for sign language. Sings can be stored and reproduced via a variety of devices and platforms, stationary or mobile, high performance or energy efficient. The actual problem is the reproduction of sign language on all these platforms, for further usage by people with hearing disabilities in particular and everyone in general. Deployment of a single unified technology on various platforms (android, ios, windows, linux, web) without need to port it or to implement it under each platform is a major problem. One way of solving the stated problem of visualization and reproduction of sign language is cross platform software development. Unlike single-platform technologies that operate only on a specific platform under which they were developed, "cross platform software provides the ability to perform on more than one platform with identical (or nearly identical) functionality" [2]. The term "platform" in this context may refer to one of or a combination of several definitions: 1) the type of operating system (such as Microsoft Windows, Mac OS X, Linux, Solaris, Android, iOS); 2) processor type (such as x86, PowerPC, ARM); 3) the type of hardware (e.g., mainframe, workstation, personal computer, mobile device) [2]. Cross-platform technologies are on a par with the platform independent technologies (those that can operate on any platform, such as Web application) [2] and cross-platform virtual machines (technologies that support individual processes or systems, depending on the level of abstraction at which is virtualization) [3]. In this article the proposed solution of the problem is via cross platform development, taking into account characteristics of different classes of devices (such as hardware, CPU power, amount of memory, presence on the Internet) and setting the number of polygons of the three-dimensional hand model and gesture animation step. Gesture modeling and gesture recognition is performed via cross platform means as a part of proposed communication technology.

II. Materials and Methods

Sign modeling is a problem that is considered both independently and as part of the problem of modeling and recognition of gestures and thus as a technology learning and evaluating sign language. One of the systems to display the sign language is American Sing Language Online Dictionary [4], which consists of a video database of words and phrases displayed via sign language. Control via gestures is an actual problem in the development of platform independent human-computer interaction [5]. These developments were involved in a number of commercial agencies [6, 7], but the systems they propose are configured to pre-determined number of gestures, and therefore do not solve the problem of modeling sign language. Also all of them lack functionality of gestures recognitions, thus not allowing to evaluate quality of sing language performed by a user. Creating a model hand is the first step in the task of sing language modeling. In their work [8], authors analyze existing approaches of hand modeling, which are divided into two main groups: spatial and temporal. Former consider the characteristics of different positions for the hand gestures, while latter refer to the description of the dynamics of gestures. Modeling hands in the spatial area can be completed in two and three dimensions. In [9] proposed system by authors is able to simulate sign animations for a given text. As a part of this system a statistical model is used to analyze input text and generative algorithm is used when creating the appropriate simulated kinematics of sign animations. Within the article, the authors have provided ANVIL tools for input text annotation, gesture generator NOVA, and DANCE library developed in [10] is used for gesture animation. The system is built on the Microsoft Windows platform and x86 processor. In [11] authors discuss the modeling of virtual character for spatial reproduction of sign language on the platform of Microsoft Windows. In [12], authors developed a system of signed language training, which consists of two modules - gesture demonstration module via video and gesture recognition module (required gloves), based on Hidden Markov Model. The training system is based on Microsoft Windows platform and x86 processor. Gesture recognition for mobile platforms is developed in [13], but gesture modeling on mobile devices is not performed.

III. Formulation of Problem

The proposed technology should perform modeling of sign units (morphemes [14]) of sign language, and reproduce animation of gestures structures (words, sentences) via state transitions between shown units using spatial virtual model hand. The proposed technology should perform recognition of sign language based on camera input from the device in order to evaluate sing language performed by user. The technology should be a combined solution for learning sing language via gesture modeling and recognition. Technology should solve the problem of running on existing platforms using cross platform development without implementing the functionality for each platform separately. The effectiveness of the proposed approach is shown in building cross platform technology for modeling and recognition of Ukrainian dactyl alphabet.

IV. Proposed Approach

To address the modeling of sign language and perform animation of sign structures using spatial virtual model hand the cross platform technology based on cross platform framework Unity3D [15] is proposed. Cross platform framework Unity3D is also used for the user interface, both libraries and technology are implemented with programming language C#. Proposed tools can solve the problem of running the technology on multiple existing platforms. The novelty of the proposed technology is that it is cross platform and has customizable level of polygons for three dimensional hand model and animation step for gesture transitions. This allows to run proposed technology without changes on multiple platforms (different types of processors, operating systems and hardware. Advantage of cross platform technology over technologies developed for a single platform is that there is no need to modify or re-implement the functionality already available for other platforms (porting) [2], which speeds up the process of developing and deploying technologies, and increases the number of potential users. The advantage of cross platform technology over cross platform virtual machine emulators is performance speed and absence of necessity to install additional software (software dependencies).

V. Infological Model

The core of the technology is composition of three cross-platform modules as shown in Figure 1: three dimensional hand model (which is implemented with cross platform framework Unity3D), user interface (implemented also with cross platform framework Unity3D) and gesture recognition module (implemented with cross platform framework Tensorflow [16]. Core functionality is implemented with C# and Python and runs on desktop OS (MacOS, Linux, Windows) and on mobile OS (Android, iOS). Hand model module is cross platform and provides hand model representation for gesture recognition module. Hand renderer receives hand model representation and gesture specifications from gesture storage module, and provides a high-polygon rendered hand model. Gesture learning module and gesture modification module are implemented with cross platform Unity3D, both taking as input results of hand model renderer. Gesture modification module provides updated gesture specifications and transmits them to gesture storage. Gesture recognition module is proposed to be implemented with Tensorflow framework and receives as input hand model, gesture specifications and input from camera.

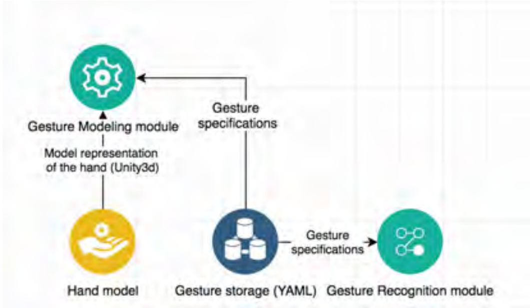

Fig. 1. Infologic model of cross platform gesture communication technology

VI. Gesture Modeling & Recognition

The hand model which is built in gesture modeling module has 27 bones, 8 of the bones are in wrist, 3 are in the thumb (one metacarpal and 2 phalanx) and 4 metacarpus and 12 phalanges are in other fingers [17]. Each bone is connected to the other through different types of joints. Designing your own cross platform engine for simulating the hand is non-trivial task, thus as the core technology for modeling three-dimensional hand model and gesture animations between morphemes cross platform framework Unity3D was selected.

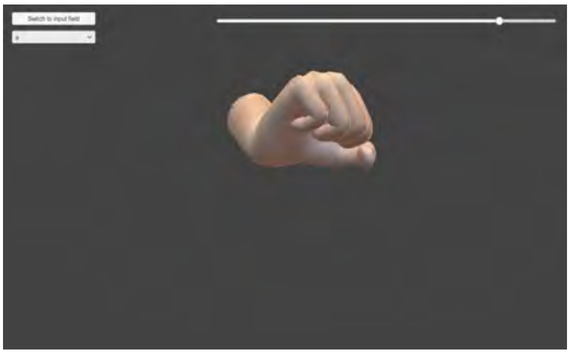

Fig. 2. Gesture modeling under iOS platform.

Unity3D framework is able to effectively reproduce a realistic hand model which consists of more than 70,000 polygons as shown in Fig. 2. Based on the anatomy of the hand within Unity3D hand model was developed with 25 degrees of mobility, 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. Gesture learning and gesture recognition modules, 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. Automatic sign language recognition can be approached similarly to speech recognition, with signs being processed similar to phones or words. Most previous work has used approaches based on hidden Markov models (HMMs) [19, 20]. Conventionally, sign language recognition consists of taking an input of video sequences, extracting motion features that reflect sign language linguistic terms, and then using pattern mining techniques or machine learning approaches on the training data. For example, Ong et al. propose a novel method called Sequential Pattern Mining (SPM) that utilizes tree structures to classify signs [21]). Convolutional Neural Networks (CNNs) have shown robust results in image classification and recognition problems, and have been successfully implemented for gesture recognition in recent years. In particular, deep CNNs have been used in researches done in the field of sign language recognition, with input-recognition that utilizes not only pixels of the images. With the use of depth sense cameras, the process is made much easier via developing characteristic depth and motion profiles for each sign language gesture [22]. Multiple existing researches done over various sing languages show that CNNs achieve state-of-the-art accuracy for gesture recognition [23, 24, 25].

Convolutional neural networks have such advantages: no need in hand crafted features of gestures on images; predictive model is able to generalize on users and surrounding not occurring during training; robustness to different scales, lightning conditions and occlusions. Although, selected approach has couple of disadvantages, which may be overcome with a relatively big dataset (1,000 images for each gesture, among more than 10 people of different age, sex, nationality and images taken under different environment conditions and scales): need to collect a rather big and labeled gesture images dataset; black-box approach which is harder to interpret. F1-score of gesture recognition on test dataset of 0.2 fraction of whole dataset is shown at Table 1 (based on number of train samples per class).

TABLE I.
F1-SCORE BASED ON NUMBER OF IMAGE SAMPLES PER CLASS

Image samples	**F1-score**
100	0.6
200	0.74
500	0.8
1000	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. Sample dataset chunk shown in Fig. 3.

VII. APPLICATION OF CROSS PLATFORM TOOLS

The hand model which is built in gesture modeling module has 27 bones, Due to selected cross platform implementation tools, the proposed technology solves the problem of executing on multiple platforms without the implementation under each platform separately. Software 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).

Fig. 3. Sample of dataset

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 sign language, 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. 4.

Further modules implementation will leverage from existing cross-platform technology. Gesture learning and gesture recognition modules, 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 analyzes the existing hardware and, depending on its capacity, conducts a series of adjustments: 1) number of polygons of the hand model changes to priority for performance speed; 2) during rotation hand model changes pitch angle at which it rotates, with priority for speed. If the available hardware does not meet the minimum requirements of information technology, the user is given the recommendation to choose “online” mode, in which the calculation is not performed on hardware.

VIII. CONCLUSION

The proposed technology is built with cross platform tools for gesture modeling, gesture transitions animation and gesture recognition. The technology uses virtual spatial model of hand. With the help of cross platform development, the technology solves the problem of execution on the existing multiple platforms without implementing functionality under each platform separately. Thus, it was shown the effectiveness of the technologies built using cross platform tools, for example modeling and recognition elements of dactyl Ukrainian alphabet sign language. Information and gesture communication technology was developed with further scaling capabilities in mind for gestures of other languages alphabets. To implement this idea, the validation mechanism of new gestures to the common database can be applied. Cross platform information and communication technology and standardized protocol and data format (YAML) allows a range of solutions for remote computing using cloud computing, Web servers, local servers using a single sign database PostgreSQL [26]. The gesture communication technology can be augmented with other cross platform modules, such as gesture recognition and gesture learning modules.

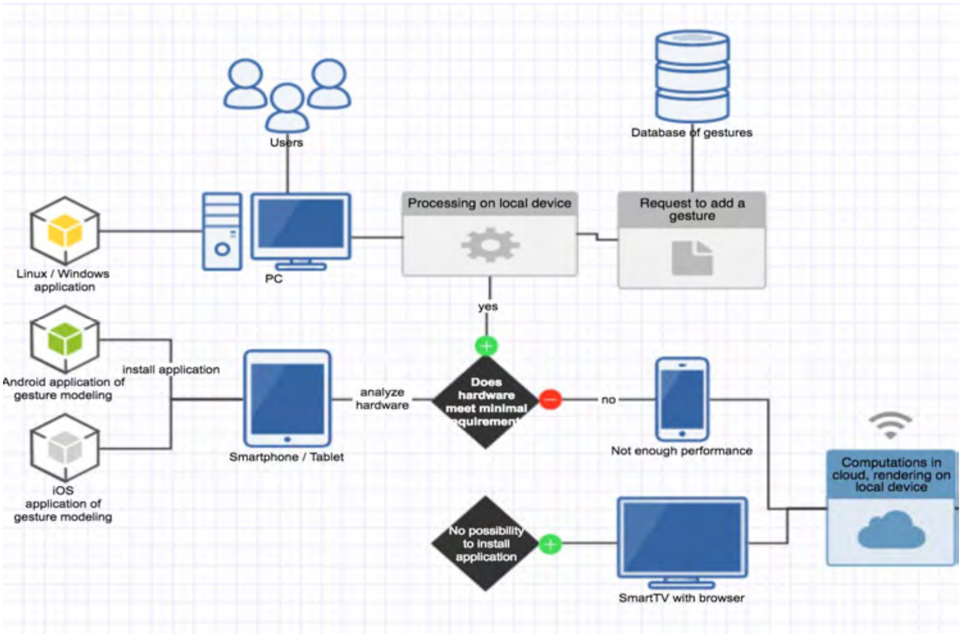

Fig. 4. General scheme of cross-platform and adaptive execution of information technology.

REFERENCES

[1] P. Mell and T. Grance, "The NIST Definition of Cloud Computing," (Technical report), National Institute of Standards and Technology: U.S. Department of Commerce, pp. 1-7, September 2011. doi:10.6028/NIST.SP.800-145. Special publication 800-145.

[2] The Linux Information Project, Cross-platform Definition.

[3] J. Smith, and R. Nair, "The Architecture of Virtual Machines," Computer, vol. 38, no 5, pp. 32–38, 2005.

[4] ASL Sing language dictionary www.signasl.org/sign/model

[5] L. A. Graschenko, A. P. Fisun and et., Teoreticheskie i prakticheskie osnovyi cheloveko-kompyuternogo vzaimodeystviya: bazovyie ponyatiya cheloveko-kompyuternyih sistem v informatike i informatsionnoy bezopasnosti: Monografiya / Red. A. P. Fisun. Orel: OGU, 2004. Dep. v VINITI 15.10.2004 g. # 1624 — V2004

[6] Samsung TV Gesture book www.samsung.com /ph/ smarttv/ common/guide_book_3p_si/waving.html

[7] Apple Touchless Gesture System for iDevices http://www. patentlyapple.com/patently-apple/2014/12/apple-invents-a-highly-advanced-air-gesturing-system-for-future-idevices-and-beyond.html

[8] R. Z. Khan, I. A. Noor, "Comparative study of hand gesture recognition system," Natarajan Meghanathan, et al. (Eds): SIPM, FCST, ITCA, WSE, ACSIT, CS & IT 06, pp. 203–213, 2012.

[9] M. Neff, M. Kipp, I. Albrecht, and H.-P., "Seidel Gesture Modeling and Animation by Imitation," Technical Report MPI-I-2006-1-005, Max-Planck-Institut Informatik, Saabrucken, Germany, 2006.

[10] A. Shapiro, D. Chu, B. Allen, and P. Faloutsos, "A Dynamic Controller Toolkit," Sandbox '07 Proceedings of the 2007 ACM SIGGRAPH symposium on Video games, , San Diego, California, pp. 15-20, August 04 - 05, 2007

[11] Iu. G. Kryvonos, Yu. V. Krak, Yu. V. Barchukova, and B. A.Trocenko, "Human Hand Motion Parametrization for Dactilemes Modeling," Journal of Automation and Information Sciences, vol. 43, no. 12, pp.1-11, 2011.

[12] O. Aran, I. Ari, A. Benoit, A. Huerta Carrillo, F.-X. Fanard, P. Campr, L. Akarun, A. Caplier, M. Rombaut and B. Sankur, "Sing language tutoring tool," 2006 eNTERFACE'06, Dubrovnik, Croatia. Final Project Report, July 17 – August 11, pp. 1-11, 2006.

[13] J.L. Raheja, A.S Sadab and A. Chaudhary, "Android based portable hand sign recognition system," Ed: A. Chaudhary, Recent Trends in Hand Gesture Recognition, GCSR, vol. 3, pp. 1-18, 2015. DOI: 10.15579/gcsr.vol3.ch1,

[14] W. C. Stokoe, "Sign Language Structure," An Outline of the Visual Communication Systems of the American Deaf., p.61-67, 1960.

[15] Unity3D framework www.unity3d.com

[16] Tensorflow framework documentation www.tensorflow.org/api/

[17] R. Tubiana, J. Thomine, and E. Mackin, Examination of the hand and wrist. 2nd ed. Martin Dunitz, 1996. ISBN: 1853175447/1-85317-544-7. Publisher: Informa Healthcare

[18] YAML – The Official YAML Web Site www.yaml.org

[19] O. Koller, J. Forster, and H. Ney, "Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers," Computer Vision and Image Understanding, vol. 141, pp. 108–125, 2015.

[20] P. Dreuw, D. Rybach, T. Deselaers, M. Zahedi, and H. Ney, "Speech recognition techniques for a sign language recognition system," In Interspeech, Antwerp, Belgium, ISCA best student paper award Interspeech 2007, pp. 2513-2516, August 2007.

[21] Eng-Jon Ong et al. "Sign language recognition using sequential pattern trees," 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2200–2207, 2012.

[22] A. Agarwal, and M. Thakur, "Sign Language Recognition using Microsoft Kinect," IEEE International Conference on Contemporary Computing, pp. 181-185, 2013.

[23] L. Pigou, S. Dieleman, P.-J. Kindermans, and B. Schrauwen, "Italian Sign language: Sign Language Recognition Using Convolutional Neural Networks. ELIS, Ghent University, Ghent, Belgium, 2015

[24] Brandon Garcia, and Sigberto Alarcon Viesca, American Sign language: Real-time American Sign Language Recognition with Convolutional Neural Networks Stanford University Stanford, CA, pp. 225-232, 2015.

[25] V. Bobic, "Hand gesture recognition using neural network based techniques," School of Electrical Engineering, University of Belgrade, 2016.

[26] PostgreSQL official web site www.postgresql.org