CONSTRUCTION OF A REALISTIC MOVEMENT ON THE 3D HUMAN MODEL FOR STUDYING AND LEARNING SIGN LANGUAGE

Iurii Kryvonos, Iurii Krak, Anton Ternov

Abstract: *Complex information technology for communications gesture and mimicry modeling and visualization is created.*

Keywords: *modeling, sign language, computer system, visual-speech synthesis and analysis.*

ACM Classification Keywords: *I.2.8 Problem Solving, And Search H.1.1 Systems and Information*

Introduction and problem statement

Among three ways that people communicate (in text, voice, gesture), the important and under-researched is the sign language, which is the basis of communication for deaf people and people with hearing disabilities [Belikov]. Because the information in sign language is transmitted through hands movements, facial expressions and articulations, the main visual tools for learning such a language are a photographic images and video of gestures. On the one hand the use of text books with a verbal description of signs leads to a different interpretation of movements and gestures by each person through his own interpretation of information, which further complicates the understanding and expression the meaning of sign communication. On the other hand, exclusive usage of photos and video of a gesture for the development of modern educational and communicational computer systems is quite difficult and not always possible, since a photographic image does not show the desired dynamics of a gesture, video does not allow to see a gesture from the other side. In addition, simple concatenation of videos does not result in a smooth signing.

Mentioned difficulties in usage in classical approaches to the learning of gestures lead to the need to develop more efficient technologies that would allow to create new computer systems for modeling and learning a sign language. One of the basic approaches is to develop a comprehensive information technology which includes the functionality of the synthesis of movements of a sign language on the spatial model [Krivonos].

It is required to implement an animation capabilities of the communication process with the help of sign language using a virtual human model, to establish appropriate information and mathematical models. These researches aim to build a model for fixing morphemes (the minimal meaningful units) of a sign language; to create technology and software to receive, save and play gestures within the framework of the model; to suggest algorithmic solutions a) to calculate the typical human hand trajectories, b) to model body movements in transitions from one gesture to another, and c) to animate facial expressions and articulations of a human face.

The spatial model of the person to fix the units of sign language

Animation of a gesture can be considered as an animation with appropriate frequency of different states of simplified skeletal model, which recreates a skeleton of a real person. It can be formalized as a hierarchical structure consisting of units connected by a kinematics pairs of the corresponding class [Kirichenko]. For example, modern three-dimensional modeling packages (Poser, 3D Studio Max) allow to generate an animation using information about changes in the respective angles of bones rotation.

A gesture can be formally described by a set H which is reflected a simplified human skeleton, Euler angles M and an order of their application. It can be used to build the corresponding bones of the skeleton with a change in time (discrete, with appropriate frequency, e.g.,1 / 30 sec):

$$H = \{H_i : H_i = \{k, d_i, M_i \in M\}\}, \quad (1)$$

where H_i – i-th bone in a skeleton ($i = \overline{0, N-1}$, N – number of bones in a skeleton); k – index of a parent bone; $d_i = [x_i, y_i, z_i]^T$ – coordinates of the point of the end of the bones in the coordinate system associated with the beginning of this bone;

$$M = \{M_i : M_i = (order_i, \theta_i)\}, \quad (2)$$

where M_i – values of Euler's angles and the order of rotations over time for the i-th bone; $order_i \in \{1, \cdots 6\}$ – an order of rotations implementation around the respective axes for the i-th bone (1-XYZ, 2-XZY, 3-YXZ, 4-YZX, 5-ZXY, 6-ZYX); $\theta_i = (\theta_i^j)$, $\theta_i^j \in \{(\varphi_{i\,X}^j, \varphi_{i\,Y}^j, \varphi_{i\,Z}^j)\}$ – set of Euler's angles for the i-th bone ($j = \overline{0, K-1}$, K – number of frames to play the gesture at a given frequency).

To obtain a set of angles (2) that characterize the change in the position of bones from the initial state of the skeleton motion capture technology is used [Lander]. Based on this approach corresponding software has been created. With the help of it capture, digitization and fixation of gestures can be done.

Modeling of facial expression and articulation of a human face during pronunciation

To display the facial expression and articulation of a person on a spatial model it's necessary to construct a mathematical model in which it is possible to identify differences in the position of human lips contour for the construction of a system to the proper articulation. The mathematical model must include the implementation of the possibilities of creating a visual alphabet for a language, visems analysis and applying these results to arbitrary models of the human head.

For the synthesis of the mathematical model let's move from the space of photographic images of the human face in the process of pronunciation to the vector space of characteristic parameters in order to display a human face on the spatial model. Such a transition is implemented in two phases:

1. Detection the inner contour of lips on the image

$$\mathrm{Im}\,L \to D, \tag{3}$$

where $\mathrm{Im}\,L = \{I_k : I_k \in FSV\}$ – ordered set of key frames of the video stream FSV (*Face Speech Video*), formed when shooting facial expressions on the human face, namely the position of lips in pronunciation of the words ($k = \overline{1,K}$ – ordinal index of the frame in the chosen sequence, where K – number of key frames); $I_k = \{col_{ij}^k\},\ i = \overline{1,m},\ j = \overline{1,n}$ – image of size $m \times n$ of a face with a mimics state of lips during articulation, m and n – length and width of image I_k respectively; $col_{ij}^k = I_k(i,j)$ – color of a pixel in system RGB with coordinates (i,j) on a image I_k; $D = \{D_k : D_k = \{d_{top}^k, d_{bot}^k\}\}$ – a set of lips contour, D_k – pair of point curves – lips contours (top d_{top}^k and bottom d_{bot}^k) for k -th frame.

2. Approximation with irregular basis splines (NURBS) [Piegl] of the inner contour of lips – obtaining the vector of characteristic features:

$$D \to P, \tag{4}$$

where: $P = \{p_k : p_k^i \in F,\ i = \overline{1,s}\}$ – space of characteristic features, , F – the characteristic features of the researched object, p_k – characteristic vector, p_k^i – coordinates of this vector, s – dimension P.

Using the transformation (3), (4) an information technology [Krak] has been created which helps in modeling facial expressions and articulations of the human face.

Modeling of gestures animation and facial expressions

For the synthesis of animation of gestures and facial expressions on human spatial model the following formal description of the corresponding set of parameters and algorithms are proposed. Three-dimensional human model, which will be implemented by the process of animation gestures and facial expressions, has the following components: $V = \{v_i : v_i = (x,y,z)\}$ – set of vertices of triangles for the triangulation surface of three-dimensional human model; $N = \{n_i : n_i = (x,y,z)\}$ – set of normales to the vertices; $T = \{t_i : t_i = (t,r)\}$ – set of texture coordinates to the vertices; $V^{ind} = \{V_i^{ind} : V_i^{ind} = (k_1,k_2,k_3)\}$ – set of indices indicating the order of construction of the triangles from the set of vertices; $I = \{I_i : I_i = \{img\}\}$ – set of photographic images of elements of the model – textures; $i = \overline{1,S}$, S – a certain number.

To modulate a skeleton animation it is necessary to be able to calculate new values of the vertices V of triangles. For this purpose, a mechanism of skinning is used. It can be defined as an algorithm for binding set of vertices V to bones' rotation angles. Then the model of skeletal animation can be formalized as follows:

$$MH = \{MH_i : MH_i = \{k, \{l_1, \ldots, l_m\}, d_i, Glb_i, Order_i\}\}, \quad (5)$$

where MH – description of simplified skeleton of a person (hierarchy of bones) for realization of a skeleton animation. Here MH_i – i-th bone of a skeleton ($i = \overline{0, N-1}$, N – number of bones in a skeleton); k – index of parent bone; $\{l_1, \ldots, l_m\}$ - set of indexes of child bone, $d_i = [x_i, y_i, z_i]^T$ – coordinates of the end of i-th bone in the coordinate system associated with the beginning of this bone; Glb_i – vector to determine the coordinates of the bones in the global coordinate system; $Order_i$ – order of rotation; $Skin = \{Skin_i : Skin_i = \{(IndexVertex_1, Weight_1), \ldots\}\}$ – set of vertices that affect the current vertex when the angles are changed.

Skinning is calculated for each vertex v as follows:

$$v' = \sum_{i=0}^{n} \{(v * IM_{H_i} * TM_{H_i}) * w_{H_i}\}, \quad (6)$$

where n – number of bones belonging to the vertex v; IM_{H_i} – inversion matrix for bone H_i; TM_{H_i} – transfer matrix relatively to the bone H_i; w_{H_i} – weight coefficient of influence of point of bone H_i on vertex v.

Pronunciation of animation and emotional coloring are modeled using morphing technique (a smooth transition from one state of object to another). Only base states are used during morphing in which the intermediate states are calculated and animation are modeled. The method of segment morphing makes it possible to form several different facial expressions based on a small number of morphs and change the state of a face on a pronunciation animation. The advantage of this morphing is an opportunity to animate jaw independently from the lips and eyes, as well as from the emotional expressions on the face. The expression for the morphing from the R morphs in the formalism of the model for fixed frame is

$$V' = V + \sum_{m=1}^{R} w_m \cdot RM_m, \quad (7)$$

where w_m – weights coefficients, V – set of vertices of the base model, RM_m – input morph for blending. The result of the operation - a linear combination of the vertex set of a model and facial expressions.

Morphing using basic visemes is made taking into account weights coefficients in next steps:

1) The number of frames of gesture animation is calculated;
2) The set of emotion that are present in the word is defined;
3) The set of visemes to visualize the process of articulation is calculated;
4) Durations of animations stages, points of visemes emergence and weights for each base morph in frame are calculated.

To synchronize the animation of gesture and mimics it is necessary that the beginning and the end of animation of articulation coincide with the beginning and the end of the animation of a gesture.

Example of the synthesis of the human face conditions by blending morphs of visemes and emotions is shown on Pic. 1.

a) b) c)

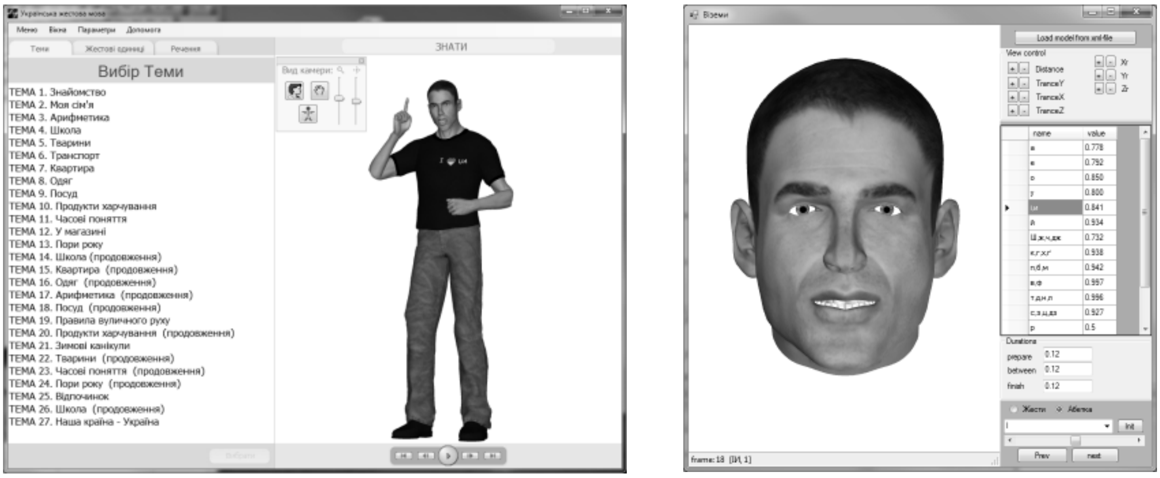
Pic. 1. Different states of human face: a) - viseme "y", b) - interrogative expression, c) the blending result during gesture animation

Information technology to play the animation process

To render the animation process of gestures and facial expressions of human three-dimensional model, software that implements the skeletal animation to peform a gesture (Pic. 2a) and morpheme animation to play facial emotions and pronunciation (Pic. 2b) is created.

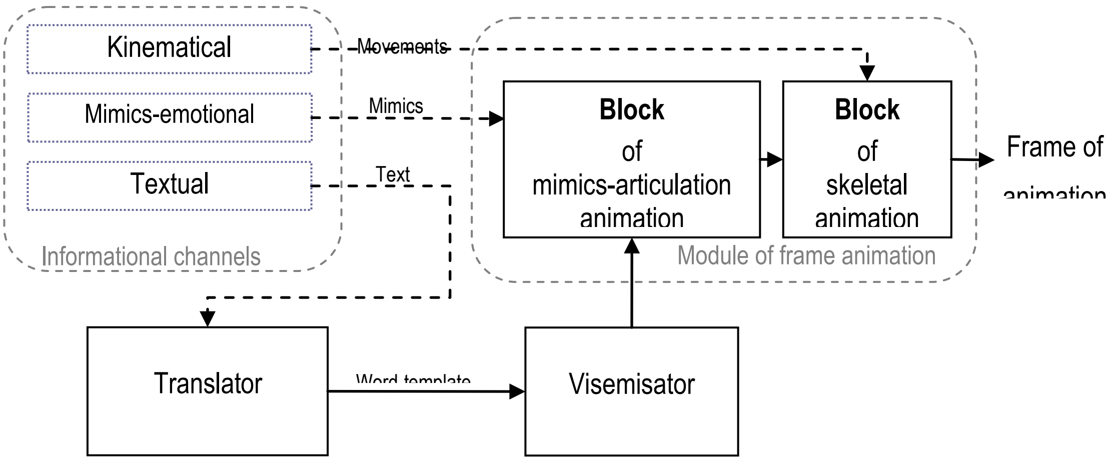
The corresponding software functionality is implemented, which, using three-dimensional API OpenGL, plays 3D human model by the specified attributes. The model of gesture animation using skinning algorithms (Block of skeletal animation) and morphing (Block of mimics-articulated animation) is shown on Pic.3.

Based on the established information technology Ukrainian Sign Language is implemented. Methodology of teaching sign language in specialized schools for deaf children is implemented. Formally, the technology consists of three informational blocks (topics, words and sentences) and a block to play gesture on virtual model, which has a specific purpose - with the help of it an opportunity to demonstrate in the learning process dynamics of the gesture is emerged. Since gestures are digitized by sign language native speaker, they, in fact, become the standards of playback of a gesture. This functionality for the frame demonstration of a gesture is a means by which it will be possible to study a gesture without a particular teacher.



a) b)

Pic. 2. a) Spatial model of a person, b) morphemic model of the human face



Pic. 3. Model of gesture animation process

Conclusions

Using the proposed model for the fixation of movements, couple dozens of gestures of Ukrainian Sign Language have been obtained. Thanks to developed technology of playing gestures from this set (using the spatial model) one can model the motion obtained from the video image of a specific person - sign language native speaker.

Developed technology and software allow to unify the study of sign language and in this sense can become the basis for creating a standard set of correct signing for a sign language.

Further studies are aimed at improving the proposed technology:

- filling the database of signs;
- development of tools for semantic binding of sentences on ordinary language with sentences on sign language (text-to-sign translation);
- improvement of the spatial human model taking into account some specifics of implementation of a sign language

Bibliography

[Belikov] Беликов В.И. Жестовые системы коммуникация // Семиотика и информатика. Вып. 20. – М.:Языки русской культуры, Русские словари,1973. – С.127-148.

[Krivonos] Кривонос Ю.Г., Крак Ю.В., Бармак О.В. [та ін.] Інформаційна технологія невербального спілкування людей з вадами слуху // Штучний інтелект. – 2008. №3. – С. 325-331.

[Kirichenko] Кириченко М.Ф., Крак Ю.В., Сорока Р.О. Оптимізація маніпуляційних роботів. – К.: Либідь, 1990. – 144 с.

[Lander] Lander J. Working with Motion Capture File Formats 3D. – LLC, Jan., 1998. – P. 30–37.

[Piegl] Piegl L., Tiller W. The NURBS Book, 2nd Edition. – Berlin, Germany: Springer-Verlag, 1996. – 645 p.

[Krak] Крак Ю.В., Бармак О.В., Тернов А.С. Информационная технология для автоматического чтения по губам украинской речи // Компъютерная математика. – 2009. – №1. – С. 86-95.

Authors' Information

Iurii Kryvonos *- V.M.Glushkov Cybernetics Institute of NASU, Director Deputy,*

Iurii Krak *V.M.Glushkov Cybernetics Institute of NASU, senior scientist*
krak@unicyb.kiev.ua,

Anton Ternov *– V.M.Glushkov Cybernetics Institute of NASU, junior researcher*
anton.ternov@gmail.com

address: 40 Glushkov ave., Kiev, Ukraine, 03680