Proceedings of ITC-CSCC 2000, Pusan, Korea

A Gesture-Emotion Keyframe Editor for Sign-Language Communication between Avatars of Korean and Japanese on the Internet

Sang-Woon KIM † Yung-Who LEE † Jong-Woo LEE † Yoshinao AOKI ††

† Div. of Computer Science & Engineering, Myongji University
San 38-2, Nam-dong, Yongin-shi, 449-728 Korea
Email : kimsw@wh.myongji.ac.kr
†† Graduate School of Engineering, Hokkaido University
N13 W8, Kita-ku, Sapporo-shi, 060-8626 Japan

Abstract— The sign-language can be used as an auxiliary communication means between avatars of different languages. At that time, an intelligent communication method can be also utilized to achieve real-time communication, where intelligently coded data (joint angles for arm gestures and action units for facial emotions) are transmitted instead of real pictures. In this paper, we design a gesture-emotion keyframe editor to provide the means to get easily the parameter values. To calculate both joint angles of the arms and the hands and to generate their keyframes realistically, a transformation matrix of inverse kinematics and some kinds of constraints are applied. Also, to edit emotional expressions efficiently, a comic-style facial model having only eyebrows, eyes, nose, and mouth is employed. Experimental results show a possibility that the editor could be used for intelligent sign-language image communications between different languages.

I. Introduction

The Internet cyberspace, a coordinate based virtual world where 3D sciences are constructed and animated, is a place of connecting millions of people around the world based on multimedia such as text, sound and animation. To enter cyberspace and become a member of a new community, an avatar which is a virtual persona for self-representation designed for navigating is needed[1]. The communication between(or among) avatars consists of two components: (1) Verbal components, which are written or spoken languages; (2) Non-verbal components, which are gestures, emotions, and intonations. For inter-personal communications, non-verbal components play an important role compared to verval ones. In [2], it is shown that 55% of the information content of a communication is being occupied by facial expressions (non-verbal information) as opposed to less than 7% of information occupied by verbal information. Also in [3], it is stated that communication between virtual humans should be equipped with the ability to recognize other virtual humans and perceive their facial expressions, gestures and postures. However, the current avatar communication is based on text only or both of text and voice[1].

Recently, therefore, a study on investigating a possibility of avatar communication using sign-language has been performed[4],[5]. From the study, it is revealed that the sign-language can also be used as an auxiliary communication means between avatars of different languages. To achieve real-time communication, first, a sign-language image is analyzed into a sequence of parameters and only intelligently coded data is transmitted instead of motion pictures or their compression. Then, the corresponding sign-language gesture is reconstructed with the received parameters using CG animation techniques. At that time, the joint angle parameters of the arms and the hands have been decided based on animator's experiences using forward kinematics. So, if the parameters have not been determined adequately, the animation generated with them become to be unnatural. And it takes long time for beginners to do that. Moreover, a facial expression is also selected arbitrarily and kept holding same emotion from start to finish of the animation.

To overcome those drawbacks, in this paper, we design a sign-language keyframe editor by which we can provide a means to get easily the parameter values using inverse kinematics. The proposed editor consists of three modules: an arm gesture editor, a hand shape editor and a facial emotion editor. With the arm gesture and hand shape editor, both joint angles of the arms and the hands are determined by using a transformation matrix of inverse kinematics[6]. In the facial emotion editor, a comic-style facial model[7] having only eyebrows, eyes, nose and mouth is employed to edit the desired emotional expression efficiently. Experimental results show a possibility that the editor could be used for intelligent sign-language image communication between avatars of different languages in cyberspace.

In Section II, we describe the intelligent sign-language communication. Section III addresses each components of the proposed editor and experimental results are discussed in Section IV. Finally, the conclusions are given in Section V.

II. Intelligent Sign-Language Communication

To achieve real-time sign-language communication, an intelligent communication method on a client-server architecture can be employed. In the method, 3D models are stored at all clients in advance and only intelligently coded data such as joint angles are transmitted instead of motion pictures or their compression. At server, sign-language images are analyzed into a sequence of parameters and transferred through the Internet. At clients, then, the corresponding CG animations are regenerated with the received parameters on their models.

As an application example, a sign-language communication system between Korean and Japanese can be considered. For inter-communication between KSL (Korean Sign-Language) and JSL (Japanese Sign-Language), we have designed a preliminary sign-language translation system, which consists of a KSL Dictionary, a JSL Dictionary and a Translation Dictionary. For example, a KSL sentence "NaNeun HakKyoEa KamNiDa (I go to school)" is translated into a JSL of "WatashiHa GakKouNi Ikimasu (I go to school)" and transmitted.

III. The Sign-Language Keyframe Editor

A. *Arm Gesture and Hand Shape Editors*

Sign-language animation is generated by adjusting joint angles of the arms, the hands, and the fingers. Even though an arm gesture is builted with seven kinds of joint angles, only four parameters among them play their roles in the arm's movement. On the other hand, each finger of the hand has four kinds of joint angles as its animation prarameters. Therefore, the twenty ($= 4 \times 5$) kinds of joint angles are used as the hand shape parameters. The total number of parameters (joint angles) for two arms and two hands amount to 54 ($= 27 \times 2$). In order to calculate their values, inverse kinematics[6] can be applied. Given the position and orientation of an end-effector (the wrist of an arm or the end of a finger), we can find its joint angles through the inverse kinematics so that the end-effector can be positioned as desired.

A 4×4 homogeneous transformation matrix equation of inverse kinematics and the geometrical analysis of the arm or the hand coordinate system are utilized to derive the solution equations for the joint angles. Using those equations, we design arm gesture editor and hand shape editor, by which the desired arm gestures and hand shapes can be edited. In those editors, both joint angles of the arms and the hands are calculated by the same transformation matrix, and the design window of arm gestures can be changed into that of hand shapes by a clicking. Then static and dynamic constraints to the motions of the arms and the fingers are applied in order to generate their keyframes realistically.

Figs. 1 and 2 show the arm gesture editor and the hand shape editor, respectively, which are implemented with Visual C++ 5.0 and Open Inventor on Windows platforms. In the figures, the left is an editing window of gestures or shapes and the right is joint angles' panel. When we build up the keyframes by dragging mouse on the left window, we can get easily the corresponding parameter values from the right panel.

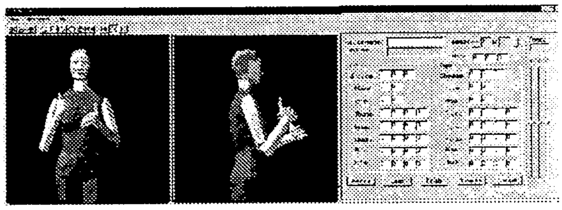

Fig. 1. A view of the arm gesture editor. The left two windows are gesture builders and the right is their 14 joint angles' panel.

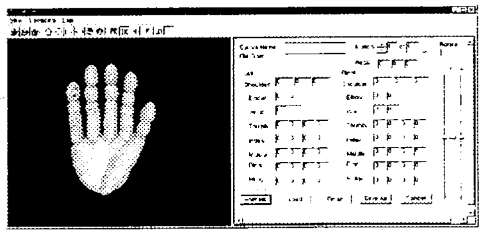

Fig. 2. A view of hand shape editor. The left window is shape builders and the right is their 20 joint angles' panel.

B. *Facial Emotion Editor*

Facial expression is an important factor in communicating emotional message. Also, it seems to be more essential to avatar communications using sign-languages.

Therefore, it is required that various kinds of facial expressions can be edited easily and transmitted together with the parameters of arm gestures. In this paper, we design a facial emotion editor by which six kinds of facial emotions such as Joy, Sadness, Disgust, Anger, Surprise, and Fear can be generated with flexible intensities. These emotions are expressed with the action units (AU's) of FACS (Facial Action Coding System) on an avatar model.

In general, a 3D avatar model is composed of many pieces of polygons while a 2D one is a line-drawing represented with small number of control points. The emotional expression cost (processing time) of the 3D model is more expensive than the 2D one, but the emotional expression capability is not improved so much. Therefore, we employ an abstract face model, a 2D comic-style emotional expression model, where almost components of a face are omitted except just the necessary parts such as eyebrows, eyes, nose, and mouth. Using this comic model, some kinds of exaggerated emotions can be edited efficiently on the 3D model.

Fig. 3 shows the facial emotion editor, which is implemented with Open Inventor on Windows platforms. In the figure, the left window is for a 3D facial model and the right is for a 2D comic model. The middle one is the buttons pannel for selecting an emotion among the six emotions and adjusting its intensity through an analog button. When we want to edit an emotion type, first, we generate the emotion on the 2D comic model using the type selection buttons and the intensity control button. Then the emotional expression of the 3D is determined according to the AUs of the 2D expression, which are shown up in the control pannel.

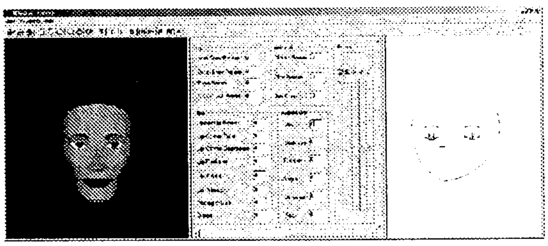

Fig. 3. A view of the facial emotion editor. The left window is for 3D facial model, the right is for the 2D comic model, and the middle is the AUs' control pannel.

IV. Experiments

We make experiment the sign-language keyframe editor implemented on Window platforms. Fig. 4 shows a set of keyframes for a KSL "NaNeun HakKyoEa KamNiDa (I go to school)", which are generated by the arm gesture editor. The parameter values (degrees) for each frames of Fig. 4 are shown in Table 1, where θ_1 is the joint angle around the shoulder Z axes; γ is the joint angle around the shoulder X axes; θ_2 is the joint angle around the shoulder Y axes; θ_3 is the joint angle around the elbow Y axes.

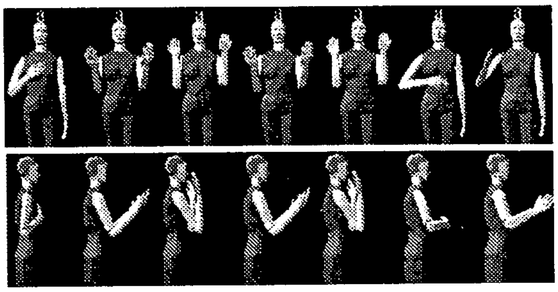

Fig. 4. Examples of KSL keyframes generated by the arm gesture editor, "Na(I)": the left-most 1 frame, "HakKyo(school)": the next 4 frames, and "GaDa(go)": the last 2 frames.

TABLE I

The values of joint angles (degrees) determined from the arm gesture editor panel for each frames of Fig. 4, where L-: Left-arm, R-: Right-arm.

No. of Frame	L-θ_1	L-γ	L-θ_2	L-θ_3	R-θ_1	R-γ	R-θ_2	R-θ_3
1	0	0	0	0	-20	60	-10	150
2	0	0	20	120	0	0	20	120
3	0	0	30	160	0	0	30	160
4	0	0	20	120	0	0	20	120
5	0	0	30	160	0	0	30	160
6	0	0	0	0	-30	60	10	120
7	0	0	0	0	10	10	-30	120

Fig. 5 shows a set of keyframes for numeric digits from "1" to "8", which are constructed by the hand shape editor. Table 2 shows the parameter values (degrees) of each frames in Fig. 5, where $\theta_i(\bullet), 1 \leq i \leq 4$ are the joint angles at the ith node of the thumb, the index, the middle, the ring and the small fingers, respectively.

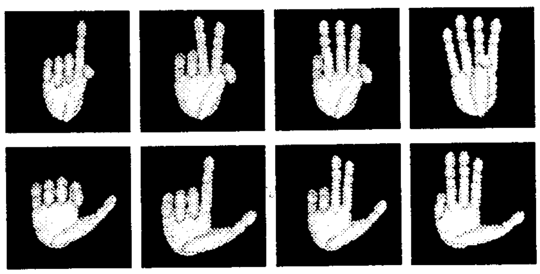

Fig. 5. Examples of hand shape keyframes for numeric digits from "1" to "8", which are generated by the hand shape editor.

TABLE II

THE JOINT ANGLES' VALUES (DEGREES) CALCULATED BY THE HAND SHAPE EDITOR FOR EACH FRAMES OF FIG. 5, WHERE T: THUMB, I: INDEX, M: MIDDLE, R: RING, S: SMALL FINGER.

Values of Angles	"1"	"2"	"3"	"4"	"5"	"6"	"7"	"8"
$\theta_1(T)$	-8	-8	-8	-32	43	43	43	43
$\theta_2(T)$	-6	-6	-6	18	-33	-33	-33	-33
$\theta_3(T)$	-6	106	106	15	14	14	14	14
$\theta_4(T)$	180	180	180	150	-18	-18	-18	-18
$\theta_1(I)$	0	0	0	0	0	0	0	0
$\theta_2(I)$	0	0	0	0	59	-22	-22	-22
$\theta_3(I)$	0	0	0	0	97	9	9	9
$\theta_4(I)$	0	0	0	0	64	6	6	6
$\theta_1(M)$	-1	-12	-4	-4	-1	-1	0	0
$\theta_2(M)$	53	-15	-33	-33	53	53	0	0
$\theta_3(M)$	102	13	21	21	102	102	0	0
$\theta_4(M)$	68	8	14	14	67	67	0	0
$\theta_1(R)$	8	8	-9	-9	8	8	8	0
$\theta_2(R)$	59	59	-9	-9	59	59	59	0
$\theta_3(R)$	95	95	9	9	95	95	95	0
$\theta_4(R)$	63	63	6	6	63	63	0	0
$\theta_1(S)$	15	15	15	-12	15	15	15	15
$\theta_2(S)$	44	44	44	-27	44	44	44	44
$\theta_3(S)$	70	70	70	87	70	70	70	70
$\theta_4(S)$	47	47	47	8	47	47	47	47

Fig. 6 shows a set of keyframes for an emotion "Sadness", which are generated with the facial emotion editor on the 2D and 3D comic models using action units of 1 and 15. Table 3 shows the number of AUs' and their intensities activated for generating each frames of Fig. 6. Finally, Table 4 shows a comparison between the gesture image transmission method and the intelligent communication method for the KSL keyframes shown in Fig. 4, "Na(I)": 1 frame, "HakKyo(school)": 4 frames, and "GaDa(go)": 2 frames.

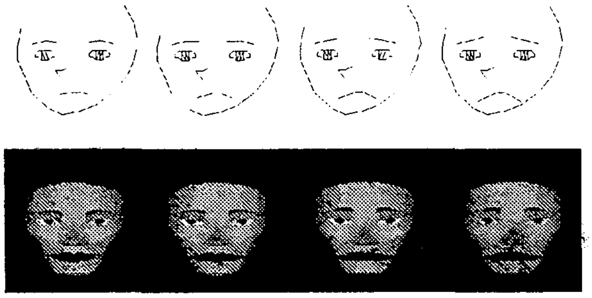

Fig. 6. Keyframes of an emotion "Sadness" generated with the facial emotion editor on 2D and 3D models using AU-1, 15. The intensities are increasing from left to right.

V. CONCLUSIONS

From the experiments, we confirmed a possibility that the editor could be used for building up the animation keyframes of sign-language communication between different languages. The future work is to apply this study into sign-language communication system implemented in virtual worlds on the Internet between Korean and Japanese, and Korean and English.

TABLE III

THE INTENSITIES OF ACTION UNITS (AUS) ACTIVATED FOR EACH FRAMES, FROM LEFT TO RIGHT, SHOWN IN FIG. 6.

No. of Frame	$AU-$ 1	$AU-$ 2, 4, 5, 6, 7, 10, 12	$AU-$ 15	$AU-$ 17, 20, 24, 25, 26, 48
1	25	0	25	0
2	50	0	50	0
3	75	0	75	0
4	100	0	100	0

TABLE IV

COMPARISON OF TRANSMITTED DATA (BYTES) AND TRANSFER RATE (KBPS) BETWEEN THE IMAGE COMMUNICATION AND THE PROPOSED INTELLIGENT COMMUNICATION FOR EACH WORDS OF FIG. 4.

KSL Words	The Gesture Image Communication	The Proposed Intelligent Communication
"Na(I)"	$1,840 \times 1 \times 2 = 3,680$	$54 \times 1 + 4 \times 2 + 11 = 73$
"HakKyo(school)"	$1,840 \times 4 \times 2 = 14,720$	$54 \times 4 + 4 \times 3 + 11 = 239$
"GaDa(go)"	$1,840 \times 2 \times 2 = 7,360$	$54 \times 2 + 4 \times 3 + 11 = 131$
Transfer rate(Kbps)	$\frac{25,760 \times 8}{0.7 \times (1+4+2)} = 42.06$	$\frac{443 \times 2 \times 8}{0.7 \times (1+4+2)} = 1.45$

ACKNOWLEDGMENTS

This work was supported by Korea Research Foundation (grants number : 1998-001-E00979). The work was also supported in part by TAO (Telecommunications Advancement Organization of Japan) as an International Collaborative Research. We wish to express our appreciations to those grants.

REFERENCES

[1] B. Damer, Avatars! Exploring and Building Virtual Worlds on the Internet, Peachpit Press, 1998.

[2] A. Mehrabian, "Communication without words," in J. A. Devito, eds., Communication: Concepts and Process, Prentice-Hall, pp. 106 - 114, 1971.

[3] N.M. Thalmann and D. Thalmann, "Creating artificial life in virtual reality," in Artificial Life and Virtual Reality, N.M. Thalmann and D. Thalmann, eds., John Wiley and Sons, pp. 1 - 10, 1994.

[4] J. Li, Y. Aoki, and S.-W. Kim, "Consideration of a CG-animation system for sign language communication between Japan and Korea," IEICE Trans., vol. J82-D-I, no. 5, pp. 613 - 621, May 1999.

[5] S. -W. Kim, J. -Y. Oh, S. Tanahashi, and Y. Aoki, "Preliminary study on a sign-language chatting system between Korea and Japan for avatar communication on the Internet," IEICE Trans., vol. E83-A, no. 2, pp. 613 - 621, Feb. 2000.

[6] K. S. Fu, R. C. Gonzalez, and C. S. G. Lee, Robotics: Control, Sensing, Vision, and Intelligence, McGraw-Hill, pp. 12 - 81, 1987.

[7] S. Tanahashi, Y. Aoki, and S.-W. Kim, "A comic emotional expression method and its applications," Proc. TENCON'99, pp. 329 - 332, Sept. 1999.