Chinese Sign Language Synthesis Technology and Its Applications Postprint

Authors: Chen Yiqiang, Liu Junfa, Yan Qingcong, He Wenjing

Date: 2016-11-02T00:00:00+00:00

Abstract

The "Chinese Sign Language Synthesis Technology" developed by our team can convert arbitrary given text into sign language and display it through multimodal behavior synthesis of virtual animated characters. Based on the analysis of current domestic and international research on sign language synthesis technologies and systems, this paper focuses on introducing our work, including key technologies such as virtual human modeling, natural language to sign language conversion, gesture smoothing, gesture motion retargeting, and multimodal behavior expression of animated characters. It also introduces the development of information accessibility application systems based on Chinese Sign Language synthesis technology, such as television sign language production systems for broadcasting programs and web-based teaching systems for commonly used Olympic sign language for online learning. Finally, it presents existing problems and our next-phase research work combining embedded terminals and ubiquitous computing.

Full Text

Preamble

Vol. 7 No. 3
Information Technology Letters
Chinese Sign Language Synthesis Technology and Its Applications
Yiqiang Chen, Junfa Liu, Qingcong Yan, Wenjing He

Abstract

We have developed a "Chinese Sign Language Synthesis Technology" that can convert arbitrary text into sign language and display it through multimodal behavior synthesis of virtual animated characters. This paper analyzes current domestic and international research on sign language synthesis technologies and systems, focusing on our contributions, including key technologies such as virtual human modeling, natural language to sign language conversion, gesture smoothing, gesture motion retargeting, and multimodal behavior expression for animated characters. We also introduce the development of information accessibility application systems based on Chinese sign language synthesis technology, such as a TV sign language production system for broadcasting programs and an online teaching system for Olympic common terms sign language. Finally, we discuss existing problems and our next steps in research integrating embedded terminals and ubiquitous computing.

Keywords: Sign language synthesis, multimodal, ubiquitous computing, information accessibility

1. Introduction

Deaf and hard-of-hearing individuals cannot conveniently transmit and obtain information through speech like hearing people, relying primarily on sign language. Although deaf individuals can now interact deeply with the outside world through television, the internet, mobile phones, and other tools, barriers to information access remain widespread. Using modern science and technology to freely obtain information like hearing people is the dream of every deaf person. Achieving this goal involves two technologies: first, sign language recognition, where deaf individuals sign, computers capture sign data through sensors, recognize the meaning, and translate it into speech for hearing people who don' t understand sign language, enabling them to "hear" sign language. Second, sign language synthesis, where hearing people express their intentions through speech, and computers translate the speech into sign language for display, enabling deaf people to "see" sound. This paper focuses on the latter technology —sign language synthesis.

Currently, sign language synthesis systems for multiple languages have emerged worldwide. The earliest computer-based sign language synthesis research dates to 1982 [1], when M. Shantz and H. Poizner described a program for synthesizing American Sign Language using a stick figure model programmed in BASIC. Subsequently, China, Japan, Germany, and many other countries conducted research on their national sign languages, achieving important results. Japanese sign language research began in 1985 [2], and by 1995, Hitachi had implemented a bidirectional conversion system from Japanese to Japanese Sign Language [3], capable of converting Japanese sign gestures into Japanese speech and vice versa. The system was applied to a residence card automatic dispenser to explain its operation to deaf individuals. In 2006, L. V. Zijl and A. Combrink implemented translation for South African Sign Language [24], using an extended synchronous tree-adjoining grammar parser to generate non-manual gestures and establish appropriate gesture space. To support Greek sign language teaching in schools, K. Karpouzis et al. designed an H-Anim-based virtual character in 2007 to visualize Greek Sign Language [25], where teachers input Greek text, which is converted to sign language performed by the virtual character. Italian researchers F.

Buttussi et al. also worked on a bilingual dictionary between sign language and written language [26], developing 3DictSL in 2007, which provides bidirectional query functionality, supports an online international sign language dictionary, allows users to learn sign languages from other countries, and enables deaf communities to expand the dictionary in their own language. In 2008, R. San-Segundo et al. from Spain first attempted to implement Spanish sign language translation in practical domains [27]. Also that year, Tirthankar et al. presented a prototype machine translation system from text to Indian Sign Language [28], which takes simple English sentences as input, performs syntactic analysis, generates corresponding Indian Sign Language syntactic structures, and outputs translation results as video streams. In 2009, M. Delorme et al. from France proposed a novel sign language synthesis process [29]. Unlike joint animation using parameter sets to describe gestures, they used geometric constraints such as vectors, planes, points, and symmetry axes for gesture modeling. Their system consists of a speech recognizer, a natural language translator, and a 3D virtual character animation module. A research group led by the University of Hamburg in Germany developed a sign language system aimed at promoting German sign language information services, using data gloves to capture motion data and employing vertex blending and joint animation techniques to drive 3D virtual human gestures, achieving good real-time performance [30].

Research on Chinese sign language synthesis systems in China began in the 1990s [4-16]. Early participants included the Gao Wen research group at the Institute of Computing Technology, Chinese Academy of Sciences, Harbin Institute of Technology, and Zhejiang University. The basic approach involves using motion capture and data gloves to collect motion data, then implementing Chinese sign language synthesis systems through coordinate transformation calculations and virtual human motion simulation techniques. Currently, the Chinese Sign Language Synthesis System developed by the Institute of Computing Technology, Chinese Academy of Sciences, has been applied in deaf school teaching, and has also been developed into a TV sign language production system for broadcasting programs and an online teaching system for Olympic common terms sign language.

2. Chinese Sign Language Synthesis Technology

Chinese sign language synthesis systems are generally based on the 30 finger alphabets and 5,586 basic sign vocabulary items defined in *Chinese Sign Language (Revised Edition, Volumes 1 & 2)* [15]. The process first uses motion tracking to establish a basic sign motion database [16], then corrects specific joint point data on this foundation to obtain an accurate basic gesture motion database. For any given Chinese text sentence, the system first uses a word segmentation system to segment the text so that these words can be recognized by the sign language synthesis system. It then extracts the corresponding gesture motion data from the database based on the recognized sign vocabulary, and finally concatenates the motion data corresponding to these sign vocabulary items into a motion sequence to drive the virtual human to play the animation. The entire sign language synthesis process is shown in Figure 1 [Figure 1: see original paper].

Specifically, sign language synthesis technology mainly includes the following research components: virtual human modeling technology, natural language to gesture motion conversion technology, gesture motion smoothing technology, gesture motion data retargeting technology, and multimodal sign language synthesis technology.

1.1 Virtual Human Modeling

Virtual human modeling generally begins by drawing the 3D appearance of the virtual human using graphics software according to actual requirements, then converting this appearance into Virtual Reality Modeling Language (VRML) data. VRML is considered the second-generation WWW standard language, primarily used to represent and share 3D virtual scenes and objects on the internet. VRML has a specialized sub-standard for describing virtual human models called H-Anim, which uses three types of nodes to represent a virtual human body model: body center of mass, human joints, and body segments, dividing the entire body into one center of mass, 77 joints, and 47 body segments. The positions of these body segments represent the virtual human' s posture.

The virtual human model includes the virtual human skeleton and the virtual human surface model. Typically, we only see the surface model of the virtual human (including muscles, skin, clothing) and not the skeleton model (as shown in Figure 2 [Figure 2: see original paper]).

The virtual human surface model consists of 53 parts, as shown in Figure 3 [Figure 3: see original paper]. A small common area is set between each pair of connected parts to prevent "cracks" during virtual human motion. As shown in Figure 4 [Figure 4: see original paper], the left image shows the connected whole, the middle shows its segmentation pattern, and the right shows the actual cross-modeling method. All segmented parts are modeled according to this pattern.

1.2 Natural Language to Gesture Motion Conversion

The general approach for natural language to gesture motion conversion uses natural language understanding technology to decompose text into basic sign words, then finds their corresponding gestures in the gesture library, and outputs the gesture motion sequence animation according to the required sign language speed. In sign language systems, basic sign words are words that can be directly displayed by the virtual human using motion capture data. The keyframe gestures representing these words are relatively smooth and gentle. If the number of basic sign words in the system is too small, many words will need to be expressed using finger-spelling gestures, making the gestures difficult for deaf people to understand. However, if there are too many basic sign words, the workload of building the sign vocabulary library increases dramatically, and the computational complexity also increases significantly during sign language synthesis due to the large search space. In fact, sign language experts point out that the number of commonly used gesture words in sign language is limited, yet sufficient to cover daily common gestures. Therefore, we selected all 5,586 vocabulary items and 30 finger alphabets from the book *Chinese Sign Language* [15] as basic sign words. For other vocabulary, we established a correspondence mechanism between synonyms and basic sign words and a synonym expansion mechanism. This enables the sign language system to express sign vocabulary that basically meets the requirements of large text corpora such as TV news programs. On the other hand, like natural language, sign language contains many words with the same form but different meanings (primarily manifested in different parts of speech for the same word). For example, the character "行" in "银行" (bank) and "行走" (walk) should be signed differently due to their different parts of speech, as shown in Figure 5 [Figure 5: see original paper]. There are even cases where the vocabulary and part of speech are the same, but the meaning differs in different linguistic contexts.

These natural language characteristics obviously pose significant challenges to sign language systems. To address these challenges, we need to vigorously leverage the latest research results from the natural language processing field. Currently, we have incorporated part-of-speech attributes into the sign vocabulary expression process, enabling different gestures to be displayed for different parts of speech.

1.3 Gesture Motion Smoothing Algorithm

Basic sign words generally lack motion capture data between them. To synthesize fluent sentence-level animation, gesture animation must be smoothed. The smoothness of virtual human gesture animation directly affects the intelligibility of gesture movements. Since virtual human gesture animation is an animation sequence concatenated from some meta-animation data, there are large differences in gesture movements between adjacent sign words and between different roots within the same sign word. Without smoothing, the gesture animation will exhibit excessive changes at the seams between adjacent roots.

Through statistical analysis of sign language sentence samples, we found that the main factors affecting gesture motion smoothness are the rotation angle change amplitudes of the shoulder and elbow joints. The shoulder joint has 3 rotational degrees of freedom, rotating around the X, Y, and Z axes respectively, as shown in Figure 6 [Figure 6: see original paper]. The elbow joint has 2 rotational degrees of freedom, rotating around the X and Y axes respectively, as shown in Figure 7 [Figure 7: see original paper].

The shoulder joint' s motion trajectory is spherical around the shoulder joint. Therefore, spherical linear interpolation (SLERP) is used for the three rotation angles of the shoulder joint:

Assume p0 and p1 are quaternions representing the three rotation angles of the shoulder joint in two adjacent keyframes of a sign motion, and n is the number of frames to be interpolated between p0 and p1. Then when interpolating the i-th ($0 \leq i \leq n$) frame: Slerp(p0, p1, t), where Ω is the rotation angle magnitude.

For the elbow joint, during motion, its rotation around the X-axis always remains in the same plane as the upper arm, while its rotation around the Y-axis only determines the palm' s orientation (as shown in Figure 7). Therefore, interpolation must be performed separately for the two rotational degrees of freedom of the elbow joint. For its rotation angle around the X-axis, due to its constraint relationship with the three degrees of freedom of the shoulder joint, we must use statistical learning methods to find the constraint relationship between the elbow joint' s rotation angle around the X-axis and the length of the spherical minor arc traversed during the elbow joint' s spherical motion (i.e., expressed through the change amounts of the three rotation angles of the shoulder joint). We selected approximately 600 relatively smooth root samples from the overall sample to form a training sample space, and used regression analysis on this training sample to identify the constraint relationship described above. This constraint relationship is then used for elbow joint interpolation. For the rotation angle around the Y-axis, simple Euler interpolation is directly used.

On the other hand, the number of keyframes to be interpolated at root seams is also a critical parameter. We calculate this value by defining a mathematical formula to measure the distance between adjacent keyframes, thereby achieving uniform and smooth transitions throughout the gesture motion. Since the rotation amplitude of the shoulder and elbow joints is directly reflected in the change amplitude of the spatial coordinate positions of the elbow and wrist joints, we use the weighted sum of Euclidean distances between elbow and wrist joints in the two keyframes as the distance measure.

The distance between frames i and j is defined as a weighted sum of Euclidean distances between elbow and wrist joint positions for both hands. Based on experimental results, we set $\alpha = 0.67$ and $\beta = 0.33$.

In experiments, using this distance formula on the training sample space described above, we statistically learned a threshold: $\Theta = 7.15$. If the distance between two frames is less than this value, the gesture motion change amplitude is considered sufficiently small and no frame interpolation is needed. Conversely, a certain number of keyframes are interpolated to achieve uniform and smooth connection throughout the gesture motion. Figure 8 Figure 8: see original paper shows the smoothing result for the test text "大家好" between the last frame of the sign word "大家" and the first frame of "好" using the above smoothing algorithm, which is better than the result obtained by direct simple quaternion interpolation (b). Due to the length of the original sequence, the results shown are sampled at equal intervals.

1.4 Gesture Motion Data Retargeting

As mentioned above, basic sign motion data is generally obtained by recording an actor' s real motion using motion capture devices, then reapplying the obtained standard motion data to various types of virtual characters. Since the size and proportions of virtual characters are basically inconsistent with those of the performer, directly applying standard motion data will result in errors. Therefore, a retargeting technology is needed to automatically migrate standard model gesture motion data to different virtual human models while preserving gesture meaning to the greatest extent [11].

1.4.1 Gesture Feature Analysis For gesture meaning, the most important characteristic is the precise positional relationship between the hands and other body parts. Minute changes in this position often lead to completely different meanings, so these important features must first be obtained. Figure 9 [Figure 9: see original paper] shows three keyframes corresponding to the sign words "人" (person), "眼" (eye), and "方" (square). It can be seen that for "人" and "方", the positional relationship between the two index fingers is the key feature —deviation affects meaning expression—while other upper limb joints such as elbows or wrists are less sensitive. For "眼", the positional relationship between the index finger and the eye is the most important feature; if the index finger points to the nose, it becomes another sign word "鼻" (nose).

To analyze feature information from gestures, we define three sets of sensitive points on the two hands and head (see Figure 10 [Figure 10: see original paper]). Major joint points are selected as sensitive point sets on the hands, and several key points such as eyes, ears, nose, and mouth are selected as sensitive points on the head. From the previous analysis, we know that the closer the distance between two points, the higher their importance and the more necessary it is to preserve their relative positional information after retargeting, and vice versa. The pair of points with the shortest distance is called key sensitive points, and the next closest pair is called secondary key sensitive points. Experiments show that selecting key sensitive points and secondary key sensitive points can capture most of the main feature information. The method is briefly introduced as follows:

Considering only three parts—the two hands and the head (the torso and arms can also be defined)—we define three sets: H represents head sensitive points; L and R represent left and right hand sensitive points respectively. We define 22 points on each hand (n) and 11 points on the head (m). First, we analyze the relative motion between the two hands; the analysis of relative motion between hands and head is similar.

For any point ir R and jl L, let D(ir, jl) be the Euclidean distance between ir and jl. Sort D and take the two smallest values d1 and d2 (d1 $\leq$ d2). Select the ir and jl corresponding to d1 as key sensitive points, and the ir and jl corresponding to d2 as secondary key sensitive points. As shown in Figure 11

[Figure 11: see original paper]: the left figure selects k1 and k2 as key sensitive points, s1 and s2 as secondary key sensitive points; the right figure has only key sensitive points k1 and k2, with no secondary key sensitive points.

1.4.2 Mapping Sensitive Points From the previous analysis, we know that the relative positional relationship between hands and other body parts in sign language is key to understanding meaning. If these relative positional relationships can be mapped during retargeting, the original meaning can be preserved. Having identified the key sensitive points and secondary key sensitive points that need to be mapped, we now discuss how to map these points.

Generally, the relative positional relationship to be mapped can be represented by spatial vectors. The vector connecting two key sensitive points can express this relative positional relationship—simply ensuring that the vector formed by corresponding sensitive points remains consistent before and after mapping.

When calculating the relative position of two sensitive points, two cases must be considered: first, one point is fixed while the other moves, such as a finger pointing to the head; second, both points are moving, such as hand-to-hand motion.

As shown in Figure 12 [Figure 12: see original paper], s1 and s2 are the original sensitive points, and t1, t2 are the corresponding sensitive points obtained by directly using the original data. To obtain new appropriate sensitive point positions: for the first case, set s2 as the fixed point, translate s1 and s2 until s2 coincides with t2, and the resulting new t1' and t2' (t2) are the retargeted sensitive point positions; for the second case, translate s1 and s2 until their center coincides with the center of t1 and t2, and the resulting new t1' and t2' are the retargeted sensitive point positions.

The corresponding calculation formulas are: Equation (3) for the first case and Equation (4) for the second case: = (3) = (4)

1.5 Multimodal Sign Language Synthesis

Facial and lip movements play an important auxiliary role in deaf communication. The intelligibility of sign language synthesis lacking facial and lip movements is greatly affected. Therefore, how to incorporate other modal behaviors such as speech facial expressions into sign language movements and achieve coordinated expression is a key challenge in realizing realistic and natural sign language synthesis. We adopt the principles and ideas of data mining, and based on a unified behavior prosody model, mine the coordination relationships among multimodal behaviors from large multimodal databases. For input text, the mined model can predict speech prosody and sign language prosody, and by organically combining the two, construct a multimodal behavior coordination control model for the coordinated control of multimodal behaviors [10].

Figure 13 [Figure 13: see original paper] shows the framework of a multimodal behavior coordinated synthesis system combining text-driven animated characters. The system inputs natural language text, uses natural language processing technology to analyze linguistic features such as grammar and morphology, and then uses speech prosody learning models from text-to-speech conversion to obtain speech prosody control parameters. Simultaneously, it extracts behavior prosody control parameters related to behavior from the recorded behavior database, thereby constructing a multimodal coordinated prosody control model. The main method uses the combination of speech prosody information and partial behavior prosody information to obtain coordinated prosody control parameters, which are then applied back to the generation of prosody control parameters for various behavior modes to complete the synthesis of various behaviors. The synthesized behavior sequence already contains coordinated control information and can be played directly.

3. Chinese Sign Language Application Systems

Some research results of Chinese sign language synthesis technology have been gradually put into practical use, providing information accessibility services for deaf people in China. Among them, the "3D Virtual Human TV Digital Sign Language Editing System" automatically synthesizes sign language video windows that can be played synchronously with TV programs. The "Chinese Sign Language Olympic Common Terms Online Teaching System" served as essential learning software for 2008 Paralympic volunteers and successfully supported the 2008 Beijing Paralympic Games.

3.1 3D Virtual Human TV Digital Sign Language Editing System

The purpose of this system is to convert text into Chinese sign language expressed by a virtual human and overlay the animated image sequence onto video programs synchronously. The system fully considers the characteristics of various TV signals (standards, signal types, resolution, etc.), and uses hardware and DirectShow technology to automatically convert electronic Chinese manuscripts into sign language animation, which is output as video signals or files and overlaid onto news videos for user viewing.

During operation, the system first imports the electronic manuscripts of each news item and the host' s spoken scripts into the system' s news list in broadcast order. Since sign language is slower than speech, the second step is to edit the manuscripts into sign language manuscripts according to the actual duration of each news segment. The system provides a function to display the sign language duration for each news segment to ensure the edited sign language manuscripts match the news duration. Finally, sign language animation is generated and overlaid with the news 画面. The system provides complete sign language playback control functions. In a typical application scenario, by using a method similar to manual captioning with reference to the program content, sign language animation can be overlaid onto the news 画面.

The figure below shows the application flowchart of the system at a TV station and the video output with sign language animation overlaid:

Figure 14 [Figure 14: see original paper]. Sign language news application flowchart
Figure 15 [Figure 15: see original paper]. Sign language news video output

3.2 Chinese Sign Language Olympic Common Terms Online Teaching System

The purpose of the Chinese Sign Language Olympic Common Terms web teaching system is to enable anyone to interactively learn Chinese sign language expressions through the web [23] and animated characters at any time. When using the system, after logging into the webpage, users input the vocabulary they want to learn, trigger the system to play the sign gesture, and simultaneously integrate mature speech synthesis technology to achieve real-time synchronized speech and gesture playback. Considering the needs of beginners, the system provides different gesture playback speeds for learners at different levels. The system also offers the function of viewing sign gestures from different angles, enabling users to understand sign gesture movements more accurately. The system integrates real-time synchronized speech playback, making learning easier for users.

This system is a WEB service based on Browser/Server (B/S) architecture, consisting of client and server parts. Figure 16 [Figure 16: see original paper] shows the system block diagram. Figure 17 [Figure 17: see original paper] shows a screenshot of the system client interface.

4. Conclusion

Research on Chinese sign language synthesis has received support from multiple projects including the Chinese Academy of Sciences Knowledge Innovation Project, National Natural Science Foundation Key Project, National 863 Program, and National Science and Technology Support Program. While conducting in-depth technical research, we have successively developed practical systems to help disabled people obtain timely and accurate information, understand current media reports, and share the enjoyment of sports competitions and cultural performances. However, Chinese sign language synthesis technology still has deficiencies in natural language to gesture conversion and realistic multimodal expression of virtual humans, requiring collaboration with researchers from other fields such as machine translation to further improve. For future research, we are continuing to integrate sign language synthesis technology into increasingly powerful embedded terminals, and gradually developing multilingual sign language synthesis, enabling one software to simultaneously express multiple national sign languages including Chinese, Japanese, Korean, American, and German through virtual humans. We will also apply the latest achievements in ubiquitous computing to provide novel, more humanized, and personalized services for disabled users, such as using wireless positioning technology and behavior recognition technology to perceive the location and status of deaf people in real time, and provide timely virtual human sign language prompt services to them or their guardians.

References

[1] Shantz, M., Poizner, H. A Computer Program to Synthesize American Sign Language. *Behavior Research Methods and Instrumentation*. 1982, 14(5): 467-474

[2] Kawei Hideo, Tamura Shinichi. Deaf-and-mute Sign Language Generation System. *Pattern Recognition* 1985, 18(3-4): 199-205

[3] Prime, M. Integrating Sign Language into Virtual Reality Environments. IEEE Colloquium on Visualization of Three-Dimensional Fields. London, UK, 1996: 1-4

[4] Shu Hongxian, Wang Shenkang, Zhang Gaoqing. Research and Implementation of a Hypermedia Sign Language Computer-Aided Teaching System. Proceedings of the 2nd China Computer Intelligent Interface and Intelligent Application Academic Conference. Weihai, 1995: 407-412

[5] Gao Wen. Framework Structure of a Multifunctional Perception Machine. Proceedings of the 2nd China Computer Intelligent Interface and Intelligent Application Academic Conference. Weihai, 1995: 17-20

[6] Multifunctional Perception Machine Stage Progress Technical Report. Department of Computer Science and Engineering, Harbin Institute of Technology, 1995

[7] Multifunctional Perception Machine Technical Report. Harbin Institute of Technology, 1996

[8] Wang Chengwei, Gao Wen, Wang Xingren. Theory, Implementation and Application of Virtual Reality (Lingjing) Technology. Tsinghua University Press. 1996

[9] Song Yibo, Gao Wen, Yin Baocai, et al. A Text-Driven Deaf Sign Language Synthesis System. *Chinese Journal of Computers*, 1999, 22(7): 733-739

[10] Chen Yiqiang, Gao Wen, Liu Junfa, Yang Changshui. A Multimodal Behavior Coordinated Prosody Model in Sign Language Synthesis. *Chinese Journal of Computers*, 2006, 29(5): 822-827

[11] Ge Chunbao, Chen Yiqiang, Yin Baocai, Gao Wen, Yang Changshui. A New Gesture Motion Data Retargeting Method. *Chinese Journal of Computers*, 2006, 29(10): 1850-1855

[12] Chen Yiqiang. Research on Virtual Human Multimodal Behavior Synthesis Based on Data Mining. Doctoral Dissertation, Institute of Computing Technology, Chinese Academy of Sciences. 2003, Beijing.

[13] ISO/IEC 14772-1:200x part26-2001 Humanoid Animation Specification[S]

[14] ISO/IEC JTC 1/SC 29/WG 11 N2802-1999 Collaborative Virtual Environments Contributions to MPEG-4 SNHC[S]

[15] China Association of the Deaf. *Chinese Sign Language (Revised Edition, Volumes 1 & 2)* ISBN: 7508030052. 2003

[16] Wang Zhaoqi, Yang Changshui, Gao Wen. 3D Display of Chinese Sign Language Based on VRML. *Journal of Computer Research and Development*, 2003, 40(3):
[17] B. Paden. Kinematics and Control of Robot Manipulators. PhD thesis, University of California, Berkeley, 1986
[18] J. U. Korein and N. I. Badler. Techniques for generating the goal-directed motion of articulated structures. *IEEE CG&A*, Nov. 1982: 71–81
[19] D. Tolani, A. Goswami, and N. Balder. Real-time inverse kinematics techniques for anthropomorphic limbs. *Graphical Models*, September 2000, vol. 62, no. 5: 353-388(36)
[20] M. Girard and A. A. Maciejewski, "Computational modeling for the computer animation of legged figures,"*Computer Graphics*, July 1985, 19(3): 263-270
[21] J. Zhao and N. I. Badler. Inverse kinematics positioning using nonlinear programming for highly articulated figures. *ACM Transactions on Graphics*, 1994, 13(4): 3–13–336
[22] W. C. Stokoe, *Sign Language Structure: An Outline of the Visual Communication System of the American Deaf.* Studies in Linguistics: Occasional Papers 8 (Revised 1978). Buffalo, NY: Linstok,
[23] http://mct.ict.ac.cn/webser/test.html
[24] Lynette Van Zijl, Andries Combrink. The South African Sign Language Machine Translation Project: Issues on Non-manual Sign Generation. October 2006 SAICSIT ' 06: Proceedings of the 2006 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries.
[25] K. Karpouzis, G. Caridakis, S.-E. Fotinea, E. Efthimiou. Educational Resources and Implementation of a Greek Sign Language Synthesis Architecture. *Computers & Education*, 2007, Vol. 49(1): 54-74.
[26] Fabio Buttussi, Luca Chittaro, Marco Coppo. Using Web3D Technologies for Visualization and Search of Signs in an International Sign Language Dictionary. April 2007 Web3D ' 07: Proceedings of the Twelfth International Conference on 3D Web Technology.
[27] R. San-Segundo, R. Barra, R. Córdoba, L. F. D' Haro, F. Fernández, J. Ferreiros, J. M. Lucas, J. Macías-Guarasa, J. M. Montero, J. M. Pardo. Speech to Sign Language Translation System for Spanish. November 2008 *Speech Communication*, Volume 50 Issue 11-12
[28] Tirthankar Dasgupta, Anupam Basu. Prototype Machine Translation System from Text-to-Indian Sign Language. January 2008 IUI ' 08: Proceedings of the 13th International Conference on Intelligent User Interfaces.
[29] Delorme, M.; Filhol, M.; Braffort, A. Animation Generation Process for Sign Language Synthesis. *Advances in Computer-Human Interaction*, 2009. ACHI ' 09, Second International Conferences on 1-7 Feb. 2009 Page(s): 386–390. Digital Object Identifier 10.1109/ACHI.2009.29
[30] http://www.sign-lang.uni-hamburg.de/

Author Biographies

Yiqiang Chen: Deputy Director and Associate Researcher, Pervasive Computing Research Center, Institute of Computing Technology, Chinese Academy of Sciences, Perceptual Computing Group.

Junfa Liu: Assistant Researcher, Pervasive Computing Research Center, Institute of Computing Technology, Chinese Academy of Sciences, Perceptual Computing Group.

Qingcong Yan: Engineer, Pervasive Computing Research Center, Institute of Computing Technology, Chinese Academy of Sciences, Perceptual Computing Group.

Wenjing He: Master' s Student, Pervasive Computing Research Center, Institute of Computing Technology, Chinese Academy of Sciences, Perceptual Computing Group.

Note: Figure translations are in progress. See original paper for figures.