Study on Translating Chinese into Chinese Sign Language

XU Lin (徐 琳)[1] and GAO Wen (高 文)[1,2]

[1] *Department of Computer Science and Engineering, Harbin Institute of Technology Harbin 150001, P.R. China*

[2] *Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, P.R. China*

E-mail: xl@vilab.hit.edu.cn; wgao@ict.ac.cn

Received May 4, 1999; revised October 6, 1999.

Abstract Sign language is a visual-gestural language mainly used by hearing-impaired people to communicate with each other. Gesture and facial expression are important grammar parts of sign language. In this paper, a text-based transformation method of Chinese-Chinese sign language machine translation is proposed. Gesture and facial expression models are created. And a practical system is implemented. The input of the system is Chinese text. The output of the system is "graphics person" who can gesticulate Chinese sign language accompanied by facial expression that corresponds to the Chinese text entered so as to realize automatic translation from Chinese text to Chinese sign language.

Keywords machine translation, Chinese, Chinese sign language, rule

1 Introduction

Research of automatic translation between natural language and visual language is of practical and academic significance. Natural language Chinese is one of the languages that have the most users in the world. Chinese sign language (CSL) is a visual-gestural language, which transmits message by means of hand movement, facial expression change and body movement.

Human language is composed of natural language and body language. Natural language includes spoken language and written language. Body language includes gesture, facial expression, body gesture and head gesture. While people communicate face to face, natural language and body language transmit message simultaneously or cooperatively. It's of significance to research the model of sensing and expressing of body language, and the model of information merging with natural language, which can promote the level of computer understanding of natural language and enhance practicability of human-computer interface[1].

Sign language is a very important component of body language. It's a kind of body language that contains the most information. It has the same expression power as natural language in taking the task of transmitting information. And sign language has more powerful effect in vision. It's lively, vivid, and directly perceived.

CSL is a relatively steady expression system which is made up of hand shapes accompanied by facial expressions and postures as symbols. It's a special language that communicates by action/vision[2]. CSL is different from Chinese and other sign languages.

In recent years, as the research of the multi-function perception, intelligent human-computer interface and virtual reality proceeds, sign language synthesis and recognition research has been paid more and more attention. Some universities, companies and research departments have studied and developed sign language synthesis and recognition systems[3–5].

The difficulties of machine translation of Chinese-CSL lie in solving the differences in word conception system, sentence conception system and modality of language expression.

2 Chinese-CSL Machine Translation Program

It needs three procedures to translate Chinese into CSL: analysis procedure, synthesis procedure and animation display procedure, i.e., grammar-analyzing procedure of Chinese, codes-producing procedure of CSL and animation display procedure of CSL (see Fig.1). The Chinese grammar analysis is composed of morphology analysis and syntax analysis.

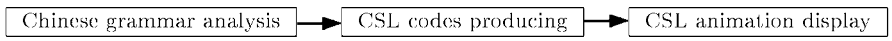

Fig.1. Chinese-CSL translation procedure.

2.1 Text-Based Chinese-CSL Transformation

To translate a Chinese sentence into CSL, we should recognize words from that sentence first. Actually, all translations must recognize words in advance, which is the task of morphology analysis (see Fig.1).

Another task that needs to be completed in morphology analysis is identifying the grammar attribute of each word in the context.

The third important task in morphology analysis is judging punctuation mark in the end of the sentence, which decides the mood of the sentence. For example, declarative sentence, exclamation sentence and interrogative sentence are represented by full stop, exclamation mark and question mark respectively. In the sentence of CSL we use facial expression to represent the mood of the sentence. The same sign language sentence can represent different moods if it is accompanied by different facial expressions.

After morphology analysis, we do syntax analysis. In this stage, the sentence structure of CSL needs to be produced. Because Chinese and CSL do not have the same word order or word number, word order adjustment and word adding and deleting are indispensable. The rules are as follows:

Noun and its modifier order rule

noun modifier + noun $\rightarrow$ noun + noun modifier

Predicate verb and its objects order rule

predicate verb + object $\rightarrow$ object + predicate verb

Question pronoun and object questioned rule

question pronoun + object questioned $\rightarrow$ object questioned + question pronoun

Measure word omitting rule

measure word $\rightarrow$ (blank)

Then, we'll think about the problem of how to find out the equivalent in CSL to every Chinese word. We can treat this problem by looking up dictionaries including Chinese dictionary and gesture synthesis dictionary. In practice, we use Chinese words to mark sign words of CSL in gesture synthesis dictionary.

As mentioned above, there is great difference in the number of words of these two languages. How can we figure out their corresponding relation? It can be discussed from the following aspects:

1) The situation that Chinese word has equivalent in CSL, i.e., the Chinese word is the mark of the sign word;
2) The situation that Chinese word has no equivalent in CSL, but its synonym has;
3) The situation that neither the word nor its synonym has equivalent in CSL.

To counter the above-mentioned situations, we propose the following equivalent transformation rules suitable for words.

For the first situation, the Chinese words can be mapped into the sign words directly. For the second situation, the Chinese words can be replaced by its synonyms that can be mapped into sign words directly. For the third situation, it needs to be divided into the following concrete situations:

1) The word is measure word;
2) The word is connective or auxiliary word of structure;
3) The word is well-known proper noun which represents the person's name or city's name;
4) The word is punctuation mark;
5) Other situations.

If it's the first situation, the word is omitted in CSL. If it's the second situation, the word is replaced by corresponding finger letters. If it's the third situation the word is replaced by initial consonant.

In the situation 4), if it's comma or slight pause, it is expressed by natural pause; if it's title mark, single (double) quotation marks , ellipsis or parentheses and square brackets, it is expressed by motioning the corresponding shape in the space.

According to the self-evident truth of language "all symbols of any natural language can be explanatorily described by symbols frequently used", we can explanatorily describe the words of situation 5), then figure out the equivalent in CSL of Chinese words present in the explanatory sentence.

In the end, the obtained codes of CSL sentences are transformed in modality, and they are displayed in computer window.

2.2 Model-Based Modality Transformation

In Subsection 2.1, after the word transformation, modality transformation is indispensable. That's another difficult key problem of translating natural language Chinese in the form of text into visual language CSL. CSL is a multi-channel parallel language. Its parallelism is embodied not only in time but also in space. The multi-channel of CSL involves gesture channel and expression channel. Only the simultaneous combination of the information of these channels can express a complete concept. The grammatical information of CSL is mainly conveyed through changes in the movement and spatial contouring of the hands and arms. Besides the motion information of hand, the changes of facial expression and the movement of body play an important part in grammar, too.

Treating so complicated information is very difficult. In Subsection 2.1, we suppose that each sign word in CSL is marked by only one equivalent in Chinese. After analysis and transformation, we get the result of the Chinese word sequence used to mark CSL. Thus modality transformation is how to transform that Chinese word sequence into gesture sequence and corresponding expression of CSL. That can be realized through gesture synthesis and expression synthesis.

2.2.1 Gesture Model

The grammatical information of CSL is mainly conveyed through changes in the movement and spatial contouring of the hands and arms. Hand has rich expression power. Two hands can express thousands of sign language words. So it's an important task to model hand and arm.

According to the physiological features of the hand and arm, we define three joints and four really used degrees of freedom (DOF) for each finger, two DOFs for wrist joint, two DOFs for elbow joint and three DOFs for shoulder joint. Thus we get 18 joints and 27 DOFs for one hand and arm totally as shown in Fig.2.

There are many individual states for each DOF. Then we can represent a specific state by a tuple having four elements:

$$(J_{\text{hand}}, J_{\text{wrist}}, J_{\text{elbow}}, J_{\text{shoulder}})$$

where

$$J_{\text{hand}} = \begin{pmatrix} j_{11} & j_{12} & j_{13} & j_{13'} \\ j_{21} & j_{22} & j_{23} & j_{23'} \\ j_{31} & j_{32} & j_{33} & j_{33'} \\ j_{41} & j_{42} & j_{43} & j_{43'} \\ j_{51} & j_{52} & j_{53} & j_{53'} \end{pmatrix} \qquad \begin{array}{l} J_{\text{wrist}} = (j_{w1}, j_{w2}) \\ J_{\text{elbow}} = (j_{fa1}, j_{fa2}) \\ J_{\text{shoulder}} = (j_{ua1}, j_{ua2}, j_{ua3}) \end{array}$$

Each j indicates a DOF, represented by joint angle. Therefore to specify one gesture demands 54 parameters.

We use T_n to represent the transformation matrix from gesture $n-1$ to gesture n.

According to this model, if initial matrixes T_0 and J_0 are given, at the time point n, we can get the current state of hands by the following equation:

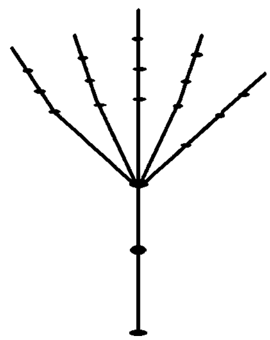

Fig.2. Model of hand and arm.

$$J_n = T_n J_{n-1}$$
$$T_n = T_n(J_0, J_1, \ldots, J_{n-1})$$

Transformation matrix is decided by the movement sequence of joints. It can be figured out by a recursive procedure:

$$T_n = f(T_{n-1})$$

Thus, we get the following equation:

$$J_n = \prod_{i=1}^{n} f(T_{n-i}) J_0$$

54 DOFs can decide a 54-dimension state space. We represent each movement state of hand and arm of one gesture action with the state of this state space. The change of state will result in the change of movements of hand and arm. Each sign word of CSL consists of an action sequence. It is modeled as a state sequence. Each sentence of CSL is a sign word sequence that is constituted according to the grammar rules.

2.2.2 Facial Expression Model

(1) Physics-based Facial Expression Model

In this paper, we take a simplified physics-based human face model. The model consists of skin model, muscle model and characteristic points on skin as shown in Fig.3.

Facial expression is produced by muscle's pulling. Each muscle can be simplified to a muscle vector, which has its own contract direction and action range. AU (Action Unit) is used to define the change of facial expression. There are 44 AUs capable of independent movement. There are six types of basic facial expressions, representing anger, abhorrence, fear, happiness, sadness and surprise respectively. An uneven 3D-mesh model is created on the skin. The points on the 3D mesh concentrate on the areas where there are great changes in the facial expression. Less changing areas are covered with large polygons. We define a group of characteristic points on skin, which consist of mesh points that reflect expression characteristics fully while expression altering.

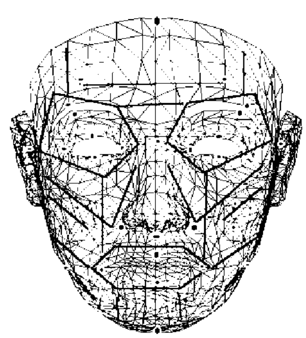

Fig.3. Model of facial expression.

(2) Product of Facial Expression in CSL

Each basic expression can be composed of several AUs' combination.

anger	$= \mathrm{AU}25^- + \mathrm{AU}4^- + \mathrm{AU}9^-$
abhorrence	$= \mathrm{AU}12^- + \mathrm{AU}4^- + \mathrm{AU}9^-$
fear	$= \mathrm{AU}12^- + \mathrm{AU}1^-$
happiness	$= \mathrm{AU}6^- + \mathrm{AU}12^-$
sadness	$= \mathrm{AU}1^- + \mathrm{AU}15^-$
surprise	$= \mathrm{AU}26^- + \mathrm{AU}1^-$

"AU number" represents the name of action unit. Superscript"$^-$" indicates contraction.

Here the facial skin is considered as ideal elastic object, which produces no inelastic eternal deformation. Expansion and contraction of muscles can be simulated by adjusting the coordination of human face characteristic points.

Based on these six types of basic expressions, we can make use of fuzzy subset concept to describe expression factor contained by each sign word.

Suppose $U = \{u_0, u_1, u_2, u_3, u_4, u_5, u_6\}$ is domain, where $u_0, u_1, \ldots, u_6$ is the seven types of expressions (six types of basic expressions plus neutral expression). Suppose V is the set of sign words, $F(U)$ is the fuzzy product set in fixed domain. Define map $F : V \to F(U)$. For each sign word a, there is a corresponding subset c. $F_c(u)$ indicates the subordination degree of u belonging to c.

Fuzzy subset c can be described as follows:

$$c = c(u_0)/u_0 + c(u_1)/u_1 + \cdots + c(u_6)/u_6$$

Here $c(u_i)/u_i$ $(i = 0, 1, 2, \ldots, 6)$ indicates the corresponding relation between element in the domain and its subordination degree.

The meshes of the expression contained in the word can be approximately considered as the combination of seven types of 3D meshes. The weight of each basic expression is direct proportional to the subordination degree of subordination function that the word corresponds to.

The concrete form that each word corresponds to is given, i.e., for arbitrary $a \in V$,

$$f(e_0, e_1, \ldots, e_6) = \sum_{i=0}^{6} k_i e_i$$

where k_i is the approximation degree in which the expression contained by sign word a is similar to the i-th basic expression, where e_i $(i = 0, 1, 2, \ldots, 6)$ is the 3D meshes of the seven types of basic expressions. Each word corresponds to a group of k_i $(0 \le k_i \le 1$, $i = 0, 1, \ldots, 6)$ that is called expression factor.

3 Experiment

According to the method mentioned in the paper, we have implemented a Chinese-CSL machine translation system.

Gesture synthesis dictionary saves motion description data of each sign word of CSL. Each item is composed of sign name, expression mark, state number, gesture data frames and expression factors. Each frame of data is composed of 54 data fields, including shoulder joint data, elbow joint data, wrist joint data and finger joint data. The key word of an item is its sign name. Each item is actually the motion description of the sign word. Expression mark indicates whether the gesture sign is accompanied by the expression or not. No expression means neutral expression. When there is expression accompanied, we get the corresponding expression factors to drive the expression module to produce corresponding expression. There are 3,330 items in gesture synthesis dictionary. All the sign words in the gesture synthesis dictionary are frequently-used ones according to "Current Chinese Frequency Dictionary".

Regarding the length of the paper, we only give a translation implementation of one Chinese word. Fig.4 is the pictures of sign word "guard". It consists of four actions. The "guard" is described as follows: stretch the index of one hand and point at the temple, show the alert appearance; and spread both palms to express "defense".

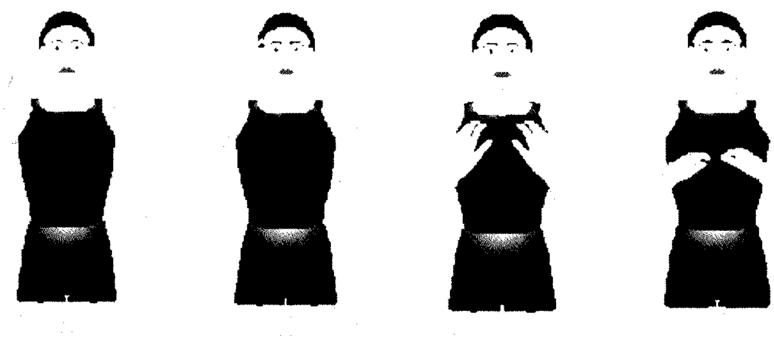

Fig.4. Gesture pictures of "guard".

4 Conclusion

In this paper, a text-based transformation method of Chinese-CSL machine translation is proposed. Gesture and facial model are created. And a practical system is implemented. The input of the system is text. The output of the system is "graphics person" who can gesticulate sign language accompanied by facial expression that corresponds to the text entered so as to realize automatic translation from text to sign language. The method is demonstrated to be practical and efficient.

References

[1] Gao Wen. Framework of multi-function perception machine. In *Proceedings of the Second Chinese Computer Intelligent Interface and Application Conference,* 1995, pp.7–20.
[2] Chinese Deaf-Mutes Society. Chinese Sign Language. Beijing: Hua Xia Press, 1995. (in Chinese).
[3] Badler N I, Morris M A. Modeling flexible, articulated objects. In *Proceeding of the ONLINE Conference,* 1982, pp.305–314.
[4] Jeffrey Loomis, Howard Poizner, Ursula Bellugi *et al.* Computer graphics modeling of American sign language. *Computer Graphics,* 1983, 17(3): 105–114.
[5] Masaru Ohki, Hirohiko Sagawa *et al.* Sign language translation system using pattern recognition and synthesis. *Hitachi Review,* 1995, 44(4): 251–254.