CSLDS: Chinese Sign Language Dialog System

Yiqiang Chen[a], Wen Gao[a, b], Gaolin Fang[b], Changshui Yang[a], Zhaoqi Wang[a]
[a]Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China
[b]Department of Computer Science, Harbin Institute of Technology, 150001, China
{ yqchen, wgao, glfang, csyang, chlwang }@jdl.ac.cn

1. Introduction

This paper presents a Chinese sign language dialog system (CSLDS) based on the technique of large vocabulary continuous Chinese sign language recognition (CSLR) and Chinese sign language synthesis (CSLS). This system can show the advance technology on gesture recognition and synthesis well and can apply to more powerful system combined with speech recognition and synthesis technology, which then can allow the convenient communication between deaf and hearing society (Figure 1).

Fig 1. **Deaf and Hearing Person Dialog System**

2. Chinese Sign Language Recognition

The aim of sign language recognition (SLR) is to provide an efficient and accurate mechanism to transcribe sign language into text. Most researchers mainly focus on the recognition of sign language with small or medium vocabulary size [1][2]. The major challenges that SLR faces now are developing methods that solve large vocabulary and continuous sign problems. In our system, these problems are addressed [3]. We use two Cybergloves and three Pohelmus 3SPACE-position trackers as input devices (Figure 2).

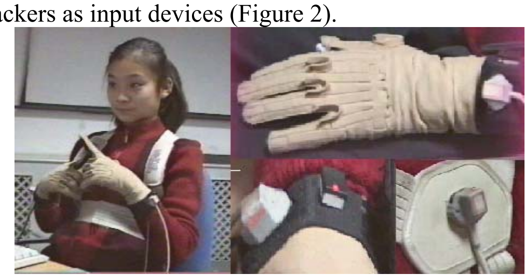


Fig 2. **The Input Device**

The system consists of two key techniques to deal with large vocabulary and continuous sign recognition problems [4]. That means, a fuzzy decision tree with heterogeneous classifiers is presented for large vocabulary sign language recognition. Gaussian mixture models based one- or two- handed classifier and finite state machine based hand shape classifier with little computational cost are first used to eliminate many impossible candidates, and then SOFM/HMM classifier in which the self-organizing feature maps (SOFM) being as an implicit different signers' feature extractor for continuous hidden Markov models (HMM) to allow some signer independence, is proposed as a special component of fuzzy decision tree to get the final results at the last non-leaf nodes that only include few candidates. To alleviate the effect of crisp classification errors, fuzzification is introduced in the decision tree. Figure 3 shows the results:

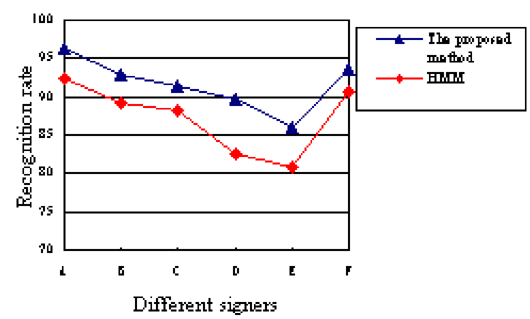


Fig 3. **The recognition results on 5113 signs**

Experimental data consist of 61356 samples over 5113 signs from 6 signers with each performing signs twice. Experimental results on a large vocabulary of 5113-signs show our proposed method has an average recognition rate of 91.6% using cross validation test. The average recognition time is 0.263 second per word and is suitable to the real-time recognition requirement.

For continuous SLR, the main issue is how to handle the movement epenthesis. To overcome this problem, large vocabulary continuous SLR based on transition movement models is proposed in this system. The proposed method employs the temporal clustering algorithm to cluster transition movements between two signs, and then the corresponding training algorithm of transition movement models is also presented.

Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG'03)
0-7695-2010-3/03 $ 17.00 © 2003 IEEE

Table 1. **The accuracy for continuous SLR**

Method	Accuracy	Time
Transition movements models	90.8% (S=279, I=53, D=127)	1.29 second/word

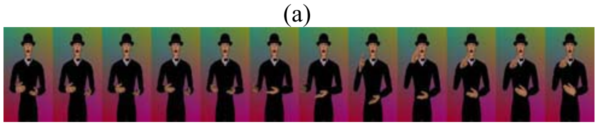
Continuous sign language database consist of the 1500 sentence samples with 750 different sentences over a vocabulary of 5113 signs. Experimental results show that continuous SLR has good performance with a recognition rate of 90.8%(S=279, I=53, D=127), where S, I and D denote the error numbers of substitution, insertion and deletion, respectively.

3. Chinese Sign Language Synthesis

In sign language synthesis system, the input is a meaningful sentence, and the output is a text-to-speech synchronized voice as well as facial animation and sign language corresponding to the input sentence [5].

For our CSLS system, We record the motion of 5500 sign words of Chinese sign language (CSL), and then edit every word we recorded by using a motion editing software we developed---GestureEdit, that can cut any frame of a motion or modify the shape and position of the hand in a frame. Of course, you need not to modify every frame but key frame. The other frames will be calculated automatically with some interpolation methods. (Figure 4)

(a)



(b)

Fig 4: **(a)The Original Display of Word "Ke Guan" (b)The Modified Display of Word "Ke Guan"**

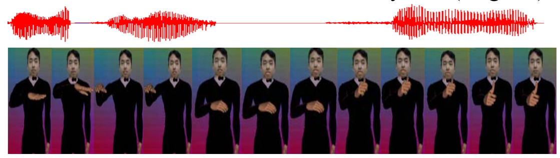
To generate the motion of a sentence, we should concatenate the motion-snippets of sign words. To make this concatenate more nature and realistic, the interpolation method based on quaternion is adopted in our system. (Figure 5)

(a) (b) (c)

Fig 5: **(a) The Last Frame of Word "DaJia" (b) The First Frame of word "Hao" (c) The Interpolation Results of Word "Dajia Hao"**

The synchronization is very important in sign synthesis system. In our system, a new visual prosody time control model is proposed to realize synchronization among speech synthesis, face animation and gesture animation [6]. The duration of visual prosody modal was obtained by incorporating the time information of recorded sign words and the predicted prosodic pattern which including time information for speech synthesis. Thus the visual prosody not only accord with the recorded gesture time to make the gesture animation more realistic, but also accord with the prosody rules of natural speech to make the gesture animation more individuate. At the same time, the visual prosody will feed back to text to speech units, which not only make the synthesis speech more nature, but also make the synthesis speech can be harmonize with gesture.(Figure 6)

Text: 大家好（Chinese）Hello, Every one (English) (a)



(b)

(c)

Fig 6. **(a) Input text (b) The synthesis speech (c) Key frame for animation of sign language**

The CSLS system was given a readable score of 92.98% for visual and understanding finger spelling, 88.98% for words, 87.58% for sentences by the students from these Deaf-mute Schools.

4. Conclusions and Future Work

All the experiments show our systems performance well, and future work will include making systems more robust and vivid.

5. Reference

[1] T. Starner, J. Weaver, and A. Pentland, "Real-time American sign language recognition using desk and wearable computer based video," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 12, pp. 1371-1375, 1998.

[2] C. Vogler and D. Metaxas, "A framework for recognizing the simultaneous aspects of American sign language," Computer Vision and Image Understanding, vol. 81, no. 3, pp. 358-384, 2001.

[3] W. Gao, J.Y. Ma, J.Q. Wu, and C.L. Wang, "Sign language recognition based on HMM/ANN/DP," Int'l J. Pattern Recognition and Artificial Intelligence, vol. 14, no. 5, pp. 587-602, 2000.

[4] Gaoling Fang, Wen Gao, Chunli Wang, Yiqiang Chen, "SOFM/HMM Based Chinese Sign Language Recognition". Journal of Computer, Vol.25, No.1, pp.16-22, 2002. (In Chinese)

[5] J. Cassell, Hannes Vihjalmsson, Timothy Bickmore, BEAT: the Behavior Expression Animation Toolkit, In: Proceedings of SIGGRAPH01, 2001.

[6] Yiqiang Chen, Wen Gao, Zhaoqi Wang, Changshui Yang, Dalong Jiang, "Text to Aatar in Multimodal Human Computer Interface", Asia-Pacific Human Computer Interface (APCHI2002), Vol.2, pp636-643, 2002.