A Comparison Between Etymon- and Word-Based Chinese Sign Language Recognition Systems*

Chunli Wang[1,2], Xilin Chen[1], and Wen Gao[1]

[1] Institute of Computing Technology, Chinese Academy of Science, 100080, Beijing, China
[2] Department of Computer Science and Engineering,
School of Electronic and Information Engineering,
Dalian University of Technology, 116023, Dalian, China
{clwang, xlchen, wgao}@jdl.ac.cn

Abstract. Hitherto, one major challenge to sign language recognition is how to develop approaches that scale well with increasing vocabulary size. In large vocabulary speech recognition realm, it is effective to use phonemes instead of words as the basic units. This idea can be used in large vocabulary Sign Language recognition, too. In this paper, Etyma are defined to be the smallest unit in a sign language, that is, a unit that has some meaning and distinguishes one sign from the others. They can be seen as phonemes in Sign Language. Two approaches to large vocabulary Chinese Sign Language recognition are discussed in this paper. One uses etyma and the other uses whole signs as the basic units. Two CyberGloves and a Pohelmus 3-D tracker with three receivers positioned on the wrist of CyberGlove and the back are used as input device. Etymon- and word- based recognition systems are introduced, which are designed to recognize 2439 etyma and 5100 signs. And then the experimental results of these two systems are given and analyzed.

1 Introduction

Sign language, as a kind of structured gesture, is one of the most natural means of exchanging information for most deaf people. It is a kind of visual language via hand and arm movements accompanying facial expressions and lip motions. The aim of sign language recognition is to provide an efficient and accurate mechanism to translate sign language into text or speech.

Attempts to automatically recognize sign language began to appear at the end of 80's. T.Starner [1] achieved a correct rate of 91.3% for 40 signs based on the image. By imposing a strict grammar on this system, the accuracy rates in excess of 99% were possible with real-time performance. Fels and Hinton [2][3] developed a system using a VPL DataGlove Mark II with a Polhemus tracker as input devices. Neural network was employed for classifying hand gestures. R.H.Liang and M. Ouhyoung[4] used HMM for continuous recognition of Tainwan Sign language with a vocabulary between 71 and 250 signs by using Dataglove as input devices. C. Wang[5] realized a continuous Chinese Sign Language (CSL) recognition system with a vocabulary of 5100 signs. C. Vogler and D. Metaxas[6] described an approach to continuous,

* This research is sponsored by Natural Science Foundation of China (No. 60533030).

S. Gibet, N. Courty, and J.-F. Kamp (Eds.): GW 2005, LNAI 3881, pp. 84 – 87, 2006.

whole-sentence ASL recognition, in which phonemes instead of whole signs were used as the basic units. They experimented with 22 words and achieved similar recognition rates with phoneme-based and word-based approaches.

From the review of the previous researches above mentioned, we know that most researches on continuous sign language recognition were done on small test vocabulary. The major challenge to sign language recognition is how to develop approaches that scale well with increasing vocabulary size. In speech recognition, using phonemes as basic unit assuredly is an effective solution to large vocabulary system. Is this idea also useful in large vocabulary sign language recognition?

In this paper, we discuss two approaches to large vocabulary Chinese Sign Language (CSL) recognition. One uses etyma and the other uses whole signs as the basic units. The results of these two approaches are compared.

2 Etymon-Based System

Two CyberGloves and a Pohelmus 3-D tracker with three receivers positioned on the wrist of CyberGlove and the back are used as input device. The raw gesture data include hand postures, positions and orientations. A sign is a sequence of frames. A frame of the raw gesture data, which in our system are obtained from 36 sensors on two datagloves, and three receivers mounted on the datagloves and the waist, are formed as 48-dimensional vector. A dynamic range concept is employed in our system for satisfying the requirement of using a tiny scale of data. The dynamic range of each element is different, and each element value is normalized to ensure its dynamic range 0-1.

Here, one etymon is defined to be the smallest unit in a sign language, that is, a unit that has some meaning and distinguishes one sign from another. For example, "Teacher" is composed by two etyma, which are shown in Fig. 1. The Bopomofo are considered as etyma, which can facilitate the CSL recognition when finger-alphabet is used accompanying with gestures. Unlike the etyma in spoken language, no explicit definition of the etymon exits in the CSL linguistics. Based on extensive and thorough analysis of 5100 signs in CSL, we find all the units that form all the signs in the CSL dictionary. Finally, about 2400 etyma are explicitly defined for CSL.

The sign data collected by the gesture-input devices is fed into the feature extraction module, and then the feature vectors are input into the training module, in which a model is built for each etymon. The signs are encoded based on the etyma, and the Etymon-sequences of signs are stored in a codebook, based on which the tree-structured network and forward index tables are built to reduce the search range [7].

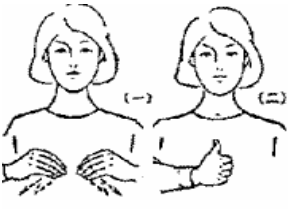

Fig. 1. Two Etyma in the word "Teacher"

The language model that is used in our system is Bi-gram model. The decoder controls the search for the most likely priority of sign appearance in a sign sequence. Then the sign sequence is exported from the decoder.

3 Experimental Results Comparison

In order to compare these two methods, some experiments are carried out. 5100 signs in CSL dictionary are used as evaluation vocabulary. Each sign was performed five times, four times are used for training and one for testing. 2439 etyma are defined for CSL. Each Etymon was performed five times for training. The number of states in HMMs is 3. 5100 signs are coded with these etyma automatically. Experiments are done based on word and etyma respectively.

One sign consists of one or more etyma. Each sign is a string of etyma. Therefore, there is movement epenthesis between two etyma, which will affect the recognition results. For the system based on etyma, recognizing a sign is similar to recognizing a sentence based on signs, so the recognition rate of isolated signs will decline.

There are about 2400 basic units and 5100 signs in Chinese sign language. The numbers of candidates in these two approaches are equivalent because of the using of Viterbi-beam algorithm. Therefore, the time of loop in the system based on etyma is half of that in the system based on signs when selecting the candidates. In the following two parts, the times of loop are the same. Besides, the algorithm based on etyma is more complex, and it takes more time to decode. The approach based on etyma does not gain better effect on the aspect of improving the speed.

The 5100 signs and 2439 etyma are analyzed and it is discovered that many etyma have low appearing frequencies in all words, namely they only appear in one or two words. For example, the etymon “electric car” is a word by itself, and does not appear in any other words. If throwing off the etyma that appear less than twice in 5100 words, there are 723 etyma left. These etyma compose 3048 words. If deleting the etyma that appear less than three times in 5100 words, there are 581etyma and 2650 words left. Experiments are done with these three vocabularies, and the results are shown in Fig. 2.

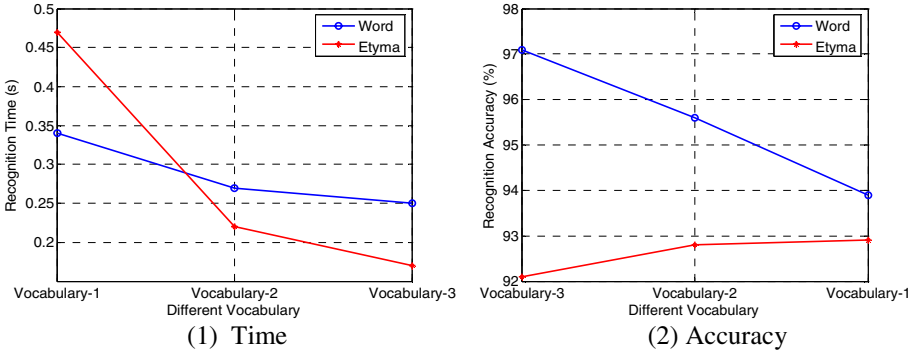

Fig. 2. The comparison of the performances of two systems based on different size of vocabulary. Vocabulary-1: 2439 Etyma and 5100 signs; Vocabulary-2: 723 Etyma and 3048 signs; Vocabulary-3: 581 Etyma and 2650 signs.

For the 5100 isolated words, the recognition rate of the approach based on etyma is lower than that of the approach based on whole signs. The recognition time based on etyma is longer than that based on words. But when the number of etyma with low appearing frequencies in all words decreases, the performances based on etyma are improved. In the case that the size of the vocabulary is large and the number of words is four times more than the number of etyma, the approach based on etyma is the proper selection.

4 Conclusions and Future Work

In speech recognition, using phonemes as basic unit assuredly is an effective solution to large vocabulary system. But is it also useful in sign language recognition? In this paper, two approaches to large vocabulary CSL recognition are introduced. Experimental results of these two approaches are compared.

There are, however, many problems that still need to be resolved. The impact of the movement between two signs is not eliminated, yet. Context-dependent Etymon (TRIPHONE) models will be built to solve this problem. But because the number of etyma is much more than that of spoken language, the effect won't be obviously. How to reduce the number of the etyma is the key.

References

1. Starner T., Weaver J., Pentland A.: Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video. IEEE PAMI, Vol 20, Issue 12, Dec 1998, pages 1371-1375.
2. S.S.Fels, G.Hinton.: GloveTalk:A neural network interface between a DataDlove and a speech synthesizer. IEEE Transactions on Neural Networks, 4(1993):2-8.
3. S.Sidney Fels.: Glove –TalkII: Mapping hand gestures to speech using neural networks-An approach to building adaptive interfaces. PhD thesis, Computer Science Department, University of Torono, 1994.
4. R.-H.Liang, M.Ouhyoung.: A real-time continuous gesture recognition system for sign language. In Proceeding of the Third International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 1998, pages 558-565.
5. Chunli Wang, Wen Gao.: A Real-time Large Vocabulary Recognition Continuous System for Chinese Sign Language. Advances in Multimedia Information Processing-PCM 2001, Beijing, China, October 2001, 150-157.
6. ChristianVogler, Dimitris Metaxas.: Toward scalability in ASL Recognition: Breaking Down Signs into Phonemes. In Proceedings of Gesture Workshop, Gif-sur-Yvette, France, 1999, pages 400-404.
7. Chunli Wang, Wen Gao, Shiguang Shan.: An approach based on phonemes to large vocabulary Chinese sign language recognition. Proceeding of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition (FG'02), Washington, USA, 2002, 411-416.