Signer Adaptation Based on Etyma for Large Vocabulary Chinese Sign Language Recognition

Yu Zhou[1], Wen Gao[1], Xilin Chen[2], Liang-Guo Zhang[2], and Chunli Wang[3]

[1] School of Computer Science and Technology, Harbin Institute of Technology, Harbin, 150001, China
[2] Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100080, China
[3] School of Computer Science and Technology, Dalian Maritime University, Dalian, 116026, China
{yzhou, wgao, xlchen, lgzhang, clwang}@jdl.ac.cn

Abstract. Sign language recognition (SLR) with large vocabulary and signer independency is valuable and is still a big challenge. Signer adaptation is an important solution to signer independent SLR. In this paper, we present a method of etyma-based signer adaptation for large vocabulary Chinese SLR. Popular adaptation techniques including Maximum Likelihood Linear Regression (MLLR) and Maximum A Posteriori (MAP) algorithms are used. Our approach can gain comparative results with that of using words, but we only require less than half data.

Keywords: SLR, signer adaptation, signer independency, MLLR, MAP.

1 Introduction

SLR aims to transcribe sign language to text so that the communication between deaf and hearing society can be convenient. SLR also helps to make the human computer interface more naturally.

By far the signer-dependent SLR has achieved remarkable results. T.Starner [1] achieved a correct rate of 91.3% for 40 signs based on the image. C.Vogler and D.Metaxas [2] described an approach to continuous, whole-sentence American SLR. C. Wang and W. Gao [3] realized a Chinese SLR system with a vocabulary of 5100 signs. But the signer-independent SLR is still a big challenge, especially in the large vocabulary SLR. The main obstacles in large vocabulary signer-independent SLR are the difficulty of availability to multi-signers' data and the vast variation of different signers' regions, habits, moods, etc. There are more than 5,000 words in Chinese sign language. It is impossible to constitute a database including all kinds of signers. Therefore, modeling one signer's sign language by sparse data becomes an urgent problem. We try to give an answer to this problem by signer adaptation based on etyma in this paper.

The rest of this paper is organized as follows. In Sect. 2, the signer adaptation method based on etyma is proposed. In Sect. 3, the experimental results are reported. Finally in Sect. 4, we give the conclusions and some possible future work.

H.H.S. Ip et al. (Eds.): PCM 2007, LNCS 4810, pp. 458–461, 2007.

2 Etyma-Based Signer Adaptation

We select etyma as the basic recognition units so as to decrease the adaptation data required, and etyma are modeled by HMMs. MLLR and MAP are used for adaptation. Now we briefly introduce MLLR and MAP, they are detailedly described in [4][5][6].

MLLR and MAP. MLLR supposes that model adaptation can be achieved by applying some affine transformations to the original model parameters. There is a regression class set in MLLR, and in each regression class the parameters share the same transformation. MLLR transforms HMMs' means by equation (1).

The conventional Maximum Likelihood (ML) estimation supposes the parameter that we want to estimate is fixed though it is not known. MAP estimation supposes that the parameter we want to estimate is under some probability distribution function, If conjugate priors are used, the MAP result of single Gaussian mixture's mean is as equation (2).

$$\hat{\mu} = A\mu + b = W\xi \tag{1}$$

$$\hat{\mu} = (N\overline{\mu} + \tau\mu)/(N + \tau) \tag{2}$$

where $\hat{\mu}$ is the adapted mean, μ is the original mean, A is the transformation matrix, b is the bias vector; $W = [A^T, b^T]^T$ is the extended transformation matrix, and $\xi = [\mu^T, 1]^T$ is the extended mean vector; τ is a weighting of the priori knowledge to the adaptation data, N is the occupation likelihood of the adaptation data, and $\overline{\mu}$ is the mean of the observed adaptation data.

Previous work. We have presented a method in which Chinese Sign Language can be recognized from models of etyma in [7]. In Chinese sign language, there are more than 5000 words in total. We break them down to about 2400 etyma. Each word is composed of 1 or more etyma in sequence. [7] has proven that the recognition accuracy of the approach based on etyma is comparable to that of the approach based on words.

Signer adaptation based on etyma. Since the approach based on etyma is comparable to that of the approach based on words, we can adapt the signer-independent system using the etyma data, whose vocabulary number is about half of the words data's vocabulary number. Our process is as below:

1) Train HMMs based on etyma;
2) Adapt HMMs using etyma data of the new speaker;
3) Generate HMM-Nets according to the word-etymon map list;
4) Recognize the test word data using HMM-Nets generated above.

Our experiments are executed with HTK [8]. Because we want to compare our method to conventional method on the level of word accuracy, we generate

HMM-Nets when we do experiments based on etyma. For more details about HMM-Nets, please read [8].

3 Experimental Results Comparison

We use etyma and words as basic units. Each etymon and word is performed four times by six signers. Four times data of all the other five signers except the test signer are used for training. The test signer's fourth time data is used for test, and the other times data are used for adaptation. We use cross validation and leave one out. The results of six signers are listed in Table 1. When we use MLLR, 1 time of adaptation data is used, and 3 times is used when we use MAP.

Table 1. The recognition accuracy comparison

Signer for test	Etyma based method			Words based method		
	SI%	MLLR%	MAP%	SI%	MLLR%	MAP%
S1	59.93	72.67	88.79	62.82	76.15	94.78
S2	56.28	68.74	90.35	60.03	73.59	95.43
S3	62.55	72.90	91.11	65.68	76.73	96.09
S4	62.55	72.87	90.99	65.36	76.46	95.80
S5	65.48	75.00	91.01	69.94	79.78	96.31
S6	59.13	71.79	88.75	62.67	74.84	94.10
Average	60.99	72.33	90.17	64.42	76.26	95.42

We only execute the global transformation adaptation during MLLR. We can find that all of the two methods improve the average recognition accuracy of words by about 12%, but the method of adaptation based on etyma only needs about half of the data compared to that of adaptation based on words.

MAP can gain better results when the adaptation data become rich, so we use 3 times data to execute MAP adaptation. Experiment results show that MAP can improve the average recognition accuracy by about 30%. Still the method we presented can save about half data.

We also experiment with different times of adaptation data and compare the accuracy improved by MLLR and MAP. Fig.1 shows the results. The improvements by MAP are much better than that of MLLR, because the adaptation data we used (even 1 time data) are plenty enough for MLLR global transformation but not for MAP.

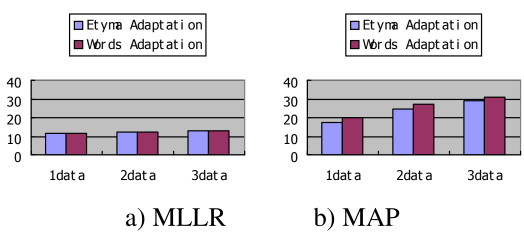

Fig. 1. The comparison between MLLR and MAP

4 Conclusions and Future Work

Adapting large vocabulary Chinese sign language based on etyma can save more than half data compared to adaptation based on whole words with both MLLR and MAP algorithms. MLLR should be adopted when adaptation data is not enough(for example only one time data or less), and MAP should be adopted after more adaptation data was collected. Breaking words to etyma is a proper way to solve signer-independent Chinese SLR.

In the future, we will break the etyma to phonemes to reduce the adaptation data required further; we will apply the adaptation to continuous Chinese SLR; we will also do some work on unsupervised adaptation.

Acknowledgment

This research is partially sponsored by Natural Science Foundation of China under contract No.60303018 and No.60603023, and Natural Science Foundation of Beijing Municipal under contract No.4061001.

References

1. Starner, T., Weaver, J., Pentland, A.: Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video. IEEE PAMI 20(12), 1371–1375 (1998)
2. Vogler, C., Metaxas, D.: Toward scalability in ASL Recognition: Breaking Down Signs into Phonemes. In: Proceedings of Gesture Workshop, Gif-sur-Yvette, France, pp. 400–404 (1999)
3. Wang, C., Gao, W., Ma, J.: A Real-time Large Vocabulary Recognition System for Chinese Sign Language. Gesture and Sign Language in Human-Computer Interaction, 86–95 (April 2001)
4. Leggetter, C.J., Woodland, P.C.: Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models. Computer Speech and Language, 171–185 (September 1995)
5. Gales, M.J.F.: Maximum Likelihood Linear Transformations for HMM-Based Speech Recognition. Computer Speech and Language, 75–98 (December 1998)
6. Gauvain, J.L., Lee, C.H.: Maximum a Posteriori Estimation for Multivariate Gaussian Mixture Observations of Markov Chains. IEEE Transactions on Speech and Audio Processing 2(2), 291–298 (1994)
7. Wang, C., Chen, X., Gao, W.: A Comparison Between Etymon- and Word-Based Chinese Sign Language Recognition Systems. In: Gibet, S., Courty, N., Kamp, J.-F. (eds.) GW 2005. LNCS (LNAI), vol. 3881, pp. 84–87. Springer, Heidelberg (2006)
8. Young, S., Evermann, G., Kershaw, D., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., Woodland, P.: The HTK Book (for HTK Version 3.2), pp. 161–177. Cambridge University (December 2002)