Communication Tool for the Hard of Hearings

A Large Vocabulary Sign Language Recognition System

Xiujuan Chai, Hanjie Wang, Fang Yin, Xilin Chen
Key Lab of Intelligent Information Processing of Chinese Academy of Sciences(CAS),
Institute of Computing Technology, CAS, Beijing, 100190, China
Cooperative Medianet Innovation Center, China

***Abstract*—Deaf person has a large social community around the world. The smooth communication is very difficult for these hard of hearings. Automatic Sign Language Recognition (SLR) can build the bridge between the deaf and the hearings and turn the seamless interaction into reality. This paper presents a visualized communication tool for the hard of hearings, i.e. a large vocabulary sign language recognition system based on the RGB-D data input. A novel Grassmann Covariance Matrix (GCM) representation is used to encode a long-term dynamics of a sign sequence and the discriminative kernel SVM is adopted for the sign classification. For continuous sign language recognition, a probability inference method is used to determine the spotting from the labels of sequential frames. Some basic evaluation and comparison of our recognition algorithms are conducted in our collected datasets. This demo will show the recognition of both isolated sign words and the continuous sign language sentences.**

***Keywords*—Sign language recognition; sentence spotting; Grassmann covariance matrix; kernel SVM; RGB-D data**

I. INTRODUCTION

There are more than 360 million hard of hearings around the world. For such a large community, the natural interaction with hearings becomes a serious social problem. In recent years, there are more and more researchers who explore SLR related technologies.

From the view point of input devices, the researches experienced the change of data glove, color glove and pure vision-based device [1-3]. The SLR with common 2D camera is very challenging since the hand tracking and segmentation is very difficult when confronted with the complex backgrounds and illuminations. Fortunately, there appears some devices which can provide the RGB and depth information simultaneously, such as the Kinect, Intel RealSence, and so on. Our demo system in this paper is based on Kinect device.

In consideration of the big successes in the field of speech recognition, HMM and its variations were introduced into SLR area [4-7] and have been dominant methods especially in the early research. Other main stream methods include DTW, CRF and so on [8-10]. These traditional methods did work when confronted with low-dimensional feature vectors and large plenty of training data. However, the visual-based features usually possess high dimensions, which make the model leant from limited training samples less effective. A lot of other methods have sprung up in recent years. E. Ong et al. proposed a multi-class classifier based on sequential pattern trees in 2012 [11] and extended it for continuous sign language sentence recognition [12]. As a typical discriminative method, SVM was also used for SLR and achieved good performance [13,14].

Fig.1 The main interface of our SLR system.

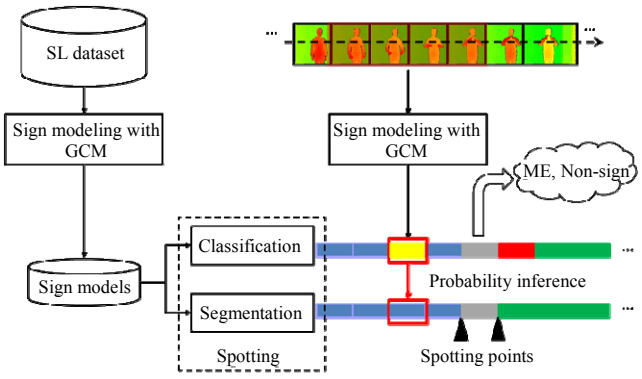

Fig.2 The main modules and framework of our SLR system .

In our work, a novel Grassmann Covariance Matrix (GCM) representation is proposed for sign modeling. Kernel SVM classifiers are trained and used to get the sign label. In continuous sign language sentence recognition, we extend our GCM hierarchically and use the probability inference technique to get the accurate spotting results. These methods are all integrated into our SLR system.

II. SIGN LANGUAGE RECOGNITION ALGORITHMS

A. Overall Framework

Figure 2 gives the overall framework and the main modules of our SLR system. With the plenty of training data, sign models are generated with GCMs. Then according to the different tasks, the classification or the spotting results can be obtained with our proposed methods.

This work was partially supported by the Microsoft Research Asia, and Natural Science Foundation of China under contract Nos. 61390511 and 61472398.

978-1-4799-9953-8/15/$31.00 ©2015 IEEE

B. GCM

To model the long-term relation of each sign, the covariance matrix is adopted in our method. The feature vector for each frame is denoted by f , whose dimension is D. Then the covariance can be computed according to Eq.(1):

$$\mathrm{C} = \frac{1}{M-1}\sum_{\mathrm{i}=1}^{M}(f_i - \mu)(f_i - \mu)^T, \quad (1)$$

where, M is the total frame number of the vedio clip and μ is the mean of all the feature vectors in this clip. Here, the vector f is the fusion of the motion (skeleton pair) and the hand shape (HOG) features.

To measure the distance between two covariance matrices, we recur to the principle angles of Grassmann manifold. The covariance matrix C is decomposed by SVD

$$\mathrm{C} = \mathrm{Y}\Sigma\mathrm{Y}, \quad (2)$$

and thus the orthogonal matrix $\mathrm{Y} = [\mathrm{y}_1, \mathrm{y}_2, \ldots, \mathrm{y}_\mathrm{D}]$ is obtained. The first d column vectors of Y are selected as ideal subspace of R^D , which is denoted as $\overline{\mathrm{Y}}$ and called as Grassmann Covariance Matrix (GCM).

C. Classification and Spotting

For sign classification task, kernel Support Vector Machine (SVM) is used for classifier training and testing. The distance between two GCMs is measured through the kernel projection as follows:

$$\mathrm{K}(\overline{\mathrm{Y}}_\mathrm{i}, \overline{\mathrm{Y}}_\mathrm{j}) = \left\| \overline{\mathrm{Y}}_\mathrm{i}^{\,T} \overline{\mathrm{Y}}_\mathrm{j} \right\|^2. \quad (3)$$

For continuous sentence spotting task, we can get the label and the corresponding reliability for each frame. Then a multiple-temporal belief propagation strategy is used to get the accurate segmentation according to the probability inference.

III. Experimental Results

To evaluate the proposed algorithms, we conduct the experiments on separate tasks and the details are given below.

A. Evaluation on Isolated Sign Word Recognition

In this experiment, we use two different datasets (D_A and D_B) to evaluate the algorithms under signer-dependent and signer-independent cases. D_A includes 1000 sign classes, and each word was recorded by the same signer with 5 repetitions. D_B has the same 1000 words played by 7 different signers. Each signer performed each sign only once.

In our experiments, the leave-one-out cross validation strategy is adopted. The experimental results are given in Table I and Table II respectively. In the experiments, the HMM [7], DTW [9] and ARMA [15] methods are used for comparison. It can be seen from the results that our method outperforms the-state-of-the-art methods tremendously.

B. Evaluation on Continuous Sentence Spotting

To evaluate the continuous sentence spotting method, we conduct the experiment on the dataset composed of 150 real continuous sign sequences with 242 signs. Here, HMM is the baseline used for comparison. The results are given in the Table III. In this table, "Corr" is the number of correctly detected signs. "Ins", "Del" and "Sub" correspond to the insertion, deletion and substitution errors. FPR and TPR are the False Positive Rate and True Positive Rate respectively. Compared with the traditional HMM, our method obtained higher TPR and lower FPR.

TABLE I. Signer-Dependent SLR Results on D_A

Methods	*HMM*	*DTW*	*ARMA*	*Our Method*
Accuracy(%)	83.2	87.0	89.1	**92.4**

TABLE II. Signer-Independent SLR Results on D_B

Methods	*HMM*	*DTW*	*ARMA*	*Our Method*
Accuracy(%)	56.2	49.8	63.2	**70.9**

TABLE III. Evaluation on Continuous Sentence Spotting

Methods	*Corr*	*Ins*	*Del*	*Sub*	*FPR %*	*TPR%*
HMM	447	235	173	251	**97.2**	**65.9**
Our Method	479	81	135	87	**44.7**	**70.6**

IV. Demo System on SLR

Based on above algorithms, a sign language recognition system is built to show the possibility of the normal communication between deaf and hearing. In this demo system, we will show two functions, which are the isolated sign word recognition and the sentence spotting respectively.

A. Isolated Sign Word Recognition

Currently, the system can realize the recognition of 1000 signs, most of which are the daily used words. Once the signer finishes the signing, the system will give the top five candidate results. If the first candidate is the right one, the system will output it and play the corresponding voice. Or else, the signer can revise the result according to the gesture interaction.

B. Continuous Sentence Spotting

In the mode of continuous sentence spotting, the user can signing several signs continuously and when the signing is finished, the system will give the recognition results for this sentence. Different from the performing of isolated sign, in the procedure of sentence signing, the hand needn't put down when one sign is ending. Thus the transition is existed between two signs, which is a big challenge in sentence spotting. Through a threshold constraining strategy for the transition problem, our system can give satisfied results.

V. Conclusion

In this paper, a tool aiming for the communication between the deaf and the hearings is developed and presented. The 3D motion and hand shape information are effectively fused by the proposed GCM representation. Further, novel algorithms are given to realize sign classification and sentence spotting. The experimental results on large vocabulary datasets convincingly show the good performance of the proposed methods. The lively demo has two main functions, which are the isolated sign word recognition and the real-time sign language sentence spotting. In the future, the special modeling for the MEs in the sentence spotting algorithm is our next step.

REFERENCES

[1] Q. Wang, X. Chen, L. Zhang, etc. Viewpoint Invariant Sign Language Recognition. Computer Vision and Image Understanding, vol.108, pp.87-97, 2007.

[2] G. Fang, W. Gao, D. Zhao. Large-vocabulary Continuous Sign Language Recognition based on Transition-Movement Models. IEEE Trans. on SMC, Part A: Systems and Humans, 37(1), pp. 1-9, 2007.

[3] M. Holte ,T. Moeslund, P. Fihl. Fusion of Range and Intensity Information for View Invariant Gesture Recognition. CVPR Workshops, 2008, pp. 1~7.

[4] W. Gao, G. Fang, D. Zhao, and Y. Chen. Transition Movement Models for Large Vocabulary Continuous Sign Language Recognition. FG, pp 553–558, 2004. 1

[5] V. Pitsikalis, S. Theodorakis, C. Vogler, and P. Maragos. Advances in Phonetics-based Sub-unit Modeling for Transcription Alignment and Sign Language Recognition. CVPR Workshop, pp. 1–6, 2011.

[6] R. Liang and M. Ouhyoung. A Sign Language Recognition System using Hidden Markov Model and Context Sensitive Search. ACM Symposium on Virtual Reality and Technology, pp. 59–66, 1996.

[7] C. Wang, W. Gao, and S. Shan. An Approach based on Phonemes to Large Vocabulary Chinese Sign Language Recognition. FG, pp. 411–416, 2002.

[8] H. Wang, A. Stefan, S. Moradi, V. Athitsos, C. Neidle, F. Kamangar. A System for Large Vocabulary Sign Search. ECCV Workshops (1), 2010, pp. 342–353.

[9] S. Celebi, A. Aydin, T. Temiz, and T. Arici. Gesture Recognition using Skeleton Data with Weighted Dynamic Time Warping. VISAPP, pp. 620–625, 2013.

[10] H.-D. Yang, S. Sclaroff, and S.-W. Lee. Sign Language Spotting with a Threshold Model based on Conditional Random Fields. PAMI, 31(7):1264–1277, 2009.

[11] E.-J. Ong, H. Cooper, N. Pugeault, and R. Bowden. Sign Language Recognition using Sequential Pattern Trees. CVPR, pp. 2200–2207. 2012.

[12] E.-J. Ong, N. Pugeault, and R. Bowden. Sign Spotting using Hierarchical Sequential Patterns with Temporal Intervals. CVPR, pp. 1931–1938. 2014.

[13] C. Sun, T. Zhang, B. Bao, C. Xu, T. Mei. Discriminative Exemplar Coding for Sign Language Recognition with Kinect. IEEE. Trans. on Cybernetics, 43(5), pp. 1418-1428, 2013.

[14] T. Pfister, J. Charles, and A. Zisserman. Domain-adaptive discriminative one-shot learning of gestures. ECCV 2014, pp. 814–829. 2014.

[15] C. Xu, T. Wang, J. Gao, S. Cao, W. Tao, and F. Liu. An Ordered Patch-Based Image Classification Approach on the Image Grassmannian Manifold. IEEE Trans. on *Neural Networks* and Learning Systems, 25(4), pp. 728–737. 2014.