Sign Language Recognition and Translation with Kinect

Xiujuan Chai, Guang Li, Yushun Lin, Zhihao Xu,
Yili Tang, Xilin Chen
Key Lab of Intelligent Information Processing
of Chinese Academy of Sciences (CAS),
Institute of Computing Technology, CAS
Beijing, China
{xiujuan.chai, guang.li, yushun.lin, zhihao.xu,
yili.tang, xilin.chen}vipl.ict.ac.cn

Ming Zhou
Microsoft Research Asia
Beijing, China
mingzhou@microsoft.com

Abstract—**Sign language (SL) recognition, although has been explored for many years, is still a challenging problem for real practice. The complex background and illumination conditions affect the hand tracking and make the SL recognition very difficult. Fortunately, Kinect is able to provide depth and color data simultaneously, based on which the hand and body action can be tracked more accurate and easier. Therefore, 3D motion trajectory of each sign language vocabulary is aligned and matched between probe and gallery to get the recognized result. This demo will show our primary efforts on sign language recognition and translation with Kinect.**

Keywords-sign language; hand tracking; 3D motion trajectory

I. INTRODUCTION

Sign language is the most important communication way between hearing impaired community and normal persons. In recent years, sign language has been widely studied based on multiple input sensors, such as data glove, web camera, stereo camera, and so on [1-3]. Although data glove based SL recognition achieves good performance even for large vocabularies, the device is too expensive to popularize. In vision-based SL recognition, the key factor is the accurate and fast hand tracking and segmentation. However, it is very difficult for the complex backgrounds and illuminations.

Different from these previous methods, our system aims to realize fast and accurate 3D SL recognition based on the depth and color images captured by Kinect.

II. 3D TRAJECTORY MATCHING FOR SIGN LANGUAGE RECOGNITION

The block diagram of our SL recognition algorithm is given in Figure 1. First, the 3D trajectory description corresponding to the input SL word is generated by hand tracking technology provided by Kinect Windows SDK [4]. Considering the difference of hand motion speed, a linear resampling is done to get the normalized trajectory by averaging the accumulated length of the whole vector. This operation aims to normalize the trajectory of each word into the same sampling point.
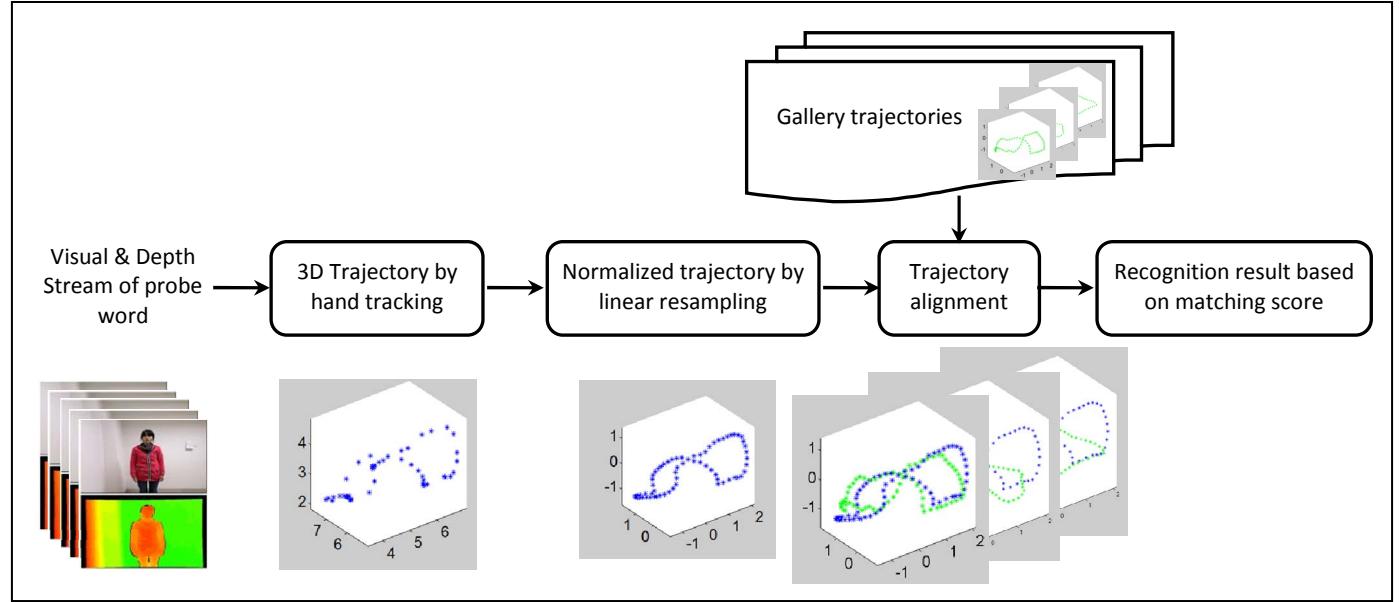

Figure 1. Block diagram of our 3D trajectory matching based sign language recognition method.

To perform the recognition, alignments between probe trajectory and gallery vectors are needed. And finally, the matching scores are computed according to the Euclidean distance measurement to give the recognition result.

To validate the performance of our sign language recognition algorithm, we conduct the experiments on a database which contains 239 Chinese SL words. The sign language corresponding to each word is recorded by 5 times. In our cross validation experiment, one group vocabularies are taken as probe and the other 4 groups of samples forms the gallery set. The rank-1 and rank-5 recognition rates are 83.51% and 96.32% respectively.

III.   SIGN LANGUAGE RECOGNITION AND TRANSLATION SYSTEM

Based on our proposed 3D trajectory matching algorithm, a sign language recognition and translation system is built to connect hearing impaired community and normal persons. The main functions of the demo system are given in figure.2 and figure. 3 shows screenshots of the demonstration of our sign language recognition and translation system.

Our system consists of two modes: Translation Mode (as shown in Figure 3. (a)), in which it translates sign language into text or speech; and Communication Mode (as shown in Figure 3 (b)), in which a normal person can communicate with the signer through an avatar.

Translation mode includes isolated word recognition and sentence recognition. In current system, raising and putting down hands are defined as the start and end gesture of each SL word to be recognized. The system gives the recognized rank-5
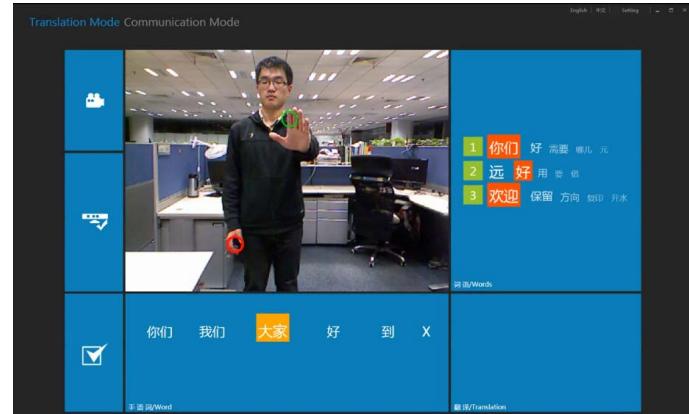
candidate words in the bottom region of the interface. If the rank-1 word is not the correct result, the signer can adjust the results by manual interaction. For sentence recognition, all the words can be input continuously and then the system will give the results by integrating both the SL matching score and the probability given by SL language model. Also, the signer can adjust or confirm the results by manual interaction.

In the Communication mode, an avatar can play the corresponding SL sentence from the keyboard text input. The impaired person makes an immediate response through sign language. The system translates the answer into text. Thus the normal person can communicate with the impaired person naturally.
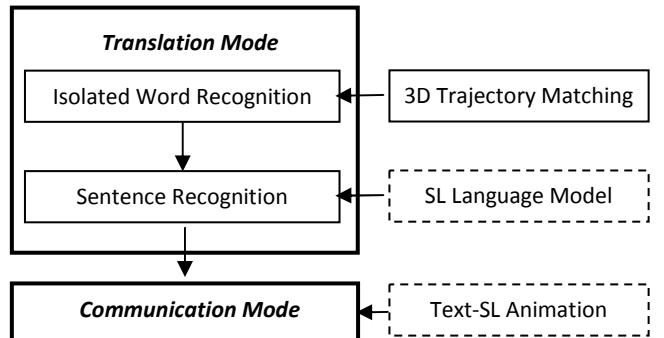

Figure.2. Main functions of our demo system.

(a)   Translation mode

(b) Communication mode

Figure 3. Screenshots of the sign language recognition and translation demonstration system.

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 S-Part A: Systems and Humans, vol, 37, no.1, pp. 1-9, 2007.

[3] M. Holte ,T. Moeslund, P. Fihl. "Fusion of range and intensity information for view invariant gesture recognition," In Proc. of the IEEE Computer Society Conf. on CVPR Workshops, 2008, pp. 1~7.

[4] J. Shotton, A. Fitzgibbon and M. Cook, etc. "Real-time human pose recognition in parts from single depth images," In Proc. of IEEE Conf. on CVPR, 2011, pp.1297-1304.