VisualComm: A Tool to Support Communication between Deaf and Hearing Persons with the Kinect

Xiujuan Chai, Guang Li, Xilin Chen
Institute of Computing Technology, CAS
Beijing, China
{xiujuan.chai, guang.li, xilin.chen}
@vipl.ict.ac.cn

Ming Zhou, Guobin Wu,
Microsoft Research Asia
Beijing, China
{mingzhou,guobinwu}
@microsoft.com

Hanjing Li
Beijing Union University
Beijing, China
tjthanjing@buu.edu.cn

ABSTRACT
With the quickly increasing of the deaf community, how to communicate with the hearing persons is becoming a serious social problem. Furthermore, the investigation indicates that the deaf community is more self-enclosed and won't exchange ideas with the hearing. To address this challenge, we develop VisualComm, a tool to support communication between deaf and hearing persons with sign language recognition technology by using the Kinect. The main contribution of the system is a holistic solution of a two-way communication between deaf and hearings, and furthermore it is a seamless experience tailored for this particular activity. Currently we have implemented the basic communication based on 370 daily Chinese words for signer.

Categories and Subject Descriptors
H.5.1[**Information Interfaces and Presentation (e.g., HCI)**]: Multimedia Information Systems – *Evaluation/methodology*.

General Terms
Algorithms, Design, Human Factors.

Keywords
Assistive Technology, Deaf Community, Communication Tool, Sign Language Recognition, Kinect

1. INTRODUCTION
Communication in daily life is a big challenge to the deaf community. Sign language is the most important communication way between deaf and hearing persons for its well understandability and insistency. Currently, for most hearing persons, sign language is difficult to understand, which results that the deaf is more isolated and they won't participate in social activity.

To address this challenge, we present VisualComm (Figure 1), a tool that assists the deaf community to communicate with the hearing. It's a two-way communication system that consists of translating the sign language to text or speech for hearing people, and conveying text or speech language to deaf person through a 3D avatar animation. Currently, the tool is able to support basic communication based on 370 daily Chinese sign words.

ASSETS'13, Oct 21–23 2013, Bellevue, WA, USA.

DOIs: http://dx.doi.org/10.1145/2513383.2513398

Figure 1. VisualComm System User Interface

2. RELATED WORK
Researchers have worked on various technologies to address communication barriers between the deaf and hearing community for many years, such as MobileASL [1], Shared Speech Interface (SSI) [4], etc. However, most of current systems don't support two-way communication between deaf and hearing individuals. Sometimes, a human sign language interpreter has to be involved [4]. Different from these works, our tool provides a holistic solution to support two-way communication automatically during the activity.

Sign language recognition (SLR) is a key technology of this solution. According to the input device, the research on SLR evolves the history from data glove to pure visual input[6][3]. However, the data glove is expensive and the visual camera makes the hand tracking and segmentation very difficult. Fortunately, Microsoft launched a motion-sensing camera called Kinect in 2010, which can provide color and depth data simultaneously and the body skeleton can be obtained easily [5]. It is a cheap but convenient 3D depth sensor. With these good features, Kinect is very fitful for sign language recognition task [7]. Therefore, we aim to develop a robust sign language recognition system by using Kinect to help deaf community to interaction with surroundings efficiently.

3. DESIGN & IMPLEMENTATION
Our goal for system design and implementation is as follows: a) two-way communication, that is to say, deaf and the hearing persons can communicate directly by using our tool. b) For the basic communication between them, frequently used words and sentences in daily scenes have to be recognized.

3.1 VisualComm

We created VisualComm, a software application includes a PC, two display (one is for the deaf person and the other is for the hearing one), and a Microsoft Kinect. VisualComm is mainly designed for modeling the remote communication between deaf and non-disabled persons. On one hand, it can translate deaf user's sign action input to text or speech. On the other hand, a 3D avatar of our system is able to convey the hearing user's text or speech information through sign language animation.

The system configuration is shown in Figure 1. In the left side of the screen is sign language translation user interface, which has a live video, real time result bar and some control buttons. The recognition results of deaf user's sign language input will be displayed on the screen in real time. System also allows the signer to adjust the results by manual interaction. In the right side of the screen is specifically for hearing users to convey information to deaf users. When a hearing user types a sentence to the system, the 3D avatar will automatically played accordingly. All communication information will be recorded and appeared in the right-bottom textbox.

3.2 Implementation

This system mainly contains two core technologies, which are sign language recognition and animation respectively and will be introduced in detail in the following sections.

3.2.1 Sign Language recognition

Different with only using the geometric feature in literature[7], hand posture and trajectory are fused into our recognition framework for their powerful representation ability of sign word.

To recognize the sign from 3D trajectory, first the probe curve is preprocessed and normalized into equally distributed vector. And then the matching score is computed with the normalized gallery trajectories[2].

For posture representation, we extract the typical posture frames from the input sign video. The typical posture frame comes from the corresponding key posture fragment which has the characteristic of relatively low and stable motion speed. Once gotten the typical posture sequence description for each sign, the sign matching is transformed into the image set matching problem. We can get the matching score by computing the similarities of two sign words image sets.

Finally the similarity is obtained by a weighted combination of the matching scores from the 3D trajectory and hand posture. In addition to the recognition framework, a language model is integrated in our VisualComm system to generate all the words into sentences.

3.2.2 Sign Language Animation

Our avatar is a 3D model that consists of 53 body segments and 53 joints, which is driven by pre-recorded SL sentence data. In addition to hand and finger movement, the avatar can also perform some facial expressions to convey information.

4. EXPERIMENT AND STUDY

In this section, we conduct on the quantitative experiments and qualitative questionnaire survey.

We collect 370 daily Chinese sign words for 5 times. In 5-folds cross-validation test, we achieve Top-1 and Top-5 recognition rate of 94.2% and 98.9% on this dataset.

Five individuals are recruited to participant in our questionnaire study, including one deaf and four non-disabled persons. The non-disabled are trained to learn some sign words. All participants play the sign words randomly and our system gives the recognition results. Through the experience of the system, all participants think that "this system is **very helpful** for the deaf to communicate with normal persons" and "the accuracy and speed of the sign language recognition and translation system is "**satisfactory**". All deaf and non-disabled participants think that "under the help of the SLR system, the hearing disabled will be **more likely to** take part in the public activity". In the free comments feedback, one subject wrote that "**It's very impressive. It can remove the barrier between the common people and the deaf community. Hopefully, it can be used for a real scenario such as hospital, bookstore, etc**". Also, one subject suggests that more vocabularies should be recognized in the future in order to help the deaf interact with external environment freely.

5. CONCLUSION AND DISCUSSION

We created and implemented a tool, named VisualComm, to support deaf person to communicate with hearing person by using the Kinect. 3D depth data and typical posture sequence are integrated to obtain the stable and accurate recognition results. Experimental results and questionnaire survey convincingly show that the VisualComm can really help the deaf to interact with the surroundings efficiently and make them more willingness to participant in public activities. In the future, extended the sign language recognition to a larger vocabulary set (more than 1000 words) is our working focus.

6. ACKNOWLEDGMENTS

This work was partially supported by the Microsoft Research Asia, the FiDiPro Program of Tekes, the Natural Science Foundation of China under contract No. 61001193 and No.61303170.

7. REFERENCES

[1] Cavender, A., Ladner, R.E., Riskin, E.A. 2006. MobileASL: Intelligibility of Sign Language Video as Constrained by Mobile Phone Technology. *ASSETS*, (2006), 71-78.

[2] Chai, X., Li, G., Lin, Y., Xu, Z., Tang, Y., Chen, X., Zhou, M. 2013. Sign Language Recognition and Translation with Kinect. *IEEE Conf. on AFGR*, (2013).

[3] Han, J., Awad, G., and Sutherland, A. 2009. Modelling and Segmenting Subunits for Sign Language Recognition Based on Hand Motion Analysis. *Pattern Recognition Letters*, 30 (2009), 623-633.

[4] Hollan, J.D., Piper, A.M. 2008. Supporting Medical Conversations between Deaf and Hearing Individuals with Tabletop Displays. *CSCW*, (2008), 147-156.

[5] Shotton,J., Fitzgibbon, A., and Cook, M., etc. 2011. Real-time Human Pose Recognition in Parts from Single Depth images. *IEEE Conf. on CVPR*, (2011), 1297-1304.

[6] Takahashi, T. and Kishino, F. 1991. Hand Gesture Coding based on Experiments using a Hand Gesture Interface Device. *SIGCHI Bulletin*, 23, (1991), 67-73.

[7] Zafrulla, Z., Brashear, H., Starner, T., Hamilton, H. and Presti, P. 2011. American Sign Language Recognition with the Kinect, *ICMI*, (2011), 279-286.