IRCITCS-266
A Real-time Sign Language Recognition System for Hearing and Speaking Challengers

Chieh-Fu Hsieh*
Department of Electrical Engineering, Tamkang University,
New Taipei, Taiwan, R.O.C.
E-mail: *Hsiehchiehfu@gmail.com

Li-Ming Chen
Department of Electrical Engineering, Tamkang University,
New Taipei, Taiwan, R.O.C.

Ku-Chen Huang
Office of Physical Education, Tamkang University
New Taipei, Taiwan, R.O.C.

Ching-Tang Hsieh
Department of Electrical Engineering, Tamkang University,
New Taipei, Taiwan, R.O.C.

Chi-Hsiao Yih
Department of Electrical Engineering, Tamkang University,
New Taipei, Taiwan, R.O.C.

Abstract

Sign language is the primary means of communication between deaf people and hearing/ speaking challengers. There are many varieties of sign language in different challenger community, just like an ethnic community within society. Unfortunately, few people have knowledge of sign language in our daily life. In general, interpreters can help us to communicate with these challengers, but they only can be found in Government Agencies, Hospital, and etc. Moreover, it is expensive to employ interpreter on personal behalf and inconvenient when privacy is required. It is very important to develop a robust Human Machine Interface (HMI) system that can support challengers to enter our society. A novel sign language recognition system is proposed. This system is composed of three parts. First, initial coordinate locations of hands are obtained by using joint skeleton information of Kinect. Next, we extract features from joints of hands that have depth information and translate handshapes. Then we train Hidden Markov Model-based Threshold Model by three feature sets. Finally, we use Hidden Markov Model-based Threshold Model to segment and recognize sign language. Experimental results show, average recognition rate for signer-dependent and signer-independent are 95% and 92%, respectively. We also find that feature sets including handshape can achieve better recognition result.

Keywords: sign language, Kinect, Human Machine Interface (HMI), hidden Markov model (HMM).

1. Introduction

Sign Languages are a gesture that has integral construction and diversity (Kelly, 2009). They were used to communicate with others by hearing and speaking challengers. People communicate with challengers are difficult when he/she is not a signer or no one knows sign language around. In general, only can be found in government agencies, hospitals or services have sign language translating server. However, it also has problem when challengers want to do something for their privacy. And utilizing writing is indirectly when challengers' writing is no good. Therefore, how to aid challengers to communicate with others that are un-signer is the key of human machine interface developing (Nam, 1996).

The human machine interface design for sign language recognition system has two ways. One is vision-based system (Mukarami, 1991; Nam, 1996; Du, 2010; Mohandes, 2001, 2012; Lang, 2011, 2012), the other is glove-based system (Li, 2012; Placidi, 2013; Kong, 2014). In glove-based system, it should equip sensor glove to generate a set of signal features for signs. The advantage of this way is extracting features easier than other algorithms (Mohandes, 2001; Du 2010). Nevertheless, most of equipment is expensive and inconvenient. In vision-based system, users recorded gesture videos by any camera. Then they extracted features after hand is located to recognize from videos. However, the features are extracted in vision-based system has to be a clean environment. In recent year, most proposals for gesture and sign language recognition in vision-based system have proposed. Nam and Wohn (Nam, 1996) designed HMM-based recognition system that considers 300 and 200 data for training and testing respectively to recognize hands movement. Murakami and Taguchi (Mukarami, 1991) utilize neural networks to develop recognition system that employs 10 curve and 3 angle descriptors analyzing. Mohandes et al. (2012) develops a HMM-based Arabic sign language recognition system that has skin detection and color grove. Lang et al. (2012) and Lang (2011) employ Kinect sensor to extract distance, position and movement as features and recognize by HMM model.

In computer sign language recognizing, the major features are defined as handshape, hand orientation, position and trajectory respectively (William, 2005). Handshape and hand orientation are gesture features. They have different mean when different combination. For example, in Taiwan sign language (TSL) (N.C.C.U., http://tsl.ccu.edu.tw/web/browser.htm), as one hand point to the other hand that handshpae is thumb or pinkie means “he” or “she” separately. Position is space feature. As “ache” sign at one body part, it means “the part is ache.” Finally, trajectory is time-space feature. It can describe the space variable in continuous time. Communicating in sign language is composed multi-sign. However, the phenomenon that calls movement epenthesis (ME) (Nam, 1996; Kong, 2014) is happened between one sign to next sign. These MEs may reduce recognize accurately in continuous sign language recognition. Therefore, ME problem is considered when continuous sign language analysis.

In this paper, we employ vision-based system to develop TSL recognition system that combines TSL database (N.C.C.U., http://tsl.ccu.edu.tw/web/browser.htm) and Kinect sensor (Microsoft Developer Network, http://msdn.microsoft.com/zh-tw/hh367958.aspx).

2. Proposed Method

In this paper, we propose a TSL recognition system to aid challengers that communize without any translators. Our system has feature extract, model training and recognition three steps. The training flow chart shows as Fig.1.

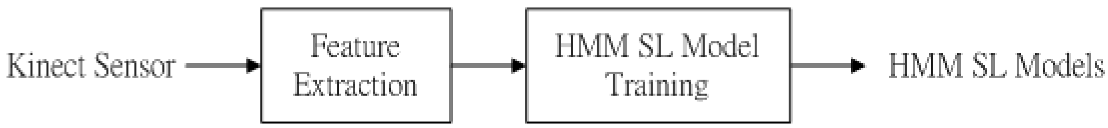

Fig.1: The flow chart of model training

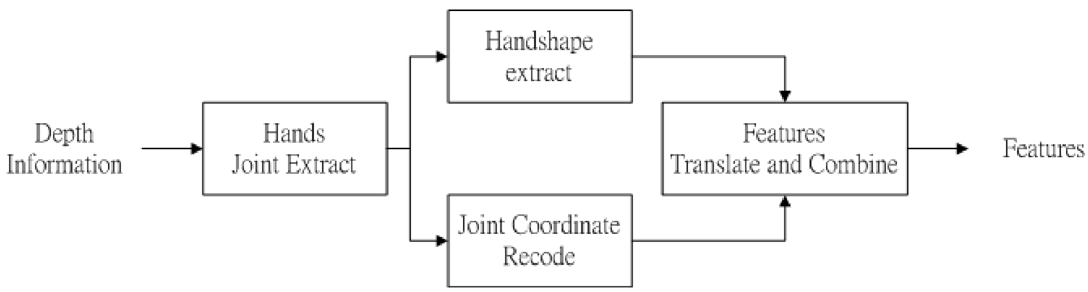

Fig.2: The detail flow chart of feature extraction

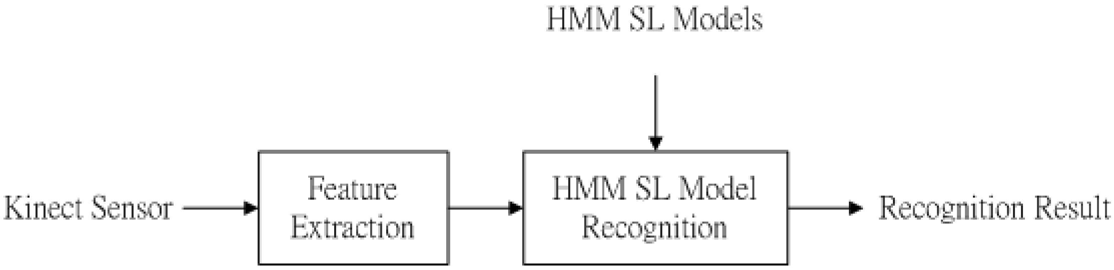

Fig.3: The flow chart of recognition

In our system, first step is feature extract. In this step, we input depth information by Kinect as Fig.2 show. Next, we employ Kinect to locate signer, then record hands 3D coordinate information to calculate features by Differential and angles between torso and hands. At the same time, we utilize (1) to quantize handshapes information to 8-bits image.

$$o(x,y) = \begin{cases} d(x,y) - d(x_H,y_H) + 127 & , \quad if\, d(x_H,y_H) - 127 \le d(x_H,y_H) \le d(x_H,y_H) + 128 \\ 0 & , \quad otherwise \end{cases} \quad (1)$$

Where $o(x,y)$ is output image, $d(x,y)$ is original depth information, (x,y) is coordinate in original depth information and (x_H,y_H) is coordinate extracted by hand joints of Kinect. After quantizing, we employ Bresenham's circle algorithm (Kao, 2011) make a circle filter and convolution with $o(x,y)$. Fig.4 is processing result.

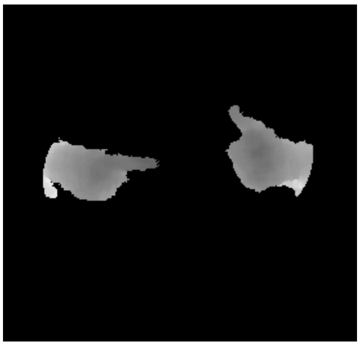

Fig.4: Extract handshapes

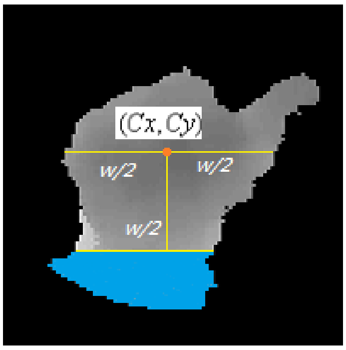

Fig.5: Wrist erase

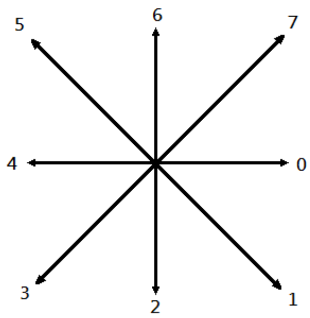

Fig.6: Direction rule

When handshapes of sign is extracted, we erase wrist part to reduce handshape recognizing error (Hsieh, 2012). First of all, we normalize angle of hand by calculating between vectors that from elbow to hand joint and y-axis. Next, the centroid (C_x, C_y) as Fig.5 shows is calculated by (2) and (3), respectively. Where T_c is total non-zero pixel, x_i and y_i is coordinate of non-zero pixels. Then, the length w that crosses centroid is found by counting.

Final, as (4) erase wrist.

$$C_x = \frac{\sum_{i=1}^{T_c} x_i}{T_c} \tag{2}$$

$$C_y = \frac{\sum_{i=1}^{T_c} y_i}{T_c} \tag{3}$$

$$o(x,y) = \begin{cases} o(x,y) & , \quad if\ y \le C_y + \frac{w}{2} \\ 0 & , \quad otherwise \end{cases} \tag{4}$$

After the wrist is erased, our method adopts boundary extraction algorithm (BEA) (Liu, 2003) to adapt Kinect information. Firstly, we scan $o(x,y)$ and find non-zero value to be starting point. Then, the coordinate is recoded by recording array. In traditional BEA, recording array has two arrays that record x and y axis of image. In our method, recording array has three arrays that record x, y and depth information of image. When starting point is found, we define search direction as Fig.6. The initial direction is set as 0, and search next non-zero pixel clockwise at 3×3 neighboring pixel. If the next non-zero is found, the algorithm changes search point to non-zero point and turns back 2 unit directions to search next non-zero pixel clockwise again. When next non-zero and starting point are the same, the algorithm would stop.

Discrete Fourier transform (DFT) descriptor (Cosgriff, 1960) is good for analyzing a close contour. Our method employs Fourier series that has robustness in size and movement to represent handshape features. Then we utilize SVM (Cristianini, 2004) to classify handshape. Finally, we employ code to correspond with these handshapes.

Final step, we employ hidden Markov model (HMM) (Rabiner, 1989) to train the sign language model. In recognition, as Fig.3, we extract features as training, then our method employs training model to recognize.

3. Experimental Result

In this section, we present experimental results for sign language recognition. The variation of the joints movement and angles between hand and torso joints are 6-D and 4-D feature vectors, respectively. The TSL signs and continuous sign language recognition experiments was obtained from five signers. The ME problem in continuous sign language recognition is manual deleted. In experimental result, we reduce some feature vectors to compare their importance.

Table I to IV are 50 signs, and every signer has 10 observations. The 50 signs include 30 signs without handshape data sets and 20 signs which 15 sign with handshape data sets and 5 continuous sign language data sets. Table I and II are signer-dependent test that has 8 observations are training and 2 observations are testing. Table III and IV are signer-independent that has 4 signers are training and 1 signer is testing.

Table 1: Signer-dependent test

Feature Set	6-D	10-D	10-D+Handshape
Accuracy	92.4%	94.2%	95.3%

Table 2: Signer-dependent test including handshape test data sets

Feature Set	6-D	10-D	10-D+Handshape
Accuracy	87.9%	89.6%	94.5%

Table 3: Signer-independent test

Feature Set	6-D	10-D	10-D+Handshape
Accuracy	90.8%	91.7%	92.8%

Table 4: Signer-independent test including handshape test data sets

Feature Set	6-D	10-D	10-D+Handshape
Accuracy	85.2%	86.3%	92.1%

We can see the accuracy without handshape test data sets as Table I can reach over than 92%. And as Table III, the accuracy is also reached 90%. In Table II and IV, the feature vectors add handshape are important in signs including handshape test data sets. The accuracy increases close to 6% when recognizing feature has handshape data sets.

4. Conclusion

We devised a TSL recognition vision-based system that employs Kinect sensor for communication between deaf people and hearing/speaking challengers. For handshape extracted, we present the algorithm to obtain it by Kinect depth image for sign recognition easily. In handshape feature translate, we modify BEA method to adapt Kinect to obtain depth feature. In recognition, our system obtained 95% and 92% accuracy in signer dependent and independent, respectively, and made vision-based sign language not only skin color but depth can obtain handshape information without equipment. We also find that feature sets including handshape can achieve better recognition result.

5. Acknowledgement

This paper is supported by National Science Council of Taiwan, and the Project Number is 102-2410-H-032-060.

6. Reference

R. L. Cosgriff, Identification of shape, *Technical Report 802-11*, Ohio State University, Columbus, OH, Dec. 1960.

N. Cristianini and J. S. Taylor, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, *Cambridge University Press, 2004*, pp. 82-108.

W. Du, J. Piater Hand modeling and tracking for video-based sign language recognition by robust principal component analysis. *Processing In: Int. workshop on sign gesture activity, Greece*; 2010.

C.T. Hsieh, C.H. Yeh, K.M. Hung, T.W. Chen and C.Y. Ke, A Real Time Hand Gesture Recognition System Based on DFT and SVM, *IEEE International Conference on Information Science and Digital Content Technology, 2012,* pp.490-495.

C.Y. Kao, C.S. Fahn. A Human-Machine Interaction Technique: Hand Gesture Recognition Based on Hidden Markov Models with Trajectory of Hand Motion. *Procedia Engineering, Volume 15, 2011*, pp. 3739-3743.

D. Kelly, J. McDonald, C. Markham. Recognizing Spatiotemporal Gestures and Movement Epenthesis in Sign Language. *Machine Vision and Image Processing Conference, IMVIP '09. 13th International. 2009*. pp. 145-150.

W.W. Kong, S. Ranganath. Towards subject independent continuous sign language recognition: A segment and merge approach. *Pattern Recognition 47. 2014.* pp. 1294–1308.

S. Lang, Sign Language Recognition with Kinect. *Ph.D. Disseration*, Bachelor, Institut fur Informatik, Freie Universitat Berlin, 2011.

S. Lang, M. Block. and R. Rojas, Sign Language Recognition Using Kinect. *Lecture Notes in Computer Science vol. 7267, 2012*, pp 394-402.

Y. Li, X. Chen, X. Zhang, K. Wang, J. Z. Wang, “A Sign-Component-Based Framework for Chinese Sign Language Recognition Using Accelerometer and sEMG Data.” *IEEE Transactions on Biomedical Engineering, Vol. 59, No. 10, October 2012*. pp. 2695-2704.

Y. H. Liu, Feature Analysis and Classifier Design and Their Applications to Pattern Recognition and Data Mining, *Ph.D. Disseration, National Taiwan Univ., R.O.C., 2003*.

Microsoft Developer Network. (n.d.). *Kinect for Windows Developer*. Retrieved from http://msdn.microsoft.com/zh-tw/hh367958.aspx.

M. Mohandes. Arabic sign language recognition. *Processing In: International conference of imaging science, systems, and technology, Las Vegas, Nevada, 2001*. pp. 753-759.

M. Mohandes, M. Deriche, U. Johar and S. Ilyas. A signer-independent Arabic Sign Language recognition system using face detection, geometric features, and a Hidden Markov Model.

Computers and Electrical Engineering vol. 38, 2012, pp. 422-433.

K. Mukarami, H. Taguchi. Gesture recognition using recurrent networks. *Processing In: Proceedings of the ACM CHI 91 human factors in computing systems. New Orleans, USA: ACM; 1991*, pp. 237–242.

Y. Nam, K. Whon. Recognition of space–time hand-gestures using Hidden Markov Model. *Processing In: ACM symposium on virtual reality software and technology Hong Kong; 1996*. pp. 51-58.

National Chung Cheng University (N.C.C.U.). (n.d.). *TSL Research group*. Retrieved from http://tsl.ccu.edu.tw/web/browser.htm.

G. Placidi, D. Avola, D. Iacoviello, L. Cinque. Overall design and implementation of the virtual glove. *Computers in Biology and Medicine 43. 2013*. pp. 1927–1940.

L. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE, vol. 77, no 2, 1989*, pp. 257–286.

C. William, Jr. Stokoe. Sign Language Structure: An Outline of the Visual Communication System of the American Deaf. *Journal of Deaf Studies and Deaf Education. Vol.10. Issue.1. 2005*. pp.3-37.