Taiwan Sign Language Recognition System Using LC-KSVD Sparse Coding Method

Ching-Tang Hsieh[1, a], Hsing-Che Liou[1, b], Li-Ming Chen[1, c]

[1]Dept. of Electrical and Computer Engineering, Tamkang University, New Taipei, 25137, Taiwan

[a]email: hsieh@ee.tku.edu.tw, [b]email: 603440073@s03.tku.edu.tw,

[c]email: ms0071799@hotmail.com

Keywords: Sign Language; Recognition System; Sparse Coding; LC-KSVD

Abstract. Sign language, for deaf-impaired people, plays an important role in communication. In this paper, we devise a Taiwan Sign Language recognition system. We use the Kinect2 sensor to get data from 94 sign morphemes shown once by 4 people, and extract hand shape features and trajectory features from depth images and joints of the body skeleton. Finally, we have each sign morpheme dictionary trained by label consistent K-SVD (LC-KSVD) sparse coding algorithm for recognition. Experiments show our system performs well and the accuracy achieves 99.47% in close test.

Introduction

Sign language, a kind of languages not via the oral vocal system, combines hand gestures with body movements and even more facial expressions to show its meaning. For many deaf-impaired people, sign language is mother language [1]. Representing manner is not only one type but also many kinds for the same meaning of sign language, such as the character of Chinese and English among spoken languages. Hundreds of sign languages are in use around the world and also are the cores of local deaf cultures, like Taiwan Sign Language.

According to previous sign language recognition research [2], sign language recognition is identified to four blocks of sign language: hand shape, hand position, trajectory orientation and movement. With four blocks in mind, hand gestures can be classified as hand postures and spatiotemporal gestures [3]; hand postures are combinations of hand shapes and orientations, and spatiotemporal gestures refers to where the hand is placed relative to the body and hand movement traces out a trajectory in space. Taiwan Sign Language (TSL) is exactly a kind of sign language, and the key points of lexical meaning resolution, e.g., hand shapes, positions, actions, directions, not hand motions, contact point, relationship of hands, repeat movements, and posture intensity, is observed from sign morphemes, rules of structure and restriction [4]. Sign language recognition system can be constructed by the foregoing factors. However, sign language recognition system for extracted methods of features can be distinguished between image-based and sensor-based approaches [5]. The advantage of image-based systems for signers uses easy equipment, but needs a lot of computing of pre-processing. This paper belongs in the image-based sign language recognition system.

A classical image-based system has five processing step: image obtained, pre-processing, segmentation, features extracted and classify [6], and so does the sign language recognition system roughly in this paper.

In the previous sign language works, S. Mo et al. [8] get hand shapes with color skin; but in their experiment result, selecting color skin for usage cannot catch hand shapes possibly, if there are similar color objects. In [9], A.-M. Cretu et al. use depth information and the threshold of distance to capture hand shapes; nevertheless, it is hard to catch hand shapes when the hand is near body. In [10], X. Wu et al. use depth joints features to draw hand location and trajectory; although, trajectory features do not used with hand shapes for gestures recognition. The above-mentioned problems are considered as the opinions of features choice. Depth images and joints are the target used to extract features in this paper, and are illustrated their characteristic in the chapter 3.

In this paper, the Histogram of Oriented Gradients (HOG) [12] algorithm referred from [11] is used to extract the hand features. In the HOG algorithm, the local oriented gradients are computed from depth image of the hands, and do normalization respectively. The normalized values are considered as hand shape features. However, the dimension of hand features computed by HOG is so large that it has to do dimension reduction via the Principal Component Analysis (PCA) algorithm to get the lower dimensional and useful hand features.

In spite of effectively reduced dimensions of hand feature via PCA, the data quantity of the sign language is still quite big and needs more time cost did classification directly. For the better effect, the method, the improved sparse coding: Label Consistent K-SVD (LC-KSVD) [14,15], is applied in the dimension reduction and the classification.

We introduce Kinect Sensor and LC-KSVD in section 2, proposed method is given in section 3. Last section is experiment and conclusion.

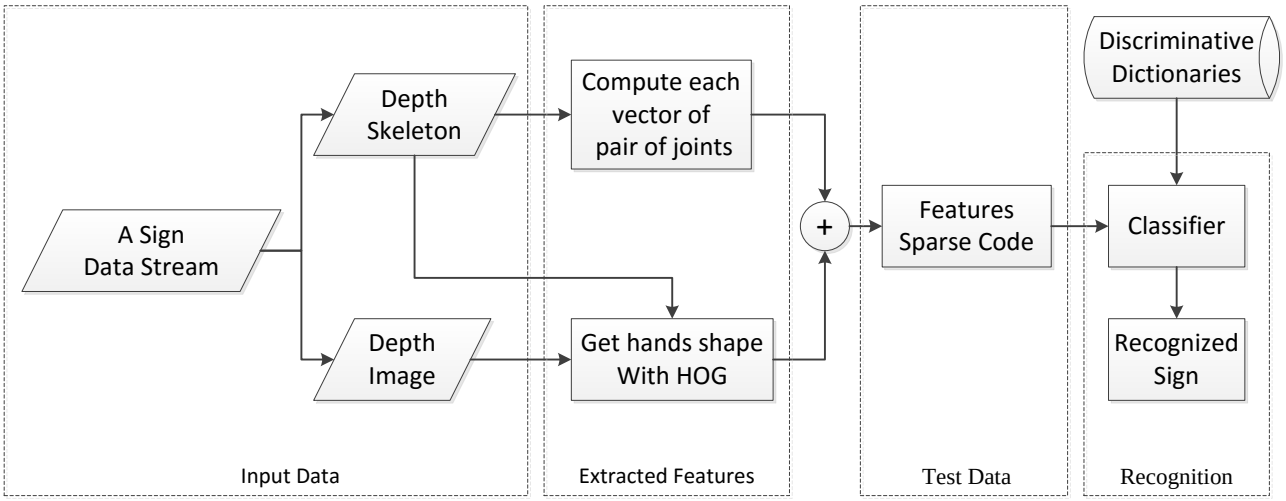

Figure 1. Flow chart of the sign language recognition system.

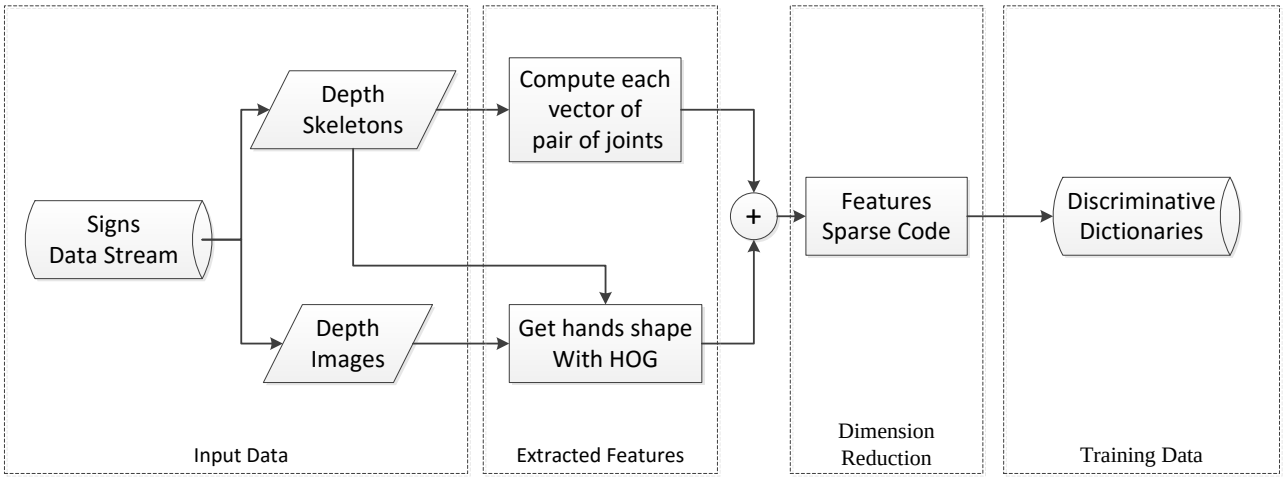

Figure 2. Flow chart of the discriminative dictionary training.

Kinect Sensor and LC-KSVD

Kinect Sensor. In the recent image-based sign language recognition researches, new human-computer interaction system is used to related works, except for the traditional image-based recognition technologies. In particular, Microsoft Kinect has attracted special attention, and has recently been used for action recognition with application in human-computer interaction [7]. With Kinect for Windows SDK 2.0 Microsoft provide, we can get the necessary features for sign language recognition, e.g., the continuous time color images, depth information, joints of the body skeleton. Therefore, we can build database with hand postures and spatiotemporal gestures

described sign language, after correspondingly dealing with original data.

LC-KSVD. Label Consistent K-SVD (LC-KSVD) is improved from the K-SVD sparse coding. Let input Y be a set of n-dimensional N input signals, i.e. $Y=[y_1...y_N]\in R^{n\times N}$, and the dimension of learning reconstructive dictionary be K items for sparse representation of Y, and Q be the 'discriminative' sparse codes of input signals Y for classification, i.e. $Q=[q_1...q_N]\in R^{K\times N}$. An objective function for dictionary construction is defined as:

$$\langle D,A,X\rangle=\arg\min_{D,A,X}\|Y-DX\|_2^2+\alpha\|Q-AX\|_2^2 \; s.t.\forall i,\|x_i\|_0\le T \quad (1)$$

where $D=[d_1...d_N]\in R^{n\times K}$ ($K>n$,making the dictionary over-complete) is the learned dictionary, $X=[x_1...x_N]\in R^{K\times N}$ are the sparse codes of input signals Y. A is a linear transformation matrix , and α controls the relative contribution between reconstruction and label consistent regularization. T is a sparsity constraint factor (each signal has fewer than T items in its decomposition). The term $\|Y-DX\|_2^2$ denotes the reconstruction error, and the term $\alpha\|Q-AX\|_2^2$ represents the discriminative sparse-code error. However, the description of LC-KSVD is not perfect in this paper, and detailed parts are referred to [14,15].

Proposed Method

In this chapter, the sign language recognition system we proposed is introduced as below, and the flow chart is shown in fig.1.

In the available data, the depth image and the joints of the body skeleton are the objects to be extracted features. The depth image is the image took the depth information as the values of distance, and it has a great characteristic that is not influenced by color difference and luminous flux to be as the hand features of sign language in this paper. The joints of the body skeleton are to be described as the situation of the body skeleton in the 3-D space, and with the relationship of time that the 3-D movement trajectory of the sign language can be obtained.

A. Training Data

To extract the hands shape features, the hands region in the depth image is caught with the joint coordinate, and the hands region segment is extracted the hands features via HOG and PCA. The trajectory feature of the sign language morpheme is computed by the normalized joints.

The feature matrix of the sign language morpheme is the chronological combination of the hands shape and trajectory features. The chronological combination is able to deal with the different sign speed of the different signer. And then, the feature matrix is trained by LC-KSVD algorithm to generate the training data that is the same and lower dimensional discriminative dictionary. The training process of discriminative dictionary is shown in fig.2.

B. Test Data

The one of the sign language morpheme is extracted via the same procedure of training data to generate the testing data. The accuracy of the testing data versus each discriminative dictionary is obtained after comparison, and the highest accuracy discriminative dictionary is considered as the classification of the input sign language morpheme.

Experimental Result

In this paper, Microsoft Kinect2 sensor is used to catch the depth image and the joint of the body skeleton from 94 sign morphemes shown once by 4 people. The depth image of every sign morphemes is the continuous time images, and the range of depth is 500 to 4500 pixels, and the image size is 512*424 pixels. The joint of the body skeleton can be got the 25 joints, and the joints spread on the head, neck, spine, shoulders, arms, hands are used in this paper. In the close test, the aforementioned data via the method we proposed can get the average accuracy 99.47% shown in

fig.3.

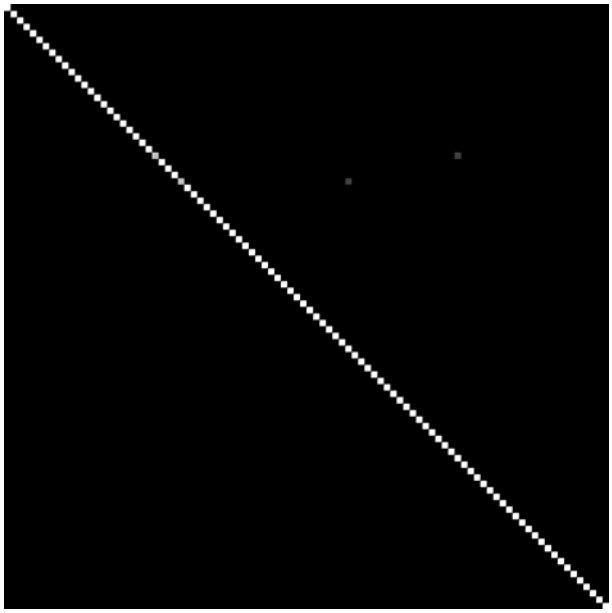

Figure 3. The confusion matrix of the experimental result. X-axis is the control group, and from top to bottom is sign1 to sign94; Y-axis is the experimental group, and from left to right is sign1 to sign94.

Conclusion and Future Work

In this paper, the depth image is used to replace the color image, and the characteristic of the depth image can solve the problems on the usage of the color image. In the close test, the great result is shown by using the pair features of hands shape and the trajectory. However, the part of fault by our examination is found that the variation of control group sign and the experimental group sign is very similar in the some short time and is emphasized by training process. In the future work, we plan to use weight to control the hands shape and trajectory features to improve the experimental result, and add the opinions of the review to the open test planning.

Acknowledgement

This work was supported by the National Science Council under grant number MOST 103-2632-E-032 -001-MY3, and the Tamkang University under grant number FDRX10-2321.

References

[1] Tsay, Jane, James H.-Y. Tai and Yijun Chen, *Taiwan Sign Language Online Dictionary* 3rd Edition, http://tsl.ccu.edu.tw/web/browser.htm, Institute of Linguistics, National Chung Cheng University, Taiwan, 2015.

[2] William C. Stokoe, Jr., Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf, Journal of Deaf Studies and Deaf Education, v10 n1 p3-37Win 2005, 2005.

[3] Y. Wu, T. S. Huang, and N. Mathews, Vision-based gesture recognition: A review, in Lecture Notes in Computer Science, pages 103–115, Springer, 1999.

[4] Kun-Ying Cai, Sin-Rong Ke, Literature Meta-analysis in A Recent Decade of Sign Language in Taiwan, Special Education for the Elementary School, vol. 46, pp. 23-24, 2008. (Written by Traditional Chinese.)

[5] M. Mohandes, M. Deriche, and J. Liu, Image-Based and Sensor-Based Approaches to Arabic Sign Language Recognition, Human-machine Systems, IEEE Transactions on, vol. 44, no. 4, Aug. 2014.

[6] R. Azad, B. Azad, and I. T. Kazerooni, Real-time and robust method for hand gesture recognition system based on cross-correlation coefficient, Adv. Comput. Sci., Int. J., vol. 2, no. 5/6, pp. 121–125, Nov. 2013.

[7] S. R. Fanello, I. Gori, G. Metta, Keep It Simple And Sparse: Real-Time Action Recognition, Journal of Machine Learning Research, vol.14 pp. 2617-2640, 2013.

[8] S. Mo, S. Cheng, X. Xing, Hand Gesture Segmentation Based on Improved Kalman Filter and TSL Skin Color Model, Multimedia Technology (ICMT), International Conference on, 2011.

[9] G. Plouffe and A.-M. Cretu, Static and Dynamic Hand Gesture Recognition in Depth Data Using Dynamic Time Warping, Instrumentation and Measurement, IEEE Transactions on, vol. 65, no. 2, Feb. 2016.

[10] X. Wu, X. Mao, L. Chen, Y. Xue, Trajectory-based view-invariant hand gesture recognition by fusing shape and orientation, IET Comput. Vis., 2015, Vol. 9, Iss. 6, pp. 797–805.

[11] C. Sun, T. Zhang, B. Bao, C. Xu and T. Mei, Discriminative exemplar coding for sign language recognition with Kinect, Cybernetics, IEEE Transactions on, vol. 43, no. 5, pp. 1418-1428, 2013.

[12] N. Dalal and B. Triggs, Histograms of oriented gradients for human detection, in Proc. CVPR, pp. 886–893, 2005.

[13] B.-K. Bao, G. Liu, C. Xu, and S. Yan, Inductive robust principal component analysis, IEEE Trans. Image Process., vol. 21, no. 8, pp. 3794–3800, Aug. 2012.

[14] Zhuolin Jiang, Zhe Lin, Larry S. Davis, Learning a Discriminative Dictionary for Sparse Coding via Label Consistent K-SVD, Computer Vision and Pattern Recognition, IEEE Conference on, 2011.

[15] Zhuolin Jiang, Zhe Lin, Larry S. Davis, Label Consistent K-SVD: Learning A Discriminative Dictionary for Recognition, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 35(11): 2651-2664, 2013.