Feature Extraction Methods in Sign Language Recognition System: A Literature Review

Conference Paper · September 2018

DOI: 10.1109/INAPR.2018.8626857

4 authors, including:

Suharjito Suharjito
Bina Nusantara University, Jakarta, Indonesia

I Gede Putra Kusuma Negara
Binus University

Feature Extraction Methods in Sign Language Recognition System: A Literature Review

Suharjito
Computer Science Department, BINUS Graduate Program – Master of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480
suharjito@binus.edu

Fanny Wiryana
Computer Science Department, School of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480
fanny.wiryana@binus.ac.id

Gede Putra Kusuma
Computer Science Department, BINUS Graduate Program – Master of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480
inegara@binus.edu

Amalia Zahra
Computer Science Department, BINUS Graduate Program – Master of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480
amalia.zahra@binus.edu

***Abstract*—Sign language is a way to communicate between deaf-mute people and normal people by performing hand gesture. Visual-based gesture recognition can help to overcome this communication limitation. Recently, several techniques and methods have been proposed in this area of research and showed some improvements. Despite all of the proposed methods, most of hand gesture recognition approaches that have been applied still lack of compatibility and have lots of limitations. For instance, hand segmentation meets the complication of distinguishing the hand and the face region by using skin detection. Motivated by those facts, this paper presents a review and explains progress of feature extraction in sign language recognition mostly in the last ten years. In this contribution, we focus on studying feature extraction methods. The literature used in this study is based on the previous published international papers which discussed sign language recognition. The main objectives from this review are to get the most effective and most compatible feature extraction method to be applied to sign language recognition system and to further research progress in the future. After reviewing various recognition techniques, we can conclude that the current works have successfully improve hand gesture recognition by inventing a technology which helps for tracking hands region precisely by using an active sensor. However, there is still room for improvements based on a markerless passive sensor, such as vision-based approaches.**

Keywords— gesture recognition, sign language recognition, feature extraction, pre-processing method

I. Introduction

Gesture is a physical movement of body [1]. It can be hands, head, or facial expression. Gesture recognition is about tracking and interpreting every movement done by our body. With the development of human computer interaction (HCI), one technology that has been created is hand gesture recognition. By using hand gesture recognition, people can give command even for translating the communication between deaf-mute and normal people. One of the applications of hand gesture recognition is sign language recognition. Recently, a perfect sign language recognition system has been considerably deliberated.

From the discussion in [2], there are several essential parts for sign language recognition. They are robustness, computational efficiency, scalability, and orientation. Orientation in sign language recognition is handled in pre-processing step. Hence, pre-processing method holds a crucial part in sign language area. It talks about how we process the raw data to be classified. Nowadays, pre-processing method in sign language field has obtained more interest from researchers. Lots of works that have been conducted focus on proposing a new method or algorithm to get a better result. Based on the survey in [3], most of the working papers up to present discuss pre-processing methods. This paper presents review for 2D and 3D processing methods and describes a brief comparison between one and other methods. Not only hand motion recognition, this paper also shows another working paper proposes a pre-processing method focusing on speed where it might operate in nearly real-time.

From all the proposed methods and experiments that have been conducted, a nearly perfect system for sign language recognition is still far away from reality. Due to the limitation of the condition handled by the system and the variety of hand shapes, preprocessing method keeps developing itself to find the most accurate method. This paper will discuss the pre-processing methods further and present the most recent technology improved for hand recognition. Section 2 explains a number of feature extraction methods. Further analysis of such methods is presented in Section 3. Finally, Section 4 concludes the paper and presents some future works.

II. Feature Extraction Method

One of the main problems in a recognition system is pre-processing issue. Sign language recognition may deal with physical movement (dynamic) of hands, face, fingers or the whole body. The hand gestures themselves are rich in variation of shapes, motions, and textures. The feature should be efficient and reliable to deal with the richness of such variations. Even by extracting features using geometric, features are not always available and reliable due to self-occlusion and room lighting [1]. The importance of a proper selection of feature extraction method is also mentioned in [4] and [2]. The robustness can be reached once the

978-1-5386-9422-0/18/$31.00 ©2018 IEEE

combination of a proper feature extraction method and the recognition algorithm has been found. The selected feature can affect the success or failure of an experiment in the field of human-computer interaction using hand gestures, since the input to a classifier is obtained from this step. The work in [5] explains the steps of feature extraction method in general that can be seen in Fig. 1.

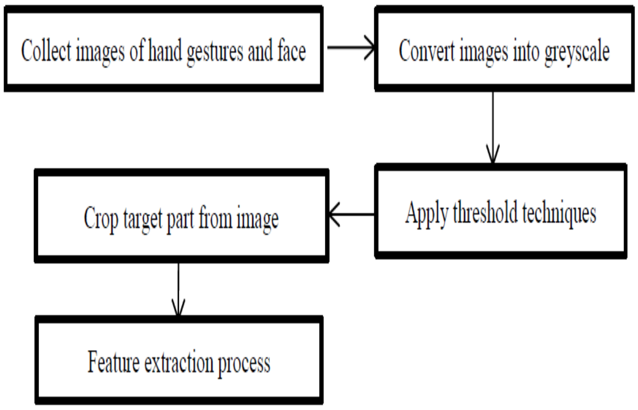

Fig. 1. Finding morphological features from image

A. Hand Segmentation

Segmentation is the first step to process the input image prior to recognition. The image that has been captured by a camera is segmented to obtain the hand area. One of the segmentation methods is using canny edge detector. Canny edge detector is used to detect the boundary from an image, in this case to detect the hand area [6], [7]. When using the Canny edge detector, a threshold plays a critical role in determining proper edges from an image. Elliptical Fourier descriptors can be an option to deal with edge detector like the work in [8]. Another method that can be used is skin detection. This method aims to identify the skin area in an image. However, the method cannot directly determine the hand area. It could solve the problem when the image provided includes the face area, which will also be detected. Thus, skin detection must be combined with hand motion tracking by calculating the boundary of hands and determine the centroid point of hands [9]. Another variation mentioned in [3] is to distinguish the head area from the face by assuming that the head area is relatively static and bigger than the face area. Similarly in [10], a face detection algorithm is applied to remove the face region. Table 1 summarizes a number of hand segmentation methods described in a number of papers along with their segmentation accuracies.

TABLE I. PERFORMANCE COMPARISON OF HAND SEGMENTATION METHODS

Author	Method	Accuracy
[6]	Canny Edge Detector	92.34%
[7]	Canny Edge Detector	100%
[8]	Elliptical Fourier Descriptors	95.1%
[9]	Skin Detection + Canny Edge Detector	-
[3]	Skin Detection	-
[10]	Skin Detection	87.33%

B. Glove-Based Hand Tracking

The experiments are intended to look for the best solution to simplify the tracking of the hand area, which leads to discovering the idea of using gloves. The use of gloves tends to make the algorithm easier to find the hands area. Gloves normally are made of common cloth [11]. As it keeps on developing, the glove-based method creates an evolution, transforming from the coloured glove-based into a sensory glove-based.

1) Coloured Glove-Based

The coloured glove-based method is an old method in sign language area for tracking hands. As shown in [12], they proposed a real-time hand recognition system using coloured gloves to extract hand position. The subject used a yellow glove on the right hand and an orange glove on the left one. The algorithm then finds the hand position by searching the appropriate pixel colour of the gloves. Once the algorithm finds the pixel, it extends the checking using eight nearest neighbours to obtain considered part of hand. Then, the feature extraction gives the result in multi-dimensional Gaussian densities by considering eight element features (each hand x, y position, angle of axis of least inertia, and bounding ellipse).

2) Sensory Glove-Based

Another way to put the complicated implementation of pre-processing method aside is by using sensory gloves. Sensory gloves are modified gloves with a sensor fitted on it [11]. The work in [11] used seven sensors. Five sensors are put in each finger, one for measuring the tilt of hand, and another one for sensing the rotation. The improvement in sensory gloves includes 10 flex sensors and an accelerometer as motion tracker [13]. Sensory glove is a weighty thing in sign recognition. It wraps all the pre-processing process in the accelerometer. The accelerometer returns the value of hands projection in feature vector as an input to the classification method. The work in [14] proposed sign language recognition by using sensory glove as an input command when using a website. Another work with sensory gloves is found in [15], which proposed Thailand sign language recognition in a single representation sign by using data gloves and motion tracker. Another improvement feature used in the paper is by adding the sensor to 14, which includes 10 sensors on fingers and 4 sensors between fingers. Table 2 summarizes some proposed works on glove-based hand tracking methods and their reported accuracies.

TABLE II. PERFORMANCE COMPARISON OF GLOVE-BASED HAND TRACKING METHODS

Author	Method	Accuracy
[12]	Coloured Gloves	95%
[11]	Sensory Gloves	88%
[13]	Sensory Gloves	90.3%
[14]	Sensory Gloves	-
[15]	Sensory Gloves	94.44%

A review conducted in [16] mentioned that glove-based approach has a disadvantage. The gloves need to be worn, and the idea of using gloves will hamper the naturalness of human computer interaction.

C. Principal Component Analysis (PCA)

PCA is one of the most favourite methods and likely to be the oldest method in scientific discipline. PCA involves mathematic procedures using matrix. PCA is selected for dimensionality reduction in matrix, thus only necessary information left in the matrix. The work in [17] describes the goals of PCA, which are:

- The most important information from the data table will be extracted;
- This important information will be compressed;
- The description of the data set will be simplified; and
- For analyzing the structure of the observations and the variables.

PCA works by computing new variables called principal components and using them to form linear combinations of original variables. One of the principal components is to produce the largest possible variance, i.e. inertia, therefore the largest part of inertia from the table will be extracted. The component is obtained from Singular Value Decomposition (SVD) of a matrix X. It is denoted as below:

$$\mathbf{X=P\Delta Q^T} \tag{1}$$

Where P is the I x L matrix of left singular vectors, Q is the J x L matrix of right singular values and Δ is a diagonal matrix of singular values. From the formula above, we can obtain another equation, which is the I x L matrix of factor scores, denoted as **F**, which is obtained from:

$$\mathbf{F=P\Delta} \tag{2}$$

The work in [18] combined PCA and glove-based data by using an Attitude Heading Reference System (AHRS) sensor. Unfortunately, AHRS sensor does not seem to give a great result. The acceleration received from the AHRS sensor is not clear and contains many noises, therefore low pass filter (LPF) is applied after comparing the result to that from Kalman filter. The paper combined PCA, *k*-means algorithm, and ABC-based Hidden Markov Model (HMM) to achieve a satisfactory recognition result. The result produced by the experiment reached the accuracy of 91.3%.

The work in [19] proposed a modification feature extraction using PCA combined with kurtosis and motion chain code (MCC). Kurtosis was used for fusing the PCA and MCC was used to represent the hand movement. The combination of these three methods achieved the least error of up to 10.91% error rate. Another combination in [6] proposed feature extraction by using PCA and Elliptical Fourier Descriptors. The Elliptical Fourier Descriptors was used to optimize and preserve the shape data, thus any kind of rotation would not change the Fourier Descriptors. This experiment yielded the accuracy of 92.34% in a recognition using Artificial Neural Network (ANN). PCA by using Otsu algorithm as image pre-processing for segmentation was implemented in [20]. Otsu algorithm produces result in the form of a feature vector which is later processed by PCA to have a compressed vector. In the paper, two main features were used for recognition: centroid of the hand (between fingers and non-fingers) and skin detection. Table 3 summarizes some proposed works on PCA-based methods and their reported accuracies.

D. Scale Invariant Feature Transform (SIFT) Algorithm

The work in [22] proposed sign language recognition by using SIFT algorithm as the feature extraction method. Another similar work is found in [23]. SIFT algorithm is an interesting method because it is robust against rotation, translation or scaling variation and it produces “large collection of local feature vectors” [24].

TABLE III. PERFORMANCE COMPARISON OF PCA-BASED METHODS

Author	Method	Accuracy
[18]	PCA + k-Means Algorithm	91.3%
[19]	PCA + Kurtosis + Motion Chain Code	89.09%
[6]	PCA + Elliptical Fourier Descriptors	92.34%
[20]	PCA + Otsu Algorithm	-

E. Serial Particle Filter

Serial particle filter has been widely used in many object recognition. The work in [10] mentioned that particle filter is well-suited for tracking single, two, or more objects. In the sign language field, particle filter is used for tracking both right and left hands. The paper combined particle filter with feature covariance matrix for hand representation and recognition. This proposed method gives the result of 87.33% accuracy and shows some improvement in noise reduction and more robust foreground detection. For future works, this method can be improved by exploring a method to reduce computational time in order to create a nearly real-time system.

F. Microsoft Kinect

Microsoft Kinect is a device designed to recognize any gesture including hand gesture. It processes every motion that has been captured into features. Microsoft Kinect is designed with 3D sensor camera which is well-suited for sign language recognition. The camera setup offers some efficiency to extract not only hands, but also head and elbows, which is a great support to distinguish the hand region and similar signs. Another advantage of using Microsoft Kinect is the independency of lighting condition. Camera will still detect equally well even in the places with less light [16].

In sign language recognition area, Kinect has been used widely to simplify the pre-processing part. The work in [25] used Kinect SDK combined with Histograms of Oriented Gradients (HOG) to portray hand figures. Kinect SDK was used to obtain the depth and coloured image. The features were then extracted by using hand trajectory and shape information that were normalized by HOG. This approach produced the accuracy of 96.75% by combining those features.

Another work using Kinect was found in [26]. The feature extraction method used in that research was spatial domain method to recognize the fingertips. The classification result stated in the paper was 97.5% accuracy by using a Support Vector Machine (SVM) classifier. A high recognition accuracy of 96.67% was also achieved by [27]. The features used in the approach were the combination of skeleton and depth image features. The extracted features included area, centroids, major and minor axis, orientation, and normalized convex hull. The combination of local and global features in [28] yields the accuracy of 99.84%. The dynamic hand gestures were handled using RGB data, depth data, and skeleton data received from Kinect. Different approach was proposed in [29], which focused on lighting normalization in Kinect sensor. It produced the accuracy of 89%. The highest accuracy was achieved in [30] with 99.9% using SVM as the classification algorithm. Table 4 summarizes some proposed methods using Kinect SDK and their reported accuracies.

TABLE IV. PERFORMANCE COMPARISON OF METHODS USING KINECT SDK

Author	Method	Accuracy
[25]	Kinect + HOG	96.75%
[26]	Kinect + Spatial Domain	97.5%
[27]	Kinect	96.67%
[28]	Kinect + Local + Global Features	99.84%
[29]	Kinect	89%
[30]	Kinect	99.9%

G. Leap Motion Controller (LMC)

The latest update on pre-processing method and experiment using active sensor has been conducted in [31]. They proposed a feature extraction method based on the data obtained by using Leap Motion Controller (LMC). LMC is a device that can detect hand movement with 200 frames per second and gives an ID whenever it detects hand movements. It converts hand movement signals into computer commands [32], [33]. LMC returns 23 features, among which only 12 features are selected in the paper. Another work in [21] proposed data transmission for 15 frames per second. LMC has been explored by [34], which mentioned that LMC someway is better than Microsoft Kinect. The LMC API can directly map the detected data to hand fingertip and movement. However, LMC is not perfect and still developing. It has some difficulties in implementing the API when hands are flipped over.

III. ANALYSIS AND DISCUSSION

After summarizing the pre-processing methods (feature extraction methods) and their experimental results in sign language recognition, we can see that a lot of progress has been made to obtain the best pre-processing method. Starting from hand segmentation, there is one experiment that produces the accuracy of 100%. Despite this result, there are a lot of limitations in the experiment, such as it can only recognize a simple single sign.

In glove-based method, there is a significant progress made with the transformation from coloured glove-based to sensory glove-based, which uses an accelerometer sensor, not an ordinary camera. PCA method gives the best result when being combined with Elliptical Fourier. Unfortunately, for SIFT algorithm, there is no paper that mentioned the accuracy explicitly. The paper only stated that by using SIFT algorithm, the features from the image are easily extracted. However, we cannot define how easy it is since there is no comparison with other methods.

For the most recent experiments, there is serial particle filter method. Actually this method has been widely used in the recognition field, but there is still lacks of attention put on gesture recognition, specifically in sign language recognition field. For the technology progress, there are Microsoft Kinect and Leap Motion Controller (LMC). Both are sensory systems that already have several feature recognitions. Microsoft Kinect has depth calculation, RGB data, and skeleton data, while LMC can detect hand movement in a high frame rate.

ACKNOWLEDGMENT

This research was supported by research grant No. 039A/VR.RTT/VI/2017 from Ministry of Research Technology and Higher Education of Republic of Indonesia.

REFERENCES

[1] G. R. S. Murthy and R. S. Jadon, "A review of vision based hand gestures recognition," International Journal of Information and Knowledge Management, 2(2), pp. 405–410, 2009.

[2] N. K. Gilorkar and M. M. Ingle, "A review on feature extraction for Indian and American sign language," International Journal of Computer Science and Information Technologies, 5(1), pp. 314–318, 2014.

[3] S. C. W. Ong and S. Ranganath, "Automatic sign language analysis: A survey and the future beyond lexical meaning," IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(6), pp. 873–891, 2005.

[4] J. P. Wachs, H. Stern, and Y. Edan, “Cluster labeling and parameter estimation for the automated setup of a hand-gesture recognition system,” IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 35(6), pp. 932–944, 2005.

[5] H. Bhavsar, "Review on feature extraction methods of image based sign language recognition system," Indian Journal of Computer Science and Engineering, 8(3), pp. 249–259, 2017.

[6] M. V. D. Prasad, V. Kishore, and A. Kumar, "Indian sign language recognition system using new fusion based edge operator," Journal of Theoretical and Applied Information Technology, 88(3), pp.574–584, August 2016.

[7] E. A. Kalsh and N. S. Garewal, "Sign language recognition system," International Journal of Computational Engineering Research, 3(6), pp.15–21, 2013.

[8] P. V. V. Kishore, M. V. D. Prasad, C. R. Prasad, and R. Rahul, "4-camera model for sign language recognition using elliptical Fourier descriptors and ANN," International Conference on Signal Processing and Communication Engineering Systems (SPACES), IEEE, pp. 34–38, 2015.

[9] P. C. Pankajakshan and B. Thilagavathi, "Sign language recognition system," Innovations in Information, Embedded and Communication Systems (ICIIECS), IEEE, pp. 1–4, 2015.

[10] K. M. Lim, A. W. C. Tan, S. C. Tan, "A feature covariance matrix with serial particle filter for isolated sign language recognition," Expert Systems with Applications, Elsevier Ltd, 54, pp. 208–218, 2016.

[11] S. A. Mehdi and Y. N. Khan, "Sign language recognition using sensor gloves," Proceedings of the 9th International Conference on Neural Information Processing (ICONIP), IEEE, vol.5, pp. 2204–2206, 2002.

[12] T. E. Starner, "Visual recognition of American sign language using hidden Markov models," Massachusetts Inst Of Tech Cambridge Dept Of Brain And Cognitive Science, 1995.

[13] L. T. Phi, H. D. Nguyen, T. Q. Bui, T. T. Vu, "A glove-based gesture recognition system for Vietnamese sign language," Proceedings of the 15th International Conference on Control, Automation and Systems (ICCAS), pp. 1555–1559, 2015.

[14] A. Ranjini S. S. and M. Chaitra, "Sign language recognition system," International Journal on Recent and Innovation Trends in Computing and Communication, 2(4), pp. 947–953, 2014.

[15] S. Saengsri, V. Niennattrakul, and C. A. Ratanamahatana, "TFRS: Thai finger-spelling sign language recognition system," 2nd International Conference on Digital Information and Communication Technology and its Applications (DICTAP), pp. 457–462, 2012.

[16] M. A. Majid and J. M. Zain, "A review on the development of indonesian sign language recognition system," Journal of Computer Science, 9(11), pp. 1496–1505, 2013.

[17] H. Abdi and L. J. Williams, "Principal component analysis," Wiley Interdisciplinary Reviews: Computational Statistics, 2(4), pp. 433–459, 2010.

[18] T. H. S. Li, M. C. Kao, and P. H. Kuo, "Recognition system for home-service-related sign language using entropy-based k-means algorithm and ABC-based HMM," IEEE Transactions on Systems, Man, and Cybernetics: Systems, 46(1), pp. 150–162, 2016.

[19] M. M. Zaki and S. I. Shaheen, "Sign language recognition using a combination of new vision based features," Pattern Recognition Letters, Elsevier B.V., 32(4), pp. 572–577, 2011.

[20] S. N. Sawant, "Sign language recognition system to aid deaf-dumb people using PCA," International Journal of Computer Science and Engineering Technology, 5(5), pp. 570–574, 2014.

[21] M. Koul, P. Patil, V. Nandurkar, and S. Patil, "Sign language recognition using leap motion sensor," International Research Journal of Engineering and Technology (IRJET), 3(11), pp. 322–325, 2016.

[22] A. Thorat, V. Satpute, A. Nehe, T. Atre, and Y. R. Ngargoje, "Indian sign language recognition system for deaf people," International Journal of Advanced Research in Computer and Communication Engineering, 3(3), pp. 5319–5321, 2014.

[23] S. Goyal, I. Sharma, and S. Sharma, "Sign language recognition system for deaf and dumb people," International Journal of Engineering Research and Technology, 2(4), pp. 382–387, 2013.

[24] P. Gurjal and K. Kunnur, "Real Time Hand Gesture Recognition Using SIFT," International Journal of Electronics and Electrical Engineering, 2(3), pp. 19–33, 2012.

[25] Y. Jiang, J. Tao, W. Ye, W. Wang, and Z. Ye, "An isolated sign language recognition system using RGB-D sensor with sparse coding," 17th International Conference on Computational Science and Engineering, IEEE, pp. 21–26, 2014.

[26] J. L. Raheja, A. Mishra, and A Chaudhary, "Indian sign language recognition using SVM," Pattern Recognition and Image Analysis, 26(2), pp. 434–441, 2016.

[27] E. Rakun, M. Andriani, I. W. Wiprayoga, K. Danniswara, and A. Tjandra, "Combining depth image and skeleton data from kinect for recognizing words in the sign system for Indonesian language (SIBI [sistem isyarat bahasa Indonesia])," International Conference on Advanced Computer Science and Information Systems (ICACSIS), pp. 387–392, 2013.

[28] E. Escobedo and G. Camara, "A new approach for dynamic gesture recognition using skeleton trajectory representation and histograms of cumulative magnitudes," SIBGRAPI Conference on Graphics, Patterns and Images, pp. 209–216, 2016.

[29] S. B. Carneiro, E. D. D. M. Santos, M. D. A. Talles, J. O. Ferreira, S. G S. Alcala, and A. F. Da Rocha, "Static gestures recognition for Brazilian sign language with Kinect sensor," SENSORS, IEEE, 2016.

[30] C. Keskin, F. Kırac, Y. E. Kara, and L. Akarun, "Real time hand pose estimation using depth sensors," International Conference on Computer Vision Workshops, IEEE, pp. 1228–1234, 2011.

[31] M. Mohandes, S. Aliyu, and M. Deriche, "Arabic sign language recognition using the leap motion controller," 23rd International Symposium on Industrial Electronics (ISIE), IEEE, pp. 960–965, 2014.

[32] M. U. Kakde, M. G. Nakrani, and A. M. Rawate, "A review paper on sign language recognition system for deaf and dumb people using image processing", International Journal of Engineering Research and Technology, 5(3), pp. 590–592, 2016.

[33] H. Bhavsar, "Review on classification methods used in image based sign language recognition system," International Journal on Recent and Innovation Trends in Computing and Communication, 5(5), pp. 949–959, 2017.

[34] L. E. Potter, J. Araullo, and L. Carter, "The leap motion controller: A view on sign language," Proceedings of the 25th Australian Computer-Human Interaction, Conference on Augmentation, Application, Innovation, Collaboration – OzCHI, ACM, pp. 175–178, 2013.