Feature Extraction For Indonesian Sign Language (SIBI) Using Leap Motion Controller

Aditiya Anwar, Achmad Basuki, Riyanto Sigit, Angga Rahagiyanto, Moh. Zikky
Department of Informatics and Computer Engineering
Graduate Program of Engineering Technology
Politeknik Elektronika Negeri Surabaya
Surabaya, Indonesia
anwaraditiya@gmail.com, basuki@pens.ac.id, riyanto@pens.ac.id, rahagiyanto@gmail.com, zikky@pens.ac.id

***Abstract*—Communication is an essential needs for people. Generally, people communicate verbally. However, some people have a deficiency of not being able to communicate verbally like deaf and mute people. They use sign language to communicate. Many research has been done to translate or recognize sign language. Key to sign language translation is hand gesture recognition. In this paper, Indonesian Sign language (SIBI) Recognition System is proposed. This system use Leap Motion Controller device and new feature extraction method to obtain more accurate gesture data. Input data taken from Leap Motion is position data within the time range *t*. This position data is calculated by relative coordinate so as to get the vector feature for a movement. Data classification processed by *k*-Nearest Neighbor (*k*-NN) and Support Vector Machine (SVM). This proposed system using new feature extraction method gives an average accuracy level of sign language recognition equal to 95.15% by using *k*-NN classification and 93.85% by using SVM classification.**

Keywords—Feature Extraction; Hand Gesture Recognition; Sign Language; Leap Motion;

I. INTRODUCTION

Communication is a need for people to interact and exchange information to each other. Communication is divided into two, they are verbal and non-verbal communication. Generally, in daily life, people use verbal communication [1]. However, some of them have disability to use verbal communication. They are deaf and mute. They communicate by sign language. In Indonesia, sign language is divided into two, they are BISINDO and SIBI. BISINDO (Bahasa Isyarat Indonesia) is natural sign language of deaf-mute so that it is often used by them to communicate to each other. SIBI (Sistem Isyarat Bahasa Indonesia) is an Indonesian sign language which standardized by government. SIBI is made to connect the communication between deaf-mute and normal people. Nevertheless, not all of normal people understand SIBI and even not a few of deaf-mute people find it difficult to use SIBI [2].

In recent years, there are many researches about sign language recognition. Previous researches offer a solution to facilitate deaf-mute to be able to communicate with normal people and facilitate normal people to understand sign language. These researches refer to hand gesture recognition of the sign language [3]. Sign language gesture is a body language formed by hand and finger movement and position [4]. There are two types of sign language gesture, they are static and dynamic gesture. Static gesture is a fixed gesture of hand and finger without any change in movement or position within a certain time. While dynamic gesture is a series of hand and finger gesture with a movement which formed a pattern in a certain time. Examples of alphabet sign language illustrated in *Fig. 1*. These alphabet are SIBI sign language. All of the alphabet has static gesture except for alphabet J and Z which have dynamic gesture.

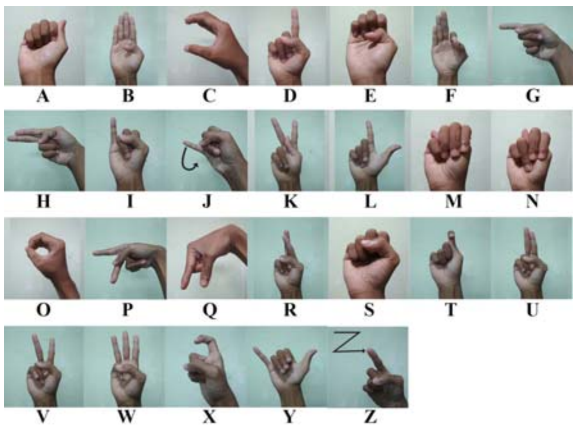

Fig. 1. SIBI sign language

This paper use pattern recognition approaches to recognize SIBI sign language gesture. The device used to support this research is Leap Motion Controller. Leap Motion has a camera sensor which can visualize a real-time 3D hand model in virtual system [5]. We proposed a new feature extraction method to recognize a pattern in hand movement. Our feature extraction involve calculation of relative coordinate in Cartesian coordinate system of the hand. This feature extraction applied in our system to recognize sign language gestures especially SIBI.

II. PREVIOUS WORKS

Many researchers tried to build hand gesture recognition system, especially for sign language recognition. They perform a variety of approaches and use different sensors to build sign language recognition systems.

978-1-5386-0787-9/17/$31.00 ©2017 IEEE

Ching-Hua Chuan, et al [6] build American Sign Language (ASL) recognition system using Leap Motion Controller. The features used for machine learning are pinch strength, grab strength, average distance, average spread, average tri-spread, extended distance, dip-tip projection, orderX, and angle. Pinch and grab strength obtained from the API while the other features derived from the API using specific calculation. They record gesture dataset of 26 ASL alphabet from two people, one of them is deaf. They apply two classification method, those are k-Nearest Neighbor (k-NN) and Support Vector Machine (SVM). By using four-fold cross validation, they obtained 72.78% classification accuracy in k-NN method and 79.83% in SVM method.

A.S. Elons, et al [7] also build sign language recognition system using Leap Motion Controller. Their system recognize Arabic sign language (ArSL). They use two different features selected from motion tracking data of the Leap Motion. These features are fingers position and distances between fingers. These features processed in Multi-layer perceptron Neural Network (MPL) for training and gesture classification. The result of the classification shows gesture recognition accuracy rate of 88% for finger position distance feature.

João Gabriel Abreu, et al [8] evaluate the use of electromyogram (EMG) sensor in Myo Armband as a feature for sign language recognition. The sign language used for evaluation is 20 Brazilian sign language (LIBRAS) alphabet which has static movement. Classification performed by using Support Vector Machine (SVM). As a result, with EMG sensor alone, sign language gesture is very difficult to recognize compared using other hand gesture recognition techniques. However, they believe that the result of EMG data sensor are significant enough to be added to sign language recognition system as additional feature. Yun Li, et al [9] build sign language recognition system using accelerometer and electromyography (EMG) sensor which worn in forearm. They propose a combination of multi-channel and multi-sensor information strategy to automatically recognize sign language at sub-word level. Decision Tree is used as gesture classification method. They perform a test of 121 frequently used Chinese sign language. Data collection in 4 research sessions obtained 4 groups of data samples. Recognition was done using four-fold cross validation with 3 groups as training data and remaining 1 group as test data. By method proposed, the average accuracy rate of the sign language recognition reaches 95.78%.

Leap Motion is a new sensor device than the other hand gesture recognition sensors. Leap motion commonly used in game or VR application. Until now, Leap Motion is still being developed to obtain better hand detection results. Therefore, Leap Motion is a promising sensor device for sign language recognition system.

III. Leap Motion Controller Sensor

In recent years, many devices have been used by researchers to research hand gesture recognition, especially sign language recognition. One of device is Leap Motion Controller. Leap Motion is a device with 2 camera sensors and infrared LED that can detect hand movements in its detection space. The advantage of Leap motion is that it can detect hand movements to the fingers in real-time [7]. Leap Motion reads hand and finger information and visualizes the information in the form of a 3D hand model which is resulting in a series of data. This data are position, speed, and orientation [10]. In this research we focus on position data. Position data are coordinate values of X, Y, and Z in Cartesian coordinate system. This data obtained from certain points of the hand, including the Palm point and the finger joints points. The points that are recognized by Leap Motion are marked by the green ball in *Fig. 2*.

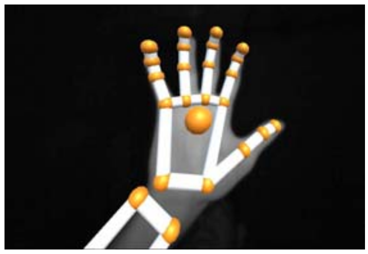

Fig. 2. Recognized points of the hand in Leap Motion

Each finger joints and hand has its own coordinate value with center coordinate reference (0, 0, 0) in the Leap Motion itself. The movement of hand gesture from time to time can be known from the coordinate of the points. Thus, this coordinate point's value can be used to recognize a particular gesture movement.

IV. Proposed System

This section presents the process of a new feature extraction method proposed by us. This feature extraction transforms the input data into a feature vector and selects the features so as to obtain efficient and effective data for later process [11]. In the proposed system, 3D handed models visualized by Leap Motion device are displayed in 3D space coordinate of x, y, and z-axis Cartesian. We provide 16 reference points of coordinate in 3D hand model. These points are scattered in the palm and finger joints as shown in *Fig. 3*.

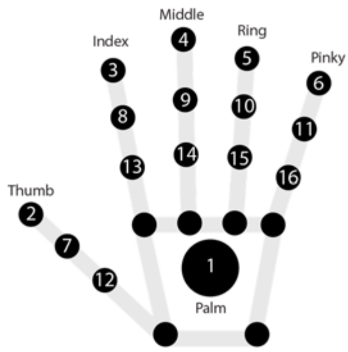

Fig. 3. Reference points in hand

The difference between this proposed system and the existing system in [6] is the features we used. The existing system use feature as explained in previous Section II while our proposed system use positions based on reference points as features.

The other existing sign language recognition system in [7] use fingers position as one of its feature. The system use the tip of each fingers position as explained in previous Section II while our proposed system use more detailed positions in each fingers called reference points as shown in *Fig. 3*.

The reference points in *Fig. 3* include Palm (1), Thumb1 (2), Index1 (3), Middle1 (4), Ring1 (5), Pinky1 (6), Thumb2 (7), Index2 (8), Middle2 (9), Ring2 (10), Pinky2 (11), Thumb3 (12), Index3 (13), Middle3 (14), Ring3 (15), and Pinky3 (16). Each of these points have 3 values (x, y, and z coordinate) so that there are 48 raw data values as shown in TABLE I.

TABLE I. 48 DATA VALUE OF COORDINATES DETAIL IN HAND

Palm			Thumb1			…	Pinky3		
x	y	z	x	y	z	…	x	y	z
Value 1	Value 2	Value 3	Value 4	Value 5	Value 6	…	Value 46	Value 47	Value 48

Leap Motion is a sensor device that captures data in real-time, so that the coordinate points collection on each gesture is limited to ten times (ten frames) in *t* seconds of capture as in *Fig. 4*. This *t* is the time it takes for people to perform a hand sign language gesture.

Fig. 4. Collection of coordinate points is performed 10 times (frames) in time range *t*

Total data values taken in a single data retrieval is 16 points multiplied by 3 coordinates (x, y, z) multiplied by 10 frames so that the total value obtained is 480. To alleviate the performance of the system, it is necessary exclude some points. The excluded points are Thumb2 (7), Index2 (8), Middle2 (9), Ring2 (10), and Pinky2 (11) in *Fig. 3*. These points are excluded because they tend to have the same data pattern as Thumb1 (2), Index1 (3), Middle1 (4), Ring1 (5), and Pinky1 (6). Thus, the total data value taken is 11 point multiplied by 3 coordinates multiplied by 10 frames equals to 330 values.

The detail of the points used are Palm (1), Thumb1 (2), Index1 (3), Middle1 (4), Ring1 (5), Pinky1 (6), Thumb3 (12), Index3 (13), Middle3 (14), Ring3 (15), and Pinky3 (16) as shown in *Fig. 3*. From these points, in single gesture data retrieval, the value obtained as shown in TABLE II.

TABLE II. DATA OBTAINED FROM A SINGLE GESTURE

Frame	Palm			Thumb1			…	Pinky3		
	x	y	z	x	y	z	…	x	y	z
1	Value1	Value2	Value3	Value4	Value5	Value6	…	Value31	Value32	Value33
2	Value34	Value35	Value36	Value37	Value38	Value39	…	Value64	Value65	Value66
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10	Value298	Value299	Value300	Value301	Value302	Value303	…	Value328	Value329	Value 330

The use of relative coordinate performed to obtain dynamic motion patterns in the data. The palm coordinate point at the beginning of the frame is used as the reference point of the relative coordinates. Thus, all the coordinates (x, y, z) in the 10 frames are subtracted by the first palm value (x, y, z) of the frame to obtain the data in TABLE III.

TABLE III. NEW SINGLE GESTURE DATA AFTER A REDUCTION OPERATION BY RELATIVE COORDINATE

Frame	Palm			Thumb1			…	Pinky3		
	x	y	z	x	y	z	…	x	y	z
1	0	0	0	Value1	Value2	Value3	…	Value28	Value29	Value30
2	Value31	Value32	Value33	Value34	Value35	Value36	…	Value61	Value62	Value63
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10	Value295	Value296	Value297	Value298	Value299	Value300	…	Value325	Value326	Value327

In TABLE III, the first frame palm point is always 0 because it is subtracted by itself which is the reference point of the relative coordinate. Values other than this relative coordinate reference point value are used as features so that there are a total of 327 features in a single gesture. These features are the values used to represent a hand gesture. With this feature, both static and dynamic gesture can be recognized. Dynamic gesture patterns can be recognized in the change of palm point values from frames 2 to 10.

TABLE IV. DATASET MODEL

Class	Feature1	Feature2	Feature3	…	Feature327
Gesture 1	Value 1	Value 2	Value 3	…	Value 327
Gesture 2	Value 1	Value 2	Value 3	…	Value 327
⋮	⋮	⋮	⋮	⋮	⋮
Gesture n	Value 1	Value 2	Value 3	…	Value 327

TABLE IV is an illustration of the system dataset model. The classes in TABLE IV are the labels of a gesture, while Feature 1 to Feature 327 are values that shape the pattern of a gesture.

The *k*-Nearest Neighbor (*k*-NN) Algorithm is used to classify new gesture inputs based on previously created datasets. The principle of k-NN is to find *k* data in a dataset that has a similarity to new data based on minimum distance. This minimum distance is calculated using the Euclidean Distance (d_i), as in (1).

$$d_i = \sqrt{\sum_{i=1}^{n}(x_{2i} - x_{1i})^2} \quad (1)$$

Variable x_1, in (1), is new test data, x_2 is training data in dataset, and n is number of data in dataset. Each data in dataset calculated with the result that each data has values $\{d_1, d_2, d_3, \ldots, d_n\}$. These values ranked based on minimal value, and then top k values are selected. The test data is classified based on the dominant label on the k-data.

V. Result

In the proposed system, we record 26 alphabet A-Z signs from SIBI. These signs performed by 5 different persons. Each person perform 10 times record for each sign. Thus, every alphabet has 50 gesture variant. As a result, the dataset has a total instance of 1300 gesture data.

We applied k-NN for sign recognition. We validate the data using leave-one-out cross validation. The highest accuracy achieved by using k-NN is 95.15% with k = 1. The detailed k-NN accuracy result is shown in TABLE V.

TABLE V. K-NN ACCURACY RESULT

Alphabet	**Accuracy**	**Miss-Classified as**	**Alphabet**	**Accuracy**	**Miss-Classified as**
A	94%	N (2%) S (4%)	N	66%	A (4%) M (22%) S (8%)
B	100%	-	O	100%	-
C	100%	-	P	96%	H (2%) T (2%)
D	96%	L (2%) X (2%)	Q	100%	-
E	100%	-	R	90%	K (2%) U (6%) X (2%)
F	98%	Z (2%)	S	84%	A (2%) D (2%) M (4%) N (8%)
G	98%	Z (2%)	T	100%	-
H	96%	O (4%)	U	96%	K (2 %) R (2%)
I	100%	-	V	96%	U (4%)
J	100%	-	W	98%	B (2%)
K	92%	R (2%) U (4%) V (2%)	X	94%	D (2%) T (4%)
L	98%	B (2%)	Y	100%	-
M	84%	N (12%) S (4%)	Z	98%	D (2%)

As for comparison, we also applied Support Vector Machine (SVM) and validate the data using leave-one-out cross validation. There are various kernel in SVM but the one resulting the best performance is Radial Basis Function (RBF) kernel. This kernel yields 93.85% accuracy. The detailed SVM accuracy result is shown in TABLE VITABLE VI.

TABLE VI. SVM ACCURACY RESULT

Alphabet	**Accuracy**	**Miss-Classified as**	**Alphabet**	**Accuracy**	**Miss-Classified as**
A	88%	M (2%) S (10%)	N	74%	A (4%) M (14%) S (8%)
B	100%	-	O	100%	-
C	100%	-	P	96%	X (2%) Z (2%)
D	96%	Z (4%)	Q	100%	-
E	98%	Z (2%)	R	94%	U (4%) X (2%)
F	98%	Z (2%)	S	66%	A (14%) E (2%) M (8%) N (8%) Z (2%)
G	98%	Z (2%)	T	98%	N (2%)
H	98%	O (2%)	U	76%	K (8%) R (16%)
I	100%	-	V	98%	R (2%)
J	100%	-	W	98%	B (2%)
K	92%	R (2%) V (6%)	X	94%	D (2%) T (4%)
L	98%	B (2%)	Y	100%	-
M	80%	N (18%) S (2%)	Z	100%	-

TABLE VI and TABLE VI show that feature extraction using relative coordinate and calculating values on frames in t periods yields a high rate accuracy. From the result, k-NN gives better performance result than SVM. The average overall accuracy rate of k-NN is 95.15%. Both static and dynamic movements can be well recognized. However, there are some miss-classification in some features of the gesture. Some of them are movements A, E, M, N, and S in *Fig. 5*.

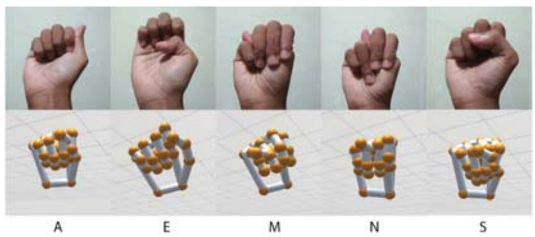

Fig. 5. Gesture A, E, M, N, and S

Although the gesture of A, E, M, N, and S in *Fig. 5* seems different, but in reality, Leap Motion Sensor detection provide almost similar information results. This can be seen in the visualization of 3D hand models from Leap Motion in *Fig. 5*. The shape of the hand gestures of the five alphabets tends to have a clenching hand. Visualization of 3D hand model by Leap Motion has limitations in the curve of the fingers which are not too deep and detail. For example, the curve of the thumb gesture of the alphabet E is not very deep and tends to be similar to the thumb gesture A and S curve. However, the system can still recognize each of the movements A, E, M, N, and S with fairly high accuracy. The amount of data in the dataset also affects the accuracy. This is proven by validating data from 2 persons. As a result, the average accuracy rate dropped to 90.19% by using *k*-NN with $k = 1$.

VI. CONCLUSION

In this study, we propose a sign language recognition system using the new feature extraction method. The sign language we use is Indonesian Sign Language (SIBI). Our feature extraction only retrieves the position data information from the points on the hand recognized by Leap Motion. Position data is calculated using the concept of relative coordinate. We have recorded the A-Z alphabet gesture dataset which each gesture has 10 data from 4 different people. We test this data using the Leave-one-out cross validation and classify it using *k*-Nearest Neighbor and Support Vector Machine. Based on the results, our feature extracted data obtained an average accuracy rate of 95.15% by using *k*-NN and 93.85% by using Support Vector Machine. This accuracy is achieved by only utilizing position data only. However, there are still some misclassification. Therefore, in future research, this system can be added by other features available from Leap Motion such as orientation, speed, or other information about fingers and hands to have a better system performance. This research also still use one hand gesture which is right hand. For the future works, this research can be developed into a two-handed sign language recognition system.

REFERENCES

[1] D. S. Ekasari, U. L. Yuhana and R. R. Hariadi, "Design and Implementation of Sign Language Recognition Module Using Kinect," *JURNAL TEKNIK POMITS,* vol. 1, no. 1, pp. 1-6, 2013.

[2] I. H. Zusfindhana, "The Use of Indonesian Sign Language System (SIBI) and Indonesia Sign Language (BISINDO) by Teenage Deaf Student in SLB-B Kota Bandung," *International Conference On Special Education in South East Asia Region,* pp. 217-223, 2017.

[3] F. Ramirez-Garibay, C. M. Olivarria, A. F. E. Aguilera and J. C. Huegel, "MyVox—Device for the communication between people: blind, deaf, deaf-blind and unimpaired," *IEEE Global Humanity Technology Conference,* pp. 506-509, 2014.

[4] S. Mitra and T. Acharya, "Gesture Recognition: A Survey," *IEEE Transactions on Systems, Man, and Cybernetics,* vol. 37, no. 3, pp. 311-324, 2007.

[5] F. Weichert, D. Bachmann, B. Rudak and D. Fisseler, "Analysis of the Accuracy and Robustness of the Leap Motion Controller," *Sensors,* vol. 13, no. 5, pp. 6380-6393, 2013.

[6] C.-H. Chuan, E. Regina and C. Guardino, "American Sign Language Recognition Using Leap Motion Sensor," *2014 13th International Conference on Machine Learning and Applications,* pp. 541-544, 2014.

[7] A. S. Elons, M. Ahmed, H. Shedid and M. F. Tolba, "Arabic sign language recognition using leap motion sensor," *2014 9th International Conference on Computer Engineering & Systems (ICCES),* pp. 368-373, 2014.

[8] J. G. Abreu, J. M. Teixeira, L. S. Figueiredo and V. Teichrieb, "Evaluating Sign Language Recognition Using the Myo Armband," *2016 XVIII Symposium on Virtual and Augmented Reality (SVR),* pp. 64-70, 2016.

[9] Y. Li, X. Chen, J. Tian, X. Zhang, K. Wang and J. Yang, "Automatic recognition of sign language subwords based on portable accelerometer and EMG sensors," *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction,* no. 17, 2010.

[10] M. Buckwald, "Introducing the Skeletal Tracking Model," Leap Motion, [Online]. Available: https://developer.leapmotion.com/documentation/python/devguide/Intro_Skeleton_API.html. [Accessed 2 July 2017].

[11] XuechuanWang and K. K. Paliwal, "Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition," *Pattern Recognition,* vol. 36, no. 10, pp. 2429-2439, 2003.