

Gesture Recognition for Indonesian Sign Language Systems (ISLS) Using Multimodal Sensor Leap Motion and Myo Armband Controllers Based-on Naïve Bayes Classifier

Khamid

Department of Informatics Engineering
Wahidiyah University
Kediri, Indonesia
perumsat@gmail.com

Adhi Dharma Wibawa, Surya Sumpeno

Department of Computer Engineering
Sepuluh Nopember Institute of Technology
Surabaya, Indonesia
adhiosa@ee.its.ac.id, surya@ee.its.ac.id

Abstract—Indonesian Sign Language System (ISLS) has been used widely by Indonesian for translating the sign language of disabled people to many applications, including education or entertainment. ISLS consists of static and dynamic gestures in representing words or sentences. However, individual variations in performing sign language have been a big challenge especially for developing automatic translation. The accuracy of recognizing the signs will decrease linearly with the increase of variations of gestures. This research is targeted to solve these issues by implementing the multimodal methods: leap motion and Myo armband controllers (EMG electrodes). By combining these two data and implementing Naïve Bayes classifier, we hypothesized that the accuracy of gesture recognition system for ISLS then can be increased significantly. The data streams captured from hand-poses were based on time-domain series method which will warrant the generated data synchronized accurately. The selected features for leap motion data would be based on fingers positions, angles, and elevations, while for the Myo armband would be based on electrical signal generated by eight channels of EMG electrodes relevant to the activities of linked finger's and forearm muscles. This study will investigate the accuracy of gesture recognition by using either single modal or multimodal for translating Indonesian sign language. For multimodal strategy, both features datasets were merged into a single dataset which was then used for generating a model for each hand gesture. The result showed that there was a significant improvement on its accuracy, from 91% for single modal using leap motion to 98% for multi-modal (combined with Myo armband). The confusion matrix of multimodal method also showed better performance than the single-modality. Finally, we concluded that the implementation of multi-modal controllers for ISLS's gesture recognition showed better accuracy and performance compared of single modality of using only leap motion controller.

Keywords—Myo armband; leap motion; Naïve Bayes; Indonesian sign language systems; dynamic hand gestures; sign language.

I. INTRODUCTION

Recently the technology for body movement control devices has been growing fast and used by many people for many purposes such as leap motion and Myo armband. One of the biggest implementation of this technology is for supporting patients or disability persons including for building communication. Communication is a basic need of all human being, no exception for disability person with hearing-impaired condition. This group of people (deaf community) have quite complex problem in communication. The only way for them to get information is via visual mode. Therefore for this group of people, Sign Language (SL) is then becoming the best option. There have been some studies previously that support the deaf-community in many forms of tools such as keypads, gloves or wearable devices [1]–[2] so that they can still establish a communication by using Sign Language. Sign language is a gesture/motion of hands for expressing words or sentences. However, since the sign language is done by making a certain motion or gesture of hands then some variations becoming an important issue when a computer-based automatic translation application is developed.

Such application is not only important for giving them information from outside of their world but also essential to educate the young generation of hearing-impaired patients the way to communicate in a correct way using sign language. Sign language (SL) in many countries of world-wide has its own pattern and methodology. Each world-wide language has its own way in developing the technique [3]–[4] including Indonesia Sign Language System (ISLS). ISLS is one of sign language standard which was adopted based on Bahasa Indonesia. Technically ISLS is most similar to American Sign Language where its sign language represents the letters, numbers, and words independently. For expressing phrase or series of words, it constructs based on its basic word. For example the word “I go to school” performed from 4 words: “I”, “go”, “to”, and “school”. More, numbers 21 would be constructed from numbers: “2”

and “1”. And so on. Then, this standard would be easy to be followed or learned for anyone who interested in this sign language.

Previously there have been several studies in the area of sign language. Some of them using the same controllers of this study, and some of them combined with other type of controllers. However, combining Myo armband and leap motion technology is a new approach. For example, [5] tried to recognize 26 standard alphabet letters by using leap motion. By implementing Decision Tree algorithm, [3] gained 82% of 24 alphabet letters, while two other letters were error. Refs [4],[6] have used leap motion and Kinect technology in recognizing hand gestures. Giulio reached 91.28% of accuracy. Meanwhile other recent study using Myo armband was done by [7]. This study showed that when the hand gesture was made in static condition, the accuracy was high, however when the gesture was done in dynamic condition, the error was high.

From those results we summarize that the main problem in developing SL speech-to-text translation is the variation of the motion, for example when people saying “I” in SL, it is done by lifting the right hand into as high as the human chest and pointing to ourselves, some people may perform it differently (in term of height) and as a result this could lead to low accuracy in translating the SL to text. In this paper, we hypothesized that by combining two modalities such as leap motion and Myo armband controllers for recognizing the gesture of ISLS, the translation result would yield better accuracy compared to translating SL using only one modality. Previous study done by [8] had used two modalities in translating SL but they implemented via online application (web base), while the novelty of this paper is the implementation of translating SL using two modalities based on portable and cheap wearable devices with recent technology.

The main goal of this paper is presenting a gesture recognition system applied to Indonesia Sign Language Systems (ISLS) which consists of various hand poses either static and dynamic gestures representing alphabet or number letters, and also words/phrases by using one modality and then being compared to using two modalities.

II. METHOD

This study used leap motion controller to read the data of hand gesture, whether in static or dynamic gesture. Leap motion controller can track and record hand motion into detail such as finger-tips motion, hand/palm positions and its orientation. One of our great challenge in implementing leap motion controller in this study is that in ISLS library, there are some alphabets and words that have similar poses with some small different on finger-tip positions, movements, and orientations. This condition certainly could lead to bigger error when translating the SL to the text due to condition that the similar data representing different corpus/words. Therefore, in order to make better performance, we used Myo armband as the 2nd modality. It can be used as 2nd references for recognizing the gesture. This controller recorded the muscle activities of each finger while contracting due to the hand pose. The advantage of applying

both controllers (leap motion and Myo armband) is because of its portability, compactness, low price, and rich of information generated through wireless technology.

By extracting features of each controller independently, there would be a complete feature dataset relevant to ISLS’s gestures. The Naïve Bayes machine learning used to classify the gestures taken from this library. There are 10 static and dynamic gestures which have similar poses acquired for data training and sample test. The subjects consist of 3 students who were asked to do the data training initially based on chosen poses and to do the sample data evaluation.

Both controllers: leap motion and Myo armband can be seen physically on the Fig. 1. leap motion ported to personal computer over USB port, and Myo armband over bluetooth connectivity.

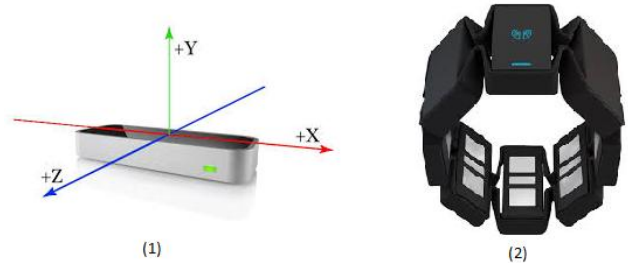


Figure 1. Leap motion (1) and Myo armband (2) controllers

Each of these controllers has the System Development Kits (SDK) libraries which enables application developer to create tools in order to use these controllers. These SDK libraries will supply the information generated by the controller based on either by pooling or event based. By pooling method means that the controller will supply data by request. But, by event means data supplied by controllers immediately generated once it processes the information.

A. Leap motion

Leap motion generates information of hand’s pose by frame detection method. It generates the hand, palm, and fingers information which consist of hands, fingers, and orientation. From hands we could extract palm center position in 3D space, and from fingers we could extract finger tips and dinger bones positions, relatively from palm center position, and for orientation information shows which way is pointed in format of roll, pitch, and yaw. This controller is sensitive of an infra-red radiation source. It answers why this controller recommended for indoor usage only because of the sun’s infrared radiation.

B. Myo armband

There are 8 channel sensors included within Myo armband controllers. Each sensor generates electrical signal relevant to muscle contraction on upper forearm. Some channels represent the contraction of fingers, and the rest for another muscles which is not relevant to finger movements, but rotation.

Each channel generates the electric signal value between 0 – 125 μ volt unit with 100 – 120 sample rate per second

which is dependent to the how many infra-redspectrum around the light. Fig. 2. shows the electric signal level of each channel for single gesture detected.

The research methodology used in this paper was divided into 2 major parts, leap motion part (yellow blocks) and Myo armband part (the blue blocks). See on Fig. 3.

C. Data acquisition

Data acquisition for both controllers is done by using time-series method. So, the data captured from one controller is synchronized to another one based on certain time marker.

In this paper the initiate time of capture is based on leap motion first time frame capture. The tools used for this data capture was written with referencing to both controller's SDK using Visual Studio 2015 C#.NET. The tools will keep receiving the data flow from Myo armband controller considering it keeps generating the muscle electric signal although the muscles are in relax state. But, the leap motion data will be triggered only when a frame captured by the controller. The data capture method for both controllers shown on Fig. 4.

Once a new frame detected, the tools cut the Myo armband data-flow and marked as first data row. And when the frame gone, the tools ends the data capture and marked as one gesture information. Then, the tools will have a set of data gesture for both controllers which is ready to be extracted for the features extraction.

D. Leap Motion Features extraction

Although the accuracy of 3D data generated by leap motion controllers is 200 μ sec which is very high, however not all fingers motion were being recognized well especially when it is touching each other, or folded with others, or being hidden from controller camera. In many cases, whenever an object is not perpendicular to camera, it suggested as a lost finger. Then, the number of fingers features cannot be used.

The most extracted features used for gesture recognizing method were:

- Angle of fingertips: which can be calculated based on following formula:

$$A_i = \angle(F_i^\pi - C, h); i = 1, \dots, 5 \quad (1)$$

where F_i^π is the F_i projection on the plane i , are the angles related to the orientation of the fingertips with the hand orientation. So, the actual number of fingers captured by controller would affect the orientation h and angle of fingertips.

- Distance of fingertips: calculated based on fingertips and palm center positions.

$$D_i = \|F_i - C\|, i = 1, \dots, 5 \quad (2)$$

It represents a 3D distance of fingertips F_i from a hand-center C .

- Elevation of fingertips: represents the distance of fingertips from the plane region.

$$E_i = \text{sgn}((F_i - F_i^\pi) \cdot n) \|F_i - F_i^\pi\|, i = 1, \dots, 5 \quad (3)$$

E. Myo armband feature extraction

There are various features that can be extracted from Myo armband which are captured from 8 available channels. However in this paper we extracted only the 5 most used features in the gesture recognition method. They were Mean Absolute Value (MAV), Zero Crossing (ZC), Root Mean Square (RMS), Waveform Length (WL), and Variance (VAR).

- Mean Absolute Value: is calculated based on raw data value above threshold. Data should be cleaned by using the predefined absolute value then being processed based on following formula:

$$MAV = \frac{1}{N} \sum_{k=1}^N |X_k| \quad (4)$$

where k is index of data sample and N is number of samples.

- Waveform Length: is calculated based on the length of cumulative of EMG waveforms on certain time period. The calculation for this feature would follow the formula below:

$$WL = \sum_{k=1}^N |X_{k+1} - X_k| \quad (5)$$

- Zero Crossing: is determined based on how many times the signal crossing the zero line for either upper or lower level. The calculation would be following the formula:

$$ZC = \sum_{k=1}^{N-1} [\text{sgn}(X_k \times X_{k+1}) \cap |X_{k+1} - X_k| \geq \text{thr}], \quad (6)$$

$$\text{sgn}(x) = \begin{cases} 1, & \text{if } x < 0 \\ 0, & \text{otherwise} \end{cases}$$

where thr is the threshold value.

- Root Mean Square: is calculated for determining the ratio between direct levels of voltage and its alternation. This will represent the power of EMG and is following the formula below:

$$RMS = \sqrt{\frac{1}{N} \sum_{k=1}^N X_k^2} \quad (7)$$

- Variance: is calculated based on average of the square value of amplitude of EMG signal. The formula regarding to this calculation would be following:

$$VAR = \frac{1}{N-1} \sum_{k=1}^N X_k^2 \quad (8)$$

F. Complete Feature-Set

Based on extracted features from both controllers described on previous section, there were 2 groups of features sets called \mathbf{D}_{leap} and \mathbf{D}_{myo} . The feature group of leap motion can be formulated as $\mathbf{D}_{\text{leap}} = [\mathbf{A}, \mathbf{D}, \mathbf{E}]$ where they are from angle, distance, and elevation features. The other can be formulated as $\mathbf{D}_{\text{myo}} = [\mathbf{M}, \mathbf{Z}, \mathbf{R}, \mathbf{W}, \mathbf{V}]$, which was representing all features extracted from Myo armband controller.

The concatenation of both feature-groups performs a complete feature sets $[\mathbf{D}_{\text{leap}}, \mathbf{D}_{\text{myo}}]$ where it represents a single gesture. Fig. 5 shows the structure of complete feature-sets performed from both controller's dataset.

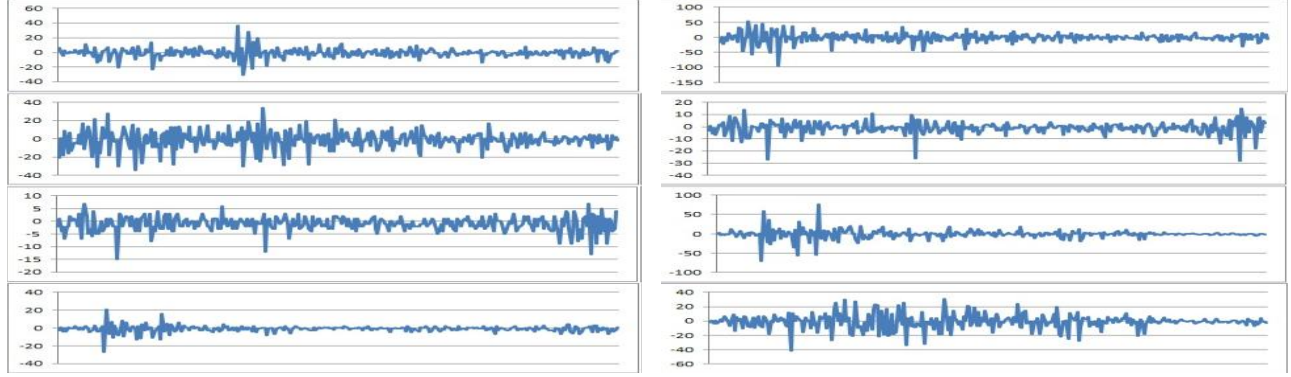


Figure 2. Myo armband electric signal

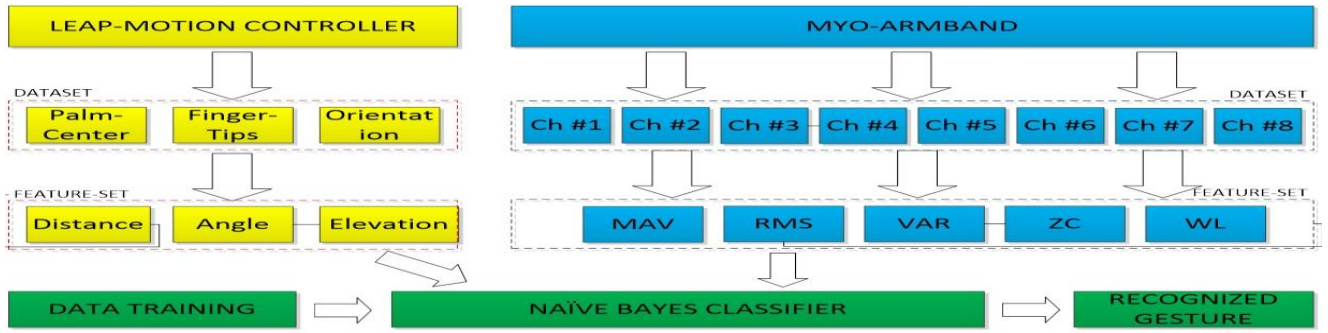


Figure 3. Classification method



Figure 4. Classification method

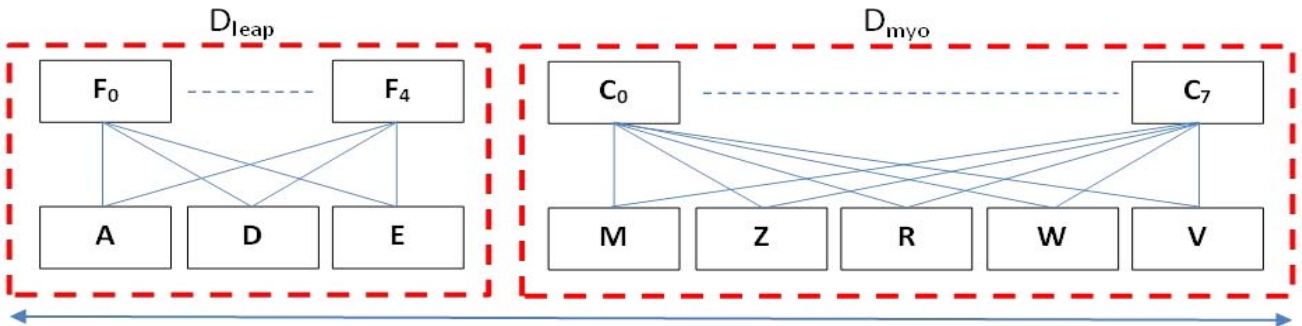


Figure 5. Leap motion and myo armband feature-sets

G. Classifier

This research used Naive Bayes [9] for its basic classifier. This classifier was chosen due to its simplicity in technique and implementation. The basic principal of Naive Bayes is assuming every feature in the feature-sets is not dependent each others. Its efficiency in using the data training, by using a small amount of data-set without reducing the accuracy of classification performance makes this classifier widely used on many gesture recognition researches.

Basically, Naive Bayes is a statistical approach with the probability formula as follows:

$$P(C_k | x) = \frac{P(C_k)P(x | C_k)}{P(x)} \quad (9)$$

where x is vectors of features and C_k is possible of classes of k . In the readable words, the equation above can be simplified as follows:

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}} \quad (10)$$

Assumed the naive independence is formulated as below:

$$P(x_i | C_k, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = P(x_i | C_k) \quad (11)$$

Then for all i , we could simplify to

$$P(C_k | x_1, \dots, x_n) = \frac{P(C_k) \prod_{i=1}^n P(x_i | C_k)}{P(x_1, \dots, x_n)} \quad (12)$$

with $P(x_1, \dots, x_n)$ is a constant then we can estimate $P(C_k)$ and $P(x_i | C_k)$ by using Maximum A Posteriori or MAP. This simplicity of MAP is called Naive Bayes.

III. EXPERIMENT AND RESULT

Using a tools titled as Weka, for presenting and evaluating the classifier method based on model generated from feature-sets which is formatted on csv file, we can do the review in quick and most efficient. Weka already presents the default Naive Bayes classifier as well, so that we just need to input the data model for further processing.

The research tried to classify the gestures into two steps: gesture classification for leap motion, which means $[D_{\text{leap}}]$ feature only involved, and for complete feature-set $[D_{\text{leap}}, D_{\text{myo}}]$. These steps were used for showing how accurate the classification between single and multi modality. In this study, 3 pre-trained subjects were involved while each of subject posed 5 gestures with 20 samples for each gesture. Totally, there were 300 samples of gestures that would be used for the data training and the sample test partially.

For the testing purpose, Weka [10] tools was also used considering its rich functions for machine learning purpose. Once captured data was processed for feature extraction, it would generate a series of features which was a paired of extracted features from both controllers. Fig. 3 shows the series of features generated by our data acquisition tools.

There were 5 distances, 5 angles, and 5 elevators for leap motion features sets. For Myo armband, there were 8 features for each MAV, Zero-Crossing, RMS, Wave-Form Length, and VAR features-sets. It means there were total 55 features-sets for every gestures captured from both

controllers. In this experiment, 20 samples were used as data training per each gestures.

On sign language, most dynamic gestures are based on the static ones. This experiment involved 10 different gestures which were basically 5 static gestures, and the other 5 were the dynamic version of static gestures. These static gestures were: number 5, letter C, U, D, and B [1]. While the dynamic gesture were the movement of the static gestures. They were: 5-hi, C-mess, U-negative, D-where, and B-come on. The 5-hi means static gesture "5" when it moved becomes "hi" gesture. When the static gesture "C" moved then it becomes "mess" gesture, and so on.

One of the reasons implementing Naive Bayes (NB) classifier is its fast processing reason, although when it involves a bunch of data. This is good for gesture recognition which is requiring a very fast processing with a lot of data on every capture session. In this experiment, which was involving 2 different controllers, the speed of either data capturing, feature extraction, and classifying process was absolutely important. Practically, this is good for sign language apps such as ISLS or any other apps which presents most realtime response.

First step of experiment initiated by classifying the gestures captured by leap motion controller. We used the ratio 60:40 of data captured for data training and data sample. The result was shown on Table I for leap motion classification controller only.

TABLE I. LEAP MOTION ONLY CLASSIFICATION

Result of	Samples	Percents
Correctly Classified Instance	134	91.78%
Incorrect Classified Instance	12	8.22%
Relative Absolute Error		8.47%
Root Mean Squared Error		98.01%
Total Instance	146	

TABLE II. CONFUSION MATRIX LEAP MOTION ONLY

	a	b	c	d	e	f	g	h	i	j
a=5	16	0	0	0	0	0	0	0	0	0
b=5-Hi	5	11	0	0	0	0	0	0	0	0
c=U	0	0	15	0	0	0	0	0	0	0
d=U-Negative	0	0	1	17	0	0	0	0	0	0
e=D	0	0	0	0	12	7	0	0	0	0
f=D-Where	0	0	0	0	3	11	0	0	0	2
g=B	0	0	0	0	0	0	17	0	0	0
h=B-Come on	0	0	0	0	0	0	0	12	0	0
i=C	0	0	0	0	0	0	0	0	9	1
j=C-Mess	0	0	0	0	0	0	0	0	0	14

The confusion matrix of this classification result can be found on Table II. In that table we found that most of gestures can reach more than 90% recognition comparable with its static or dynamic version. But, some gestures fall below that level. For example, for gesture 5-Hi, from 11+5 taken samples, 11 times correctly classified as 5-Hi, 5 times classified as 5, which means 31.25% in-accurately classified. The other is gesture D compared with D-Where. The D has

12+7 taken samples, 12 times correctly classified, and the rest classified incorrectly which means 36.84% in-accurately classified.

The next step is classifying the gestures based on the multimodal controller inputs which can be found on Table III. And the confusion matrix for this classification can be found on Table IV.

TABLE III. LEAP MOTION AND MYO ARMBAND CLASSIFICATION

Result of	Samples	Percents
Correctly Classified Instance	134	98.63%
Incorrect Classified Instance	12	1.37%
Relative Absolute Error		1.72%
Root Mean Squared Error		17.68%
Total Instance	146	

TABLE IV. CONFUSION MATRIX OF LEAP MOTION AND MYO ARMBAND

	a	b	c	d	e	f	g	h	i	j
a=5	16	0	0	0	0	0	0	0	0	0
b=5-Hi	2	14	0	0	0	0	0	0	0	0
c=U	0	0	15	0	0	0	0	0	0	0
d=U-Negative	0	0	0	18	0	0	0	0	0	0
e=D	0	0	0	0	12	0	0	0	0	0
f=D-Where	0	0	0	0	0	16	0	0	0	0
g=B	0	0	0	0	0	0	17	0	0	0
h=B-Come on	0	0	0	0	0	0	0	12	0	0
i=C	0	0	0	0	0	0	0	0	10	0
j=C-Mess	0	0	0	0	0	0	0	0	0	14

IV. DISCUSSION AND CONCLUSION

As shown on the confusion matrix above (table IV) most of gestures classification generated above 95% (based on precision and recall calculation) accuracy and average accuracy rate for all gestures is 98.63%. It means the combination of features from both leap motion and Myo armband generate the better performance and accuracy of classification compared to one modality.

During leap motion usage only, it is frequently failed/unable to recognize the hidden or overlapped fingers considering its hardware limitation which is the infra-red camera view. In contrast the Myo armband controller was always generating electric signal relevant to what fingers did. The characteristics of Myo armband controller which generates the electrical signal of the muscle activities provide better recognition of fingers's poses. The monitored muscles of Myo armband are reflecting the fingers movements such as rotating, elevating, or simple moving.

This study has been proving that by implementing two modalities for translating ISLS into text showed the better result. This is because one modality will cover the other modality in translating the corpus. We did not translate all the corpus or words in ISLS library due to our time limitation, however in principle this result showed the potential of using two modalities in translating ISLS library into text with high accuracy so that some other applications,

for helping deaf community or patients with hearing-impaired then, can be made such as application for training them using ISLS, or educate them in the form of game, etc. Implementing hundreds of corpus of ISLS and combining them into sentence would be our near future challenge in using two modalities.

Some of ISLS gestures need the motion of two hands, this could also be another challenge in the future works because one hand can block another hand from the leap motion view and this could cause error in recognizing. Variation in position of those two hands would also influence the data in leap motion for example, which hand is on top of another, the right hand or the left hand. During our experiment, we found that when the gestures were performed by non-skilled ISLS practitioner the variation of the gestures were high, and this could causing error in the process of translating. The accuracy will remain high when the practitioner was well-trained in performing ISLS.

ACKNOWLEDGEMENT

There is no conflict of interest in performing this study. And very thanks to persons who supported this research, for either their abilities for becoming the research's subjects, providing any permissions for legal administrations, and giving the knowledge sharing.

REFERENCES

- [1] Departemen Pendidikan Nasional, Kamus Sistem Isyarat Bahasa Indonesia [Indonesian Sign Language Dictionary], 3rd ed., Jakarta (Indonesia): Department of National Education, 2001.
- [2] W. Lu, Z. Tong, and J. Chu, "Dynamic Hand Gesture Recognition with Leap Motion Controller," IEEE Signal Processing Letters, vol. 23, no. 9, pp. 1188–1192, Sep. 2016, doi: 10.1109/LSP.2016.2590470.
- [3] P.R. Cavanagh and P.V. Komi, "Electromechanical Delay in Human Skeletal," European J. Applied Physiology and Occupational Physiology, vol. 42, no. 3, pp. 159–163, Nov. 1979.
- [4] K. Englehart and B. Hudgins, "A Robust, Real-Time Control Scheme for Multifunction Myoelectric Control," IEEE Trans. on Biomedical Engineering, vol. 50, no. 7, pp. 848–854, Jul. 2003, doi: 10.1109/TBME.2003.813539.
- [5] D. Naglot and M. Kulkarni, "ANN Based Indian Sign Language Numerals Recognition Using the Leap Motion Controller," Proc. Int. Conf. on Inventive Computation Technologies (ICICT), Coimbatore (India), Aug. 2016, vol. 2, doi: 10.1109/INVENTIVE.2016.7824830.
- [6] Leap Motion, "Leap Motion SDK and Plugin Documentation." [Online]. Available: <https://developer.leapmotion.com>.
- [7] Thalmic Labs, "Myo Gesture Control Armband." [Online]. Available: <https://www.myo.com>.
- [8] E. Fujiwara, M.F.M. dos Santos, and C.K. Suzuki, "Flexible Optical Fiber Bending Transducer for Application in Glove-Based Sensors," IEEE Sensors J., vol. 14, no. 10, pp. 3631–3636, Oct. 2014, doi: 10.1109/JSEN.2014.2330998.
- [9] 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, May 2013, doi: 10.3390/s130506380.
- [10] K.A. Sunitha, P.A. Saraswathi, M. Aarthi, K. Jayapriya, and S. Lingam, "Deaf-Mute Communication Interpreter – A Review," Int. Journal of Applied Engineering Research, vol. 11, no. 1, pp. 290–296, 2016.