

HAND GESTURE RECOGNITION FOR INDONESIAN SIGN LANGUAGE INTERPRETER SYSTEM WITH MYO ARMBAND USING SUPPORT VECTOR MACHINE

Aditiya Anwar¹, Achmad Basuki², Riyanto Sigit³

Graduate School of Informatics and Computer Engineering

Politeknik Elektronika Negeri Surabaya

Jl. Raya ITS Sukolilo Surabaya 60111, Indonesia

Telp: 6231 5947280 Fax: 6231 5946114

E-mail: anwaraditiya@gmail.com¹, basuki@pens.ac.id², riyanto@pens.ac.id³

Abstract

Hand gestures are the communication ways for the deaf people and the other. Each hand gesture has a different meaning. In order to better communicate, we need an automatic translator who can recognize hand movements as a word or sentence in communicating with deaf people.

This paper proposes a system to recognize hand gestures based on Indonesian Sign Language Standard. This system uses Myo Armband as hand gesture sensors. Myo Armband has 21 sensors to express the hand gesture data. Recognition process uses a Support Vector Machine (SVM) to classify the hand gesture based on the dataset of Indonesian Sign Language Standard. SVM yields the accuracy of 86.59% to recognize hand gestures as sign language.

Keywords: *Hand Gesture Recognition, Feature Extraction, Indonesian Sign Language, Myo Armband, Moment Invariant*

1. INTRODUCTION

Communication is an important element in society. Communication is used to pass information from a person to another person or a person to a group of people or community. People can understand each other with good communication. Unfortunately, many people in this world don't have the ability to fully communicate to other people such as deaf and mute. They have a limitation on verbal communication. They use sign language as a means to communicate.

Sign language has different standards and forms based on the country itself. In Indonesia, there are two sign language standards that are mostly used by society, which are Bahasa Isyarat Indonesia (BISINDO) and Sistem Isyarat Bahasa Indonesia (SIBI). Deaf people can communicate with each other using one of two sign language standards. Nevertheless, many people in society do not understand the sign languages in standard forms [1]. This matter has been another problem from then until now.

As technology advances rapidly in the past few years, many researchers have found many means to overcome the problem of communication between deaf and mute using sign language to people who do not understand or have knowledge of sign languages. They created a sign language interpreter system. This system can interpret or translate sign languages to texts or voices. In this way, people who do

not have knowledge of sign languages can understand what the deaf and mute spoke or conveyed.

Sign languages interpreter systems are still being developed until now. There are still many shortcomings for a sign languages interpreter system to run perfectly fine. These shortcomings are about the instability and the performance differences. This will create an opportunity for researchers to research on how to develop a fine and stable sign language interpreter system, as well as this paper.

There were many researches about the sign language interpreter system, especially SIBI Sign Languages. Precisely, a sign language interpreter system uses a hand gesture recognition method. There are different approaches to hand gesture recognition, be it the method or the devices used, which many researchers have studied.

R. Angga, *et al* [2] from Politeknik Elektronika Negeri Surabaya, Indonesia. Their research is what we optimize in the following section. Their research uses Myo Armband as a device to translate hand gestures into text. They use all of Myo Armband sensors, which are Accelerometer, Gyroscope, Orientation, Orientation Euler, and Electromyograph. The dataset used in their research is alphabet A-Z based on SIBI sign language. Alphabet A-Z of SIBI sign language consists of 2 dynamic gestures and 24 static gestures. The 2 dynamic gestures are alphabet J and alphabet Z, while the other alphabets are static gestures. The hand gesture recognition in their research uses a method called moment invariant. They use calculation of moment invariant (mean, median, standard deviation, and skewness). Sequentially, Myo Armband captures the hand gesture and sends the raw value of the data to the system. The system calculates the raw data using the statistical method of Moment Invariant. The result of the calculation created a feature, a series of values, which represent the gesture. From all the features, a dataset is created. They apply min-max normalization to guarantee the range of data. At last, they use *k*-Nearest Neighbor classification to classify new gesture data based on the dataset. Their method produced an accuracy of 82.31% by using leave-one-out cross validation.

A. Aditiya, *et al* [3] from Politeknik Elektronika Negeri Surabaya, Indonesia. The research developed a SIBI sign languages interpreter system. The device they used is Leap Motion Controller. This device is used as hand gesture recognition. They use all alphabet (A-Z) gestures as a test. They propose a new method using relative coordinate to identify the gesture and use *k*-Nearest Neighbor method as a classifier. The performance of the dataset they created had an accuracy rate of 95.58% by using leave-one-out cross validation.

M. Suresh Anand, *et al* [4] from Anna University, India. They study recognition and translation of sign language. They used Indian Sign Languages as gestures which will be recognized and translated. The result of recognition and translation is readable text or a hearable audio. The device they used is a camera. The disabled person gesture of sign language captured by camera and the image is processed by the system. They implement Support Vector Machine (SVM) in their system as a method of hand gesture recognition. Precisely, they use edge detection on image.

M. M. Chandra, *et al* [5] from National Institute of Technology, India. They study sign language to speech conversion. The sign language which they used to test their system prototype is American Sign Language (ASL) and Indian Sign Language (ISL). They use glove with embedded flex sensor, gyroscopes, and accelerometer sensor, which is also included in Myo Armband except flex sensor, to capture gesture data. The data is processed by signal processing and classified using Support Vector Machine (SVM) classifier. Their research uses 11 gesture in ISL and 22 gesture in ASL and produces 100% accuracy for ISL database and 98.91% accuracy for ASL database which both of these databases tested with 25% test data and 75% training data.

The research in this paper is about an optimization of previous sign language interpretation systems. The previous sign language interpreter system using SIBI as a sign language which will be translated. The hand gesture recognition used is moment invariant feature extraction. The research uses one person as a sample. The dataset used is alphabet A-Z which consists of 2 dynamic gestures (alphabet J and Z) and 24 static gestures (all alphabet except J and Z). On the other hand, the classification method used in the research is k -NN [2]. From this point on, the optimization proposed in this research from the previous research are the optimization method for hand gesture recognition on Myo Armband data, as well as an addition of sample data and variation of gestures, and also the optimization of classification method. This method can improve the accuracy of hand gesture recognition and differentiate more gestures. This research uses 20 persons as sample data while the previous research uses only 1 person. Therefore, the previous research is user-dependent, while this research is user-independent. The gesture data captured is using alphabet A to Z and addition of 26 words consists of static gestures and dynamic gestures. In this research, there is addition of filter and calculation in raw data captured by Myo Armband. In EMG raw data, we apply a rectification filter to remove the negative values. In orientation raw data, we apply relative orientation calculation to make the data robust to user orientation position in the real world. The last improvement, we used Support Vector Machine as a classifier while the previous work used k -Nearest Neighbor. As a test, we use leave-one-out cross validation to calculate the performance of the dataset. The result will be analysed to find out the extent of this improvement on the system.

2. RESEARCH METHODOLOGY

The System design divided into 3 parts, Input, Process, and Output, as shown in figure 1.

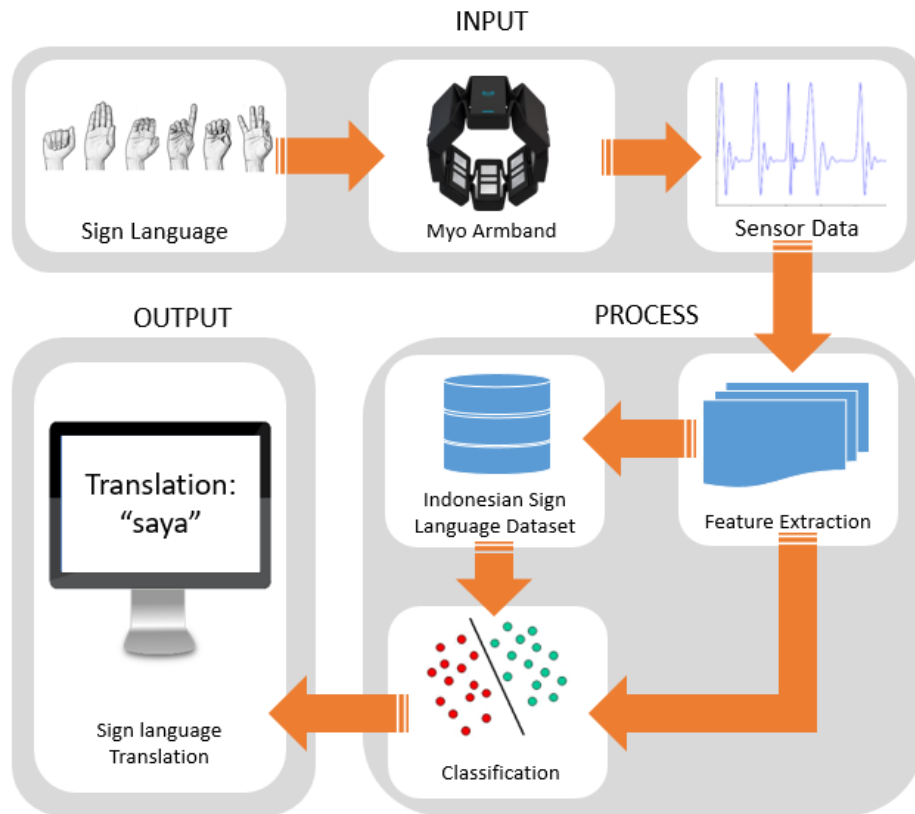


Figure 1. System Design

2.1. Input

In Input, the user wore Myo Armband and did a gesture of sign language. The gesture captured by Myo Armband device sensors resulting in a time-series of data with predetermined time $t=1$. In a second, Myo captured about 49-51 frames which each frame consists of raw data. A frame of raw data captured were accelerometer data, Gyroscope data, Orientation data, Orientation Euler data, and Electromyograph (EMG) data. There were 21 value features in 1 frame as shown in table 1.

Table 1. Raw data values

No	Data	Value Feature
1	Accelerometer	X
2		Y
3		Z
4	Gyroscope	X
5		Y
6		Z
7	Orientation	X
8		Y
9		Z
10		W
No	Data	Value Feature

11	Orientation Euler	Roll
12		Pitch
13		Yaw
14	EMG	1
15		2
16		3
17		4
18		5
19		6
20		7
21		8

These 21 values multiplied by the number of frames captured would be passed to the computer system as shown in Process diagram in figure 1.

2.2. Process

Process diagram consists of 3 parts, they are Feature Extraction, Dataset, and Classification. Feature extraction is the main process of the system which consists of several sub processes as shown in figure 2.

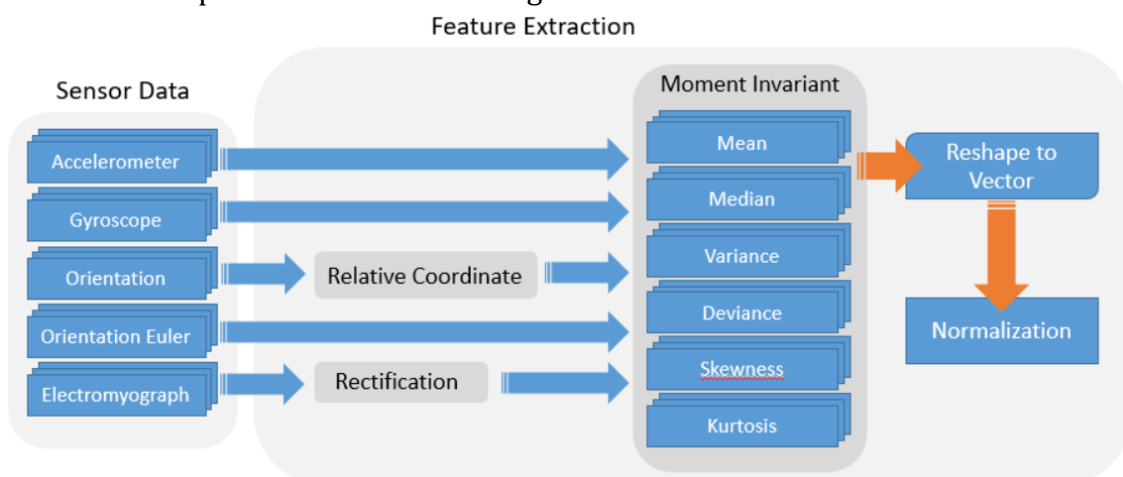


Figure 2. Feature extraction sub processes

Feature extraction is a process to create a vector feature from raw sensor data to obtain valuable data to be processed in the next stage [6]. The captured sensor data calculated by Moment Invariant method, except orientation and electromyograph which applied some filters first. Orientation data in Myo produce data of quaternion (X, Y, Z, W) [7]. This orientation was processed by relative orientation. By using relative orientation, the value of orientation was robust to wherever the user faces or world orientation. But first, we must determine the origin value of orientation. To calculate origin, we use the following formula:

$$Q_{origin} = Q_{identity} * Q_{value}^{-1} \quad (1)$$

Where Q_{origin} = quaternion origin value, $Q_{identity}$ = quaternion identity (0, 0, 0, 1), and Q_{value}^{-1} = inverse of quaternion data. Then, we applied the relative orientation formula with the following formula:

$$Q_{relative} = Q_{origin} * Q_{value} \quad (2)$$

Where $Q_{relative}$ = quaternion relative orientation value, Q_{origin} = quaternion origin value, Q_{value} = quaternion value of raw data.

EMG data processed by rectification filter. Rectification filter was used to remove the negative value from EMG data. This was necessary to know when the muscle reached maximal power, the data would represent it as positive, not negative. There were two kinds of rectification, Half-wave rectification and Full-wave rectification. The example of rectification shown in figure 3.

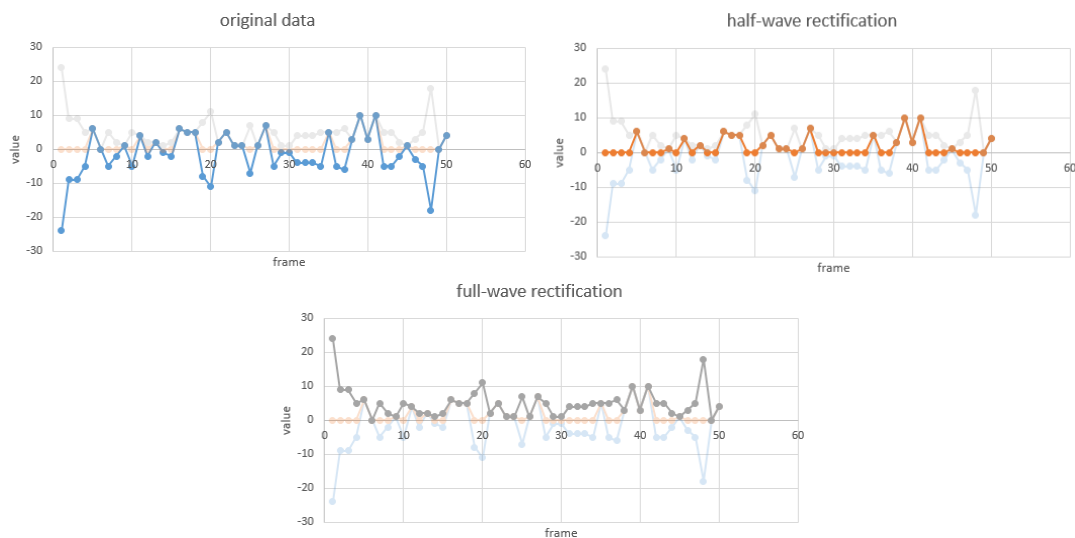


Figure 3. Half-wave and Full-wave rectification on one of EMG data

Half-wave rectification excludes all negative values from data. In this case, the negative value becomes 0. Full-wave rectification converted all negative values to positive by using absolute calculation. Full-wave rectification was more recommended because there was no missing value in EMG data. [8]

After orientation and EMG data applied some filter of its own, all the data ready to calculate by Moment invariant method. Moment invariant method consists of Mean, Median, Variance, Standard Deviation, Skewness, and Kurtosis. In this research, we added Kurtosis as part of the moment invariant method which excluded in previous research. The purpose of moment invariant method is to extract unique information from the time-series data to be able to create feature data which could recognize the difference between or among classes [9]. Mean is one of the most common formula, with the following formula:

$$Mean(x) = \frac{1}{N} \sum x_i \quad (3)$$

Where x_i = all sum of sample data and N = number of sample data. Median is the middle value of ascending-ordered sample data. Variance and standard deviation represent variability and extent around sample value, with the following formula:

$$Variance(x_0, x_1, \dots, x_{N-1}) = \frac{1}{N-1} \sum_{i=0}^{N-1} (x_i - Mean(x))^2 \quad (4)$$

$$\sigma(x_0, x_1, \dots, x_{N-1}) = \sqrt{Variance(x_0, x_1, \dots, x_{N-1})} \quad (5)$$

Where standard deviation is square root of variance. Skewness formula represents the distribution of the data skewed to, whether it is skewed to the right, symmetric, or skewed to the left. The example of skewness shown in figure 4.

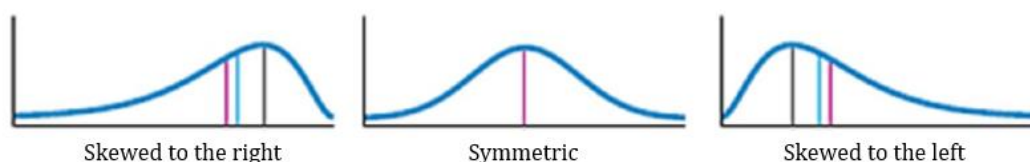


Figure 4. Skewness characteristic on data

Skewness has the following formula:

$$Skewness(x_0, \dots, x_{N-1}) = \frac{1}{N} \sum_{i=0}^{N-1} \left[\frac{x_i - Mean(x)}{\sigma} \right]^3 \quad (6)$$

The last formula of moment invariant is Kurtosis. We added Kurtosis calculation in this research, which had the following formula:

$$Kurtosis(x_0, \dots, x_{N-1}) = \frac{1}{N} \sum_{i=0}^{N-1} \left[\frac{x_i - Mean(x)}{\sigma} \right]^4 \quad (7)$$

Kurtosis represents the flatness or sharpness of the peak of data distribution which is shown in figure 5. The flatter the peak of data distribution, the more negative the value of kurtosis is, while the sharper the peak of data distribution, the more positive the value of kurtosis is.

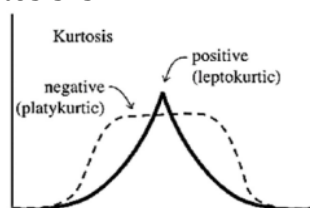


Figure 5. Kurtosis example in data distribution

Every data series of value features of accelerometer data until EMG data calculated by Mean, Median, Variance, Standard Deviation, Skewness, and Kurtosis sequentially. The calculation illustration shown in figure 6.

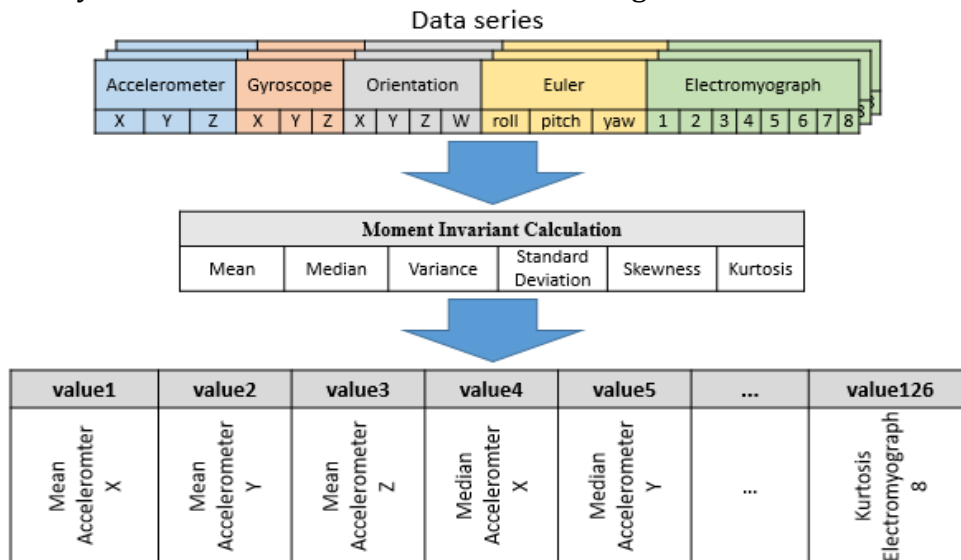


Figure 6. Data series and moment invariant calculation illustration

The result of the calculation would be a series of values with a total of 126 values which represent a gesture feature. This gesture feature was used to create either a dataset or data test for classification.

Dataset created by saving several gesture features. Each gesture feature represents a certain sign language hand gesture. The model of the dataset shown in table 2.

Table 2. Dataset model

No	Label	Gesture Feature				
		Value1	Value2	Value3	...	Value126
1	Gesture 1	value	value	value	...	value
2	Gesture 2	value	value	value	...	value
3	Gesture 3	value	value	value	...	value
...
N	Gesture 4	value	value	value	...	value

After the creation of the dataset, we applied normalization on the dataset. We use min-max normalization to scale data to a certain range [10]. Min-max normalization defined as follow:

$$newdata = \left(\frac{data - min}{max - min} \right) * (newmax - newmin) + newmin \quad (9)$$

Where *newdata* = value of normalization result, *max* = maximum value of data in the columns, *min* = minimum value of data in the columns, *newmax* = maximum limit of new value, *newmin* = minimum limit of new value, and *data* = current sample data. This normalization was also used on gesture data before it was classified based on the dataset. On the classification, we proposed Support Vector Machine (SVM) as a classifier. SVM is one of many machine learning methods. SVM mapped data into certain classes or categories which have been previously defined. This is also known as a supervised learning system. With SVM, the new gesture sample, which did not have a label or class, could be classified based on the dataset and then this new gesture sample label or class could be determined [11]. This label or class of the new gesture sample displayed in the output system.

4.3. Output

The output of the system is a text. This text displays the classification result of the new gesture sample from the system. The system would show the label of the new gesture sample which had similarity to a gesture on the dataset. To test the system performance, we used leave-one-out cross validation. This validation divided the dataset into 2 groups of data, testing data and training data. Testing data used as test samples which were classified based on training data. The result of this validation was estimating the accuracy of the gesture feature created by hand gesture recognition and classified by SVM.

3. RESULT AND DISCUSSION

In this research, we recorded hand gestures of SIBI sign language. The gestures recorded were 26 alphabet A-Z gestures and 26 words which both had a mixture of static and dynamic gestures shown in table 3 [12].

Table 3. List of gestures

Static Gestures			Dynamic Gestures		
A	B	C	J	Z	Ayah
D	E	F	Anak	Nenek	Paman
G	H	I	Suami	Kamu	Mereka
K	L	M	Siapa	Apa	Mengapa
N	O	P	Hai	Halo	Makan
Q	R	S	Minum	Bagus	Bau
T	U	V	Biru	Oranye	Putih
Static Gestures			Dynamic Gestures		
W	X	Y	Pesawat	Ular	Anjing
Ibu	Istri	Saya	Cacing		

We recorded SIBI hand gestures from 20 people. Each gesture was recorded 5 times to get the variations of a gesture from a person. As a result, the number of gestures recorded from a person were 260 gestures (26 gesture * 5 times). By the end of the gesture recording section, we got a total of 5200 gestures from 20 people.

These gestures were saved as a dataset and normalized by min-max normalization with a range of 0-1.

We did some tests in this research. We put some different conditions on the dataset to know the difference between the modifications we proposed and the previous system. First, we tried to compare with just one person sample of 260 gestures. By using leave-one-out cross validation, the dataset accuracy validation performance results shown on table 4.

Table 4. Performance comparison on one person data

Dataset	Dataset Accuracy Validation Performances (%)			
	Without Kurtosis & <i>k</i> -NN Classifier	Without Kurtosis & SVM Classifier	With Kurtosis & <i>k</i> -NN Classifier	With Kurtosis & SVM Classifier
Without Filter	85.77%	96.54%	77.31%	90.77%
With Fullwave Rectification	85.38%	91.54%	77.69%	88.46%
With Relative Orientation	85.77%	94.23%	74.62%	88.85%
With Relative Orientation & Fullwave Rectification	82.31%	89.62%	73.08%	86.15%

The highest accuracy performance owned by the dataset without filter. This result means that there was no increase in accuracy performance by filtering the raw data of Myo sensors before Moment Invariant calculation. Instead, some dataset had decreased accuracy performance. Therefore, it was enough just by using Moment Invariant. This made Moment Invariant calculation results robust on raw data series of Myo sensors. On the other hand, we test the dataset by adding Kurtosis calculation on Moment Invariant or not. The results yield that the dataset without Kurtosis on Moment Invariant had higher accuracy performance than the dataset with Kurtosis included. It means that the addition of Kurtosis calculation on raw data series of Myo sensors was not needed. By using just Mean, Median, Variance, Standard Deviation, and Skewness on Moment Invariant calculation were enough to give better accuracy performance. Afterwards, we compared the proposed classification method, Support Vector Machine classifier, with previous work classification method, *k*-Nearest Neighbor (*k*-NN). From the results on table 4, the accuracy differences on both classifiers was quite significant. The accuracy result by SVM classifier gave quite a significant increase to the accuracy result by *k*-NN classifier in all test cases. It means that the SVM classifiers were suitable for the numerical feature dataset produced by Moment Invariant calculation. At last, we tested all 20 person sample dataset which had a total of 5200 gestures. In this test, we also compared the dataset with Kurtosis included and the dataset with Kurtosis

excluded on Moment Invariant and compared the classification method of k -NN and SVM. The results shown in table 5.

Table 5. Performance comparison on 20 person data

Dataset	Accuracy Validation Performance on Classifiers (%)	
	k -NN	SVM
With Kurtosis calculation	78.58%	85.12%
Without Kurtosis calculation	82.40%	86.75%

With all 20 person sample dataset (5200 gestures), the dataset without Kurtosis calculation had a higher accuracy performance than the dataset with Kurtosis calculation. The classification using SVM classifier also produced higher accuracy performance than the classification using k -NN classifier. The highest accuracy, with a value of 86.75%, achieved by the dataset without Kurtosis calculation and classified by SVM classifier. The accuracy difference gap between these two classifiers on all 20 person sample dataset was not as much on one person dataset. It means that, in this case, SVM classifiers are best used on fewer sample dataset. The more the sample of dataset, the less the gap range of the accuracy performance on the both of classifiers. Nonetheless, SVM classifier had higher compatibility on the dataset calculated by Moment Invariant and had higher accuracy performance than k -NN classifier.

4. CONCLUSION

According to the experiment results, we conclude that the Support Vector Machine classifier produced higher accuracy performance than k -NN classifier. SVM classifier had higher compatibility on the dataset calculated by Moment Invariant. Moment Invariant used calculation of mean, median, variance, standard deviation, skewness, and kurtosis on a series of data. Moment invariant used as feature extraction on hand gesture recognition which was a powerful method in this case of research. Moment Invariant extracts the unique information from time-series data produced by Myo sensors such as Accelerometer, Gyroscope, Orientation, Orientation Euler, and EMG which each of the sensors had different ranges of value. Moment Invariant calculation was robust on raw data sensor of Myo Armband. We could use just the Moment Invariant method on feature extraction of hand gesture recognition without adding filters on raw data of Myo sensors. However, not all Moment Invariant calculations included could give the best result of feature extraction. Excluding Kurtosis calculation proved that the accuracy performance of the gesture feature produced higher results than including Kurtosis calculation.

REFERENCES

- [1] 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," in *International Conference On Special Education in South East Asia Region*, 2017, pp. 217–223.
- [2] A. Rahagiyanto, A. Basuki, and R. Sigit, "Moment Invariant Features Extraction for Hand Gesture Recognition of Sign Language based on SIBI," *Emit. Int. J. Eng. Technol.*, vol. 5, no. 1, pp. 119–138, Jul. 2017, doi: 10.24003/emitter.v5i1.173.
- [3] A. Anwar, A. Basuki, R. Sigit, A. Rahagiyanto, and M. Zikky, "Feature Extraction for Indonesian Sign Language (SIBI) Using Leap Motion Controller," 2018, doi: 10.1109/ICSEC.2017.8443926.
- [4] S. A. M., M. N., M. S., S. P., V. K. K., and K. A., "Recognition and Translation of Sign Language Using SVM," *Adv. Nat. Appl. Sci.*, vol. 10, no. 5, pp. 180–185, 2016.
- [5] M. M. Chandra, S. Rajkumar, and L. S. Kumar, "Sign Languages to Speech Conversion Prototype using the SVM Classifier," in *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, Oct. 2019, vol. 2019-October, pp. 1803–1807, doi: 10.1109/TENCON.2019.8929356.
- [6] X. Wang and K. K. Paliwal, "Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition," *Pattern Recognit.*, vol. 36, no. 10, pp. 2429–2439, Oct. 2003, doi: 10.1016/S0031-3203(03)00044-X.
- [7] E. Fresk and G. Nikolakopoulos, *Full Quaternion Based Attitude Control for a Quadrotor*. 2013.
- [8] L. Ricardo, J. Luiz, M. Bigliassi, T. F. Dias Kanthack, A. C. de Moraes, and T. Abrao, "Influence of Different Strategies of Treatment Muscle Contraction and Relaxation Phases on EMG Signal Processing and Analysis During Cyclic Exercise," in *Computational Intelligence in Electromyography Analysis - A Perspective on Current Applications and Future Challenges*, InTech, 2012.
- [9] J. Flusser, T. Suk, and B. Zitová, *Moments and Moment Invariants in Pattern Recognition*. Chichester, UK: John Wiley and Sons, 2009.
- [10] S. G. K. Patro and K. K. sahu, "Normalization: A Preprocessing Stage," *IARJSET*, pp. 20–22, Mar. 2015, doi: 10.17148/iarjset.2015.2305.
- [11] M. Awad, R. Khanna, M. Awad, and R. Khanna, "Support Vector Machines for Classification," in *Efficient Learning Machines*, Apress, 2015, pp. 39–66.
- [12] Departemen Pendidikan Nasional Indonesia, *Kamus Sistem Isyarat Bahasa Indonesia*, 3rd ed. Jakarta: Departemen Pendidikan Nasional, 2001.