OPEN ACCESS
International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-5, January 2020

Optimization of Modified Hidden Markov Model for Vision-Based Indonesian Sign Languages Recognition

Suharjito, Herman Gunawan, Narada Thiracitta

Abstract: ***Sign language is one of the most popular language which is used as a communication bridge that depends on hand movement. As it is used worldwide, the variety of sign languages is very large and most of the people doesn't understand it which makes the communication between deaf and normal people interrupted. This makes sign language recognition popular as it makes people doesn't need to understand the sign but still understand the meaning of the sign. But sign language itself has many problems such as the possibility of different dataset has the same movement but different meaning, the method used of each dataset could be ineffective in other dataset, and many other else which makes it difficult to be implemented. Other than that, vision-based recognition is not as popular as sensor-based recognition because of the difference in feature accuracy even though it could give more area of improvement. That's why the aim of this paper is to presents the combination of methods used to recognize vision-based Indonesian Sign Language and enhance the method using optimization technique. The methodology used in this study follows four steps framework of sign language recognition which is dataset collection, preprocessing, feature extraction, and recognition. Each method is improved in order to be compared and checked what is the method combination and optimization method is the best for sign language recognition. For dataset collection, the dataset that is used is formal Indonesian Sign Language which is called Sistem Isyarat Bahasa Indonesia (SIBI) with some constraint in order to make sure the quality of the dataset is good. The preprocessing methods are differentiated into two categories which is image enhancement and hand detection. Image enhancement methods include no image enhancement, Gaussian blur, Bilateral blur, and Contrast Limited Adaptive Histogram Equalization and hand detection methods include skin detection using YCbCr color space and edge detection using Canny algorithm. Feature that is used in this study is pair of left and right-side movement for each frame that is extracted by calculating the average value of each pixels from left and right-side image that could be a representative value of each frame. Lastly the recognition system that is used is Gaussian Hidden Markov Model and its' state optimization which includes no optimization, Latin Hypercube optimization, Hill Climbing optimization, and Combination of Latin Hypercube and Hill Climbing optimization. The experiment result of the proposed method could recognize up to 82% accuracy rate. The improvement of this research could be implemented on educational studies or game development that need vision-based sign language recognition system.***

Keywords: ***Hidden Markov Model, Indonesian Sign Language, Optimization, Recognition***

Manuscript published on January 30, 2020.
* Correspondence Author

Suharjito*, Computer Science Department, Binus Online Learning Program, Bina Nusantara University, Jakarta, Indonesia. Email: suharjito@binus.edu

Herman Gunawan, Computer Science Department, Binus Graduate Program – Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia. Email: herman.gunawan@binus.ac.id

Narada Thiracitta, Computer Science Department, Binus Graduate Program – Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia. Email: narada.thiracitta@binus.ac.id

I. INTRODUCTION

In the last decade, the development in automatic sign language recognition system has become a popular topic in pattern recognition. Not only does it have applications on communication purpose, sign language recognition has also implemented in e-learning system, smart system, and game development. However, the task of sign language recognition from a video is extremely challenging because of the large amount of sign language, individual style in performing the sign, and external factors such as environment, noise, and many other else.

During the last few years, the field of sign language recognition got enormous development. It is caused by the availability of many public dataset which could be used for sign language recognition, the development of computer vision field which lead to more accurate features, and the improvement of artificial intelligence field especially on machine learning and deep learning. With this development, there are researchers that review these studies to make it easier to be compared and improved [1-2].While the result is indeed improved, the improvement of vision-based recognition is still far behind sensor-based recognition because sensor-based recognition gives far more accurate feature using additional equipment like Kinect [3]-[6], Leap motion [4], and Myo armband [7]. Consequently, vision-based recognition remains as a distant solution for actual application.

On the other hand, sensor-based recognition is very unlikely to be implemented in daily life because it cost a lot to get an additional hardware while vision-based recognition only need video as an input, so it is easier to be implemented. Because of the simplicity of vision-based recognition, there are many studies that implement vision-based recognition with different method.

All of these methods give birth to a fact that there are four main components which is used in vision-based sign language recognition, which are hand configuration, place of articulation, hand orientation, and hand movement [8].

It also gives a fact that there are two main ways for the machine to recognize the sign, which is machine learning and deep learning. While deep learning is good in video recognition, it usually needs large amount of dataset and the complexity of the computation is far higher than machine learning which makes it need far better hardware. That is why machine learning is more applicable to be used in the system of sign language recognition.

As sign language recognition is also a type of sequence recognition, machine learning method which is very good in sequence recognition is Hidden Markov Model, which is also one of the most popular techniques in sign language recognition studies. While it is popular, machine learning model can't automatically train and classify a video. This condition makes a framework for sign language which consists of four sections, which is dataset collection, preprocessing, feature extraction, and modelling. This framework has been fulfilled in previous studies with many different methods and different dataset which makes it difficult to get a fair comparison because the same method used in different dataset affect the result.

In this paper, we proposed to use the combination of methods and its optimization to fulfill the sign language recognition framework. it shows the result of each combination with the same dataset which gives a fair comparison between each method. The contribution of the study lies in two points. (1) we used state optimization method for modified Hidden Markov Model. (2) We used combination of methods that's been performing well in vision or sensor-based sign language recognition on vision based Indonesian sign languages. (3) We present comparison of each method by implementing sign language application developed using computer vision techniques which could improve the performance of sign language recognition. The previous researches which relate to this research are explained in Section 2. Proposed system and its modules are explained in Section 3. Result and evaluation are presented in Section 4. Eventually conclusion of the study is presented in Section 5.

II. RELATED WORKS

Sign language recognition which is a type of sequence recognition in general could be differentiating into four sections which is dataset collection, preprocessing, feature extraction, and modelling. In this section, we present the development of each section that has been designed.

A. Dataset Collection

Dataset collection is the first section of sign language recognition framework. In dataset collection, there are two type of dataset that is used in the studies, which are vision-based data and sensor-based data. Sensor-based data need additional equipment which is used to get more accurate feature to be used in the researches, like Kinect [3]-[6], Leap Motion [4], Myo Armband [7]. On the other hand, vision-based data consists of video which is used as an input in sign language recognition to make it easier to be implemented.

Other than the type of dataset, each country has its own sign language, means that there are many different datasets that has been used in the studies. While it is certainty that each sign of a single dataset is unique, the complexity of each dataset is not in the same level. It makes the comparability of different dataset with the same methods is not possible. Even if it is sign language with a same country, if the dataset is not the same, it is still not comparable because the sign, signer, environment, and external factor are different. There are many studies which has been done for example American Sign Language [3, 5], Indian Sign Language [4, 9], Arabic Sign Language [10] where the result can't be compared with other result.

The amount of dataset and classes also affect the result of the sign language recognition. The higher the amounts of classes' means the harder the system is going to recognize the sign. The studies on sign language is very vast with different amount of dataset and classes, for example on Indian Sign Language with 25 classes on Kinect dataset got an accuracy of 90.8% [4], American Sign Language with 30 classes got an accuracy of 89% [8], Indian Sign Language with 5 classes with an accuracy of 80% [10].

Dataset collection could also be differentiated into two categories based in the source, primary dataset dan secondary dataset. Primary dataset is a dataset where the researcher gets the dataset themselves, in other word recording the sign themselves. It is very good in order to set the environment, consistency, and other external factor that could affect the result of the sign language recognition. Some researcher uses it to set the dataset based on the proposed method for example the use of two depth sensors which are Kinect and Leap Motion to make multi-sensor fusion framework for sign language recognition [4]. Secondary dataset means that the researcher uses public dataset which is shared by certain organization which is very good in order to get the comparable result without duplicating the result of other studies. The example secondary studies on sign languages recognition field is RWTH-BOSTON-50 database which is an American Sign Language dataset that was used because of it's realistic with high number of classes and data [8]. Another example of secondary studies on sign language field is LSA64 which provide 3200 Argentina Sign Language Video with colored gloves [11, 12].

B. Preprocessing

Sign language recognition is a system to recognition a video. As a video, there must be noise that doesn't give any information. That's why there is a need for preprocessing method. To ensure the consistency and uniqueness for each sign of the dataset, there is a need to enhance the data, so the characteristic of the data is not gone. Other than that, in order for the video could be changes into numerical value in feature extraction, there is a need to prepare the video to be able to be calculated in the next section. In sign language, the characteristics which are usually used are hands movement and position [8].

Preprocessing method in the studies has been very varied with the similar target in mind. Some researcher used Principal Component Analysis (PCA) and skin color detection to detect face and hands.

He also uses kurtosis position to detect inaccurate hand detection and motion chain code to represent hand trajectory [8]. There are also some that used skin colors to extracts face and hands region blobs to differentiate it. On the other hand, sensor-based give more accurate data such as Kinect give skeletal data [3]-[5], Myo armband that gives hand and finger movement called electromyography data [7], and Leap motion to sense finger and hand movement in smaller field of view [4, 9] so there is no need to preprocess the data.

C. Feature Extraction

To get the features which are needed, there is a need to know what type of feature is unique in the dataset. in sign language, some study used hand geometry shape of a hand for each sign using Kalman filter [13], electromyography and inertial measurement data using Myo armband [7], stated of hand with different point of view with Leap Motion [14], hand and fingertip position using Kinect and Leap Motion [4]. There are many ways to get it with their own accurate rate, but most of the research used hand position and movement as features in sign languages recognition.

D. Training and Classification

As a video recognition, there is a need to make a model that could represent the sign language recognition based of the extracted features. Hidden Markov Model is the most popular method that is used in sign language recognition because it can process time series data with temporal length [8]. As the development of sign language recognition field, Hidden Markov Model also got optimized along with the studies. Some studies that implement it are vision-based recognition with Hidden Markov Model on American Sign Language with an accuracy of 89% [8], sensor-based recognition with the combination of Low rank approximation to determine the number of hidden states in Hidden Markov Model which is called Light Hidden Markov Model with an accuracy of 83.6% [3], sensor-based recognition with combination of Gaussian Mixture Model and Hidden Markov Model called Gaussian Mixture Model Hidden Markov Model on American Sign Language with an accuracy of 96.15% [7], and sensor-based recognition by fusing two sensors with asynchronous and temporal dependencies between two different channel which is called Coupled Hidden Markov Model on Indian Sign Languages with an accuracy of 90.8% [4].

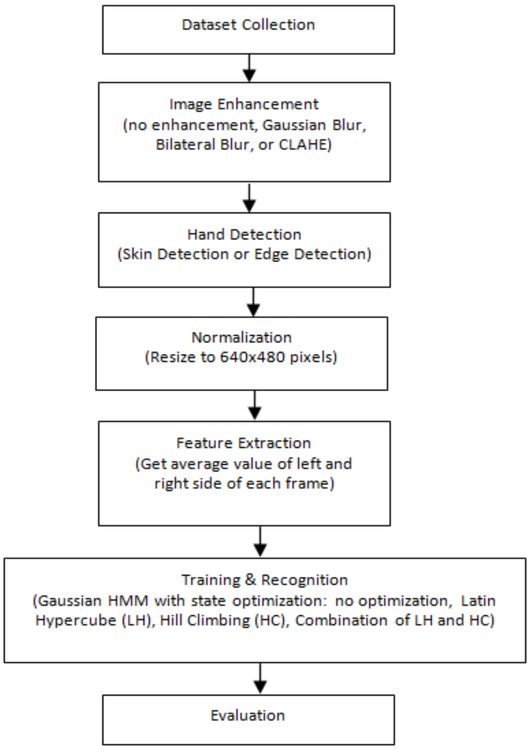

Fig. 1.Experimental design on sign language recognition

Other than that, there is also some researcher that has tried sign language recognition on Indonesia Sign Language. On 2013, researchers use flex sensors and accelerometer sensor with deep learning method called Adaptive Neighborhood based Modified Backpropagation with histogram feature extraction with an accuracy of 91.6% [15]. Another one on 2014 that use vision-based recognition with HSV skin detection, contour detection using region of interest, and image matching using speeded up robust features with an accuracy of 63% [16]. Then on 2018, sensor-based recognition using Kinect by using Hidden Markov Model with some number of hidden states with K-Fold cross validation with and accuracy of 70% has been achieved.

III. EXPERIMENTAL DESIGN

In this study, we also used four section of method to be used on sign language recognition which is Dataset Collection, Preprocessing, Feature Extraction, and Training & Classification. Overview of the experimental design could be seen on Fig. 1. and the detail of each section is presented below.

A. Dataset Collection

In sign language, the main target on dataset is consistency and uniqueness of the data. In order to get it, we used primary dataset that is taken on Bina Nusantara University in 2018/2019. The dataset consists of Indonesian Sign Language known as Sistem Isyarat Bahasa Indonesia which is usually called SiBi. The dataset consists of 15 signs with 2 signers and 10 iterations for each sign with a total of 300 video. The duration range of video is variants with the range of 1 to 4 seconds. The frame range of the video is also variants with the range of 51 to 117 frames. The dataset is collected using Samsung Galaxy S6 Camera, 16 megapixels Sony Exmor RS IMX240 sensor.

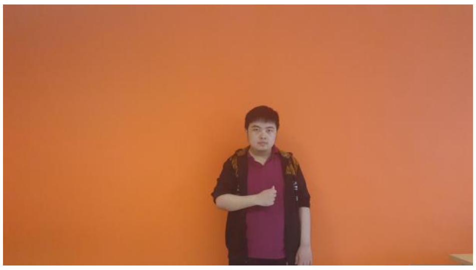

Fig. 2.Sample of a frame in dataset

On dataset collection, there are some restrictions on environmental area to make sure the dataset is valid and accurate. In this case, in order to differentiate hand and other object, there is a need to make the environment to have different color with the hands. Other than that, there is also a need to set aside other object to minimize noises in the dataset. And lastly, to get the movement of the right and left hand, there is a need for the signer to perform the sign at the middle of the video. The sample of dataset could be seen on Fig. 2.

B. Preprocessing

The target of the preprocessing technique is to enhance the video and prepare the video to change it into numerical value in feature extraction. For image enhancement, we used popular technique such as gaussian blur, bilateral blur, and contrast limited adaptive histogram equalization while for video extraction preparation, we used skin detection using YCbCr Color Space as it is better than HSV color space for skin detection [17] and edge detection using Canny edge detection as it is the most popular technique for edge detection. We also try to not use image enhancement as a comparison whether it is needed or not. Gaussian blur is used as the most popular image enhancement technique, bilateral blur is used because it gives more accuracy line which could affect the result in edge detection, and contrast limited adaptive histogram equalization is used because is it the advanced enhancement technique which is used in recent medical research and fog clearing algorithm. Sample output after preprocessing could be seen on Fig. 3.

Fig. 3.Preprocessing result using contrast limited adaptive histogram equalization and skin detection

C. Feature Extraction

Feature that is used in sign language recognition is mainly hand movement. To get the movement, the result of the preprocessing already gives a list of black and white images. Then to change it into numerical value, we calculate the center point of left side and the center point of right side. It changes the list of images into list of a pair of numbers. Each pair of numbers represents a frame and each number represents the left hand and right-hand location.

D. Training and Classification

For training and classification, we used Gaussian Hidden Markov Model which gives highest accuracy rate on sensor-based recognition by previous researcher. Gaussian Hidden Markov Model is very good in sign language recognition because the possibility of each frame is not a definite value. We make the model based on how many signs are needed to be trained and classified means that in this case there is 15 models. Each model will represent a sign to be trained. Then test video is inserted to be classified for each model. Each model will give an error value where smaller error value means higher similarity rate so model that got smallest error value is the sign. There is also hidden states optimization by using Latin hypercube, hill climbing, and the combination between Latin hypercube and hill climbing. We also used four hidden states which were shown as the best number of hidden states on previous study as a comparison [7].

Table- I: Detail of combination used on the experiment

No	Image Enhancement	Hand Detection	Optimization states of HMM
1	No Enhancement	Skin Detection	No Optimization
2	No Enhancement	Skin Detection	Latin Hypercube (LH)
3	No Enhancement	Skin Detection	Hill Climbing (HC)
4	No Enhancement	Skin Detection	Combination of LH and HC
5	No Enhancement	Edge Detection	No Optimization
6	No Enhancement	Edge Detection	Latin Hypercube (LH)
7	No Enhancement	Edge Detection	Hill Climbing (HC)
8	No Enhancement	Edge Detection	Combination of LH and HC

9	Gaussian Blur	Skin Detection	No Optimization
10	Gaussian Blur	Skin Detection	Latin Hypercube (LH)
11	Gaussian Blur	Skin Detection	Hill Climbing (HC)
12	Gaussian Blur	Skin Detection	Combination of LH and HC
13	Gaussian Blur	Edge Detection	No Optimization
14	Gaussian Blur	Edge Detection	Latin Hypercube (LH)
15	Gaussian Blur	Edge Detection	Hill Climbing (HC)
16	Gaussian Blur	Edge Detection	Combination of LH and HC
17	Bilateral Blur	Skin Detection	No Optimization
18	Bilateral Blur	Skin Detection	Latin Hypercube (LH)
19	Bilateral Blur	Skin Detection	Hill Climbing (HC)
20	Bilateral Blur	Skin Detection	Combination of LH and HC
21	Bilateral Blur	Edge Detection	No Optimization
22	Bilateral Blur	Edge Detection	Latin Hypercube (LH)
23	Bilateral Blur	Edge Detection	Hill Climbing (HC)
24	Bilateral Blur	Edge Detection	Combination of LH and HC
25	CLAHE	Skin Detection	No Optimization
26	CLAHE	Skin Detection	Latin Hypercube (LH)
27	CLAHE	Skin Detection	Hill Climbing (HC)
28	CLAHE	Skin Detection	Combination of LH and HC
29	CLAHE	Edge Detection	No Optimization
30	CLAHE	Edge Detection	Latin Hypercube (LH)
31	CLAHE	Edge Detection	Hill Climbing (HC)
32	CLAHE	Edge Detection	Combination of LH and HC

IV. EXPERIMENTAL RESULT AND DISCUSSION

A. Experiment Environment

The experiment is conducted on Intel(R) Core™ i3-4030U CPU @1.90GHz (4 CPUs), ~1.9GHz with 4 GB RAM using Python 3.6 version. The dataset that is used in this study is Indonesian Sign Language known as SIBI. As there are not any SIBI public dataset yet, there is a need to do dataset collection. In order to makes the dataset valid and accuracy, some restrictions has been applied. The number of classes that is used is 15 classes which consist of 15 different isolated sign languages. With a total of 300 video, the data is split into training data and testing data with a ratio of 8:2. The splitting data was done manually to make sure that each class has 80% data training and 20% data testing.

B. Result and Discussion

Table- II: Result of each combination

Method	Average Accuracy (%)	Average Time Cost (s)	Method	Average Accuracy (%)	Average Time Cost (s)
1	82	65	17	76	65
2	81	2101	18	78	1891
3	79	1979	19	74	1973
4	82	9490	20	81	9042
5	65	59	21	64	79
6	68	2090	22	66	2310
7	67	1943	23	62	2264
8	69	8964	24	66	9243
9	79	66	25	71	71
10	79	2273	26	74	2498
11	76	2276	27	70	2021
12	80	9347	28	76	8678
13	63	84	29	60	73
14	64	2023	30	62	2134
15	59	2368	31	57	2358
16	63	9142	32	63	9144

As the experiment used many kinds of methods, the result of each method could be compared as an evaluation. In this study, there are a total of 32 combinations of methods which is shown in Table I. On data preprocessing, there are four types of image enhancement which are used, which are no image enhancement, Gaussian Blur, Bilateral Blur, and Contrast Limited Adaptive Histogram Equalization. There are also two types of image preparation to be used on data preprocessing which are Skin Detection and Edge Detection. For skin detection, the one that is used in this study is YCbCr Color Space. After that, there is a feature extraction to differentiate left-side and right-side movement as a feature. Then lastly training and classification method using Gaussian Hidden Markov Model with four types of state optimization, which is using four hidden states which is the best on previous study [7], Latin hypercube method, hill climbing method, and the combination of Latin hypercube and hill climbing method. The experiment runs each combination of method five times in order to get the average value of accuracy and average value of time cost. The detail on each value for each combination could be seen on Table II and the detail on average result of each method could be seen of Table III, Table IV and Table V. From the experiments, the result of the combination gives variant accuracy value between 50s until 80s while the value of time cost depends on the optimization method used. In this case, skin detection method gives better performance in almost any combination rather than edge detection with wide gap. On other hand, for image enhancement method, the result of not using image enhancement is giving the best performance, followed by Gaussian Blur, Bilateral Blur, and Contrast Limited Adaptive Histogram Equalization. Lastly in respect to HMM State Optimization method, the best performance is taken by the combination of Latin Hypercube and Hill Climbing, followed by Latin Hypercube, no enhancement, and Hill Climbing method.

Table- III: Average result based on hand detection method

Hand Detection Method	Average Accuracy (%)
Skin Detection	77.375
Edge Detection	63.5

Table- IV: Average result based on image enhancement method

Image Enhancement Method	Average Accuracy on Skin Detection (%)	Average Accuracy on Edge Detection (%)
No Image Enhancement	81	67.25
Gaussian Blur	78.5	62.25
Bilateral Blur	77.25	64.5
CLAHE	72.75	60.5

Table- V: Average result based on HMM state optimization method

HMM State Optimization Method	Average Accuracy on Skin Detection (%)	Average Time Cost on Skin Detection (s)	Average Accuracy on Edge Detection (%)	Average Time Cost on Edge Detection (s)
No Enhancement	77	66.75	63	73.75
Latin Hypercube	78	2190.75	65	2230.25
Hill Climbing	74.75	2062.25	61.25	2233.25
Combination of Latin Hypercube and Hill Climbing	79.75	9139.25	65.25	9123.25

In this study, skin detection gives far more accurate performance than edge detection. It is caused by the restriction of the dataset that support both methods. For skin detection, the restriction is that the recorded video must avoid any object with the same color as skin which provides very good data. While for edge detection, the restriction is to avoid any object to avoid any useless edges which is noises. But for edge detection, the additional edge couldn't be avoided such as clothes, which makes the feature of edge detection is not very good in hand detection. Other than that, because of the feature extraction method which calculates any white pixels shown after skin or edge detection, the sensitivity of edge detection is far higher than skin detection. It happens because skin detection shows many white pixels so some additional pixels such as some noises doesn't affect much on feature extraction while on edge detection, the pixels shown is only edges which is low in number. This makes additional edge on clothes, face, and other place gives much impact on the features which makes it not as accurate as skin detection.

The reason the performance decrease with image enhancement is because the dataset of the video is already in high quality. As the dataset is recorded manually with some restriction, the result of the dataset is already clear. Other than that, we also check the quality of the data after the recording and the movement of the sign is not fast enough to create disturbing noise which gives the image enhancement method is not useful. In this case, the area detection cause by the image enhancement could be disturbed because it affects the quality of the pixels with its surrounding pixels. In other hand, four hidden state in Gaussian hidden Markov model gives high performance and low time cost as stated in previous study. Latin hypercube is giving slightly higher performance because it takes sample between high number of hidden states and low number of hidden states as comparison so it gives high performance but the time cost is far higher because the time cost for high number of hidden state takes much longer time. Hill climbing method gives low performance because the method relied too much on random value at the start, which makes it unstable and the value is very dependent on the random number. Lastly the combination of Latin hypercube and hill climbing gives highest accuracy because in this case hill climbing doesn't rely on random value, but the sample of value on each area of sample which is going to be compared with other area of sample, which gives high performance with the cost of high time cost.

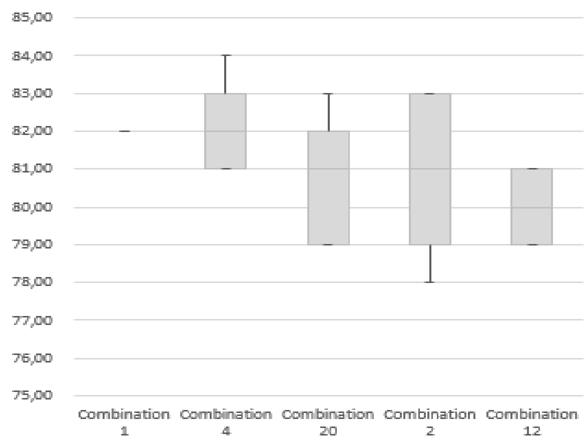

Fig. 4.Boxplot of five best combinations based on average performance in percent (%)

Lastly, Fig. 4 show the boxplot of five best combination of method based on average accuracy value of each combination of method. In the boxplot, the left side means the minimum accuracy value out of five times the program run, left side of the box left box mean the value of quartile 1, line between two box mean the value of quartile 2, right side of the right box mean the value of quartile 3, and right side mean the maximum value out of 5 box. Based on these combinations, no image enhancement takes three out of five best combinations, skin detection takes all five best combinations, and combination of Latin Hypercube and Hill Climbing takes three out of five best combinations which mean these methods are the most stabilized method out of all method. While no image enhancement with skin detection and no HMM optimization gives the same accuracy value with no image enhancement with skin detection and combination of LH & HC HMM optimization, the one with no optimization method gives unstable value to other method that makes its average accuracy low means it is not recommendable to use in most case.

But combination of LH & HC is also not perfect because while it is giving good accuracy value, the time cost of this method takes far more than the other which is a disadvantage of using this method.

V. CONCLUSION AND FUTURE WORKS

In this study, we proposed to used combination of method with its' optimization to be used on vision-based sign language recognition system for Indonesian Sign Language known as SIBI. We used single camera as an input with the signer at the middle of the camera. The proposed system enhanced and prepared the result of the dataset and then the recognition was carried out using modified Gaussian Hidden Markov Model, where the best overall accuracy of 82% shown. In term of accuracy, the optimization using combination of Latin Hypercube and Hill Climbing has the advantages but on the other hand, in term of running time, this method is far higher than other methods. In future, there could be some improvement on the number of features used in this study or by applying other optimization method in order to increase the accuracy and lower the running time.

ACKNOWLEDGMENT

This research was supported by Research Grant No. 036/CR.RTT/IV/2018 from Ministry of Research Technology and Higher Education of Republic of Indonesia.

REFERENCES

1. Suharjito, A. Ricky, F. Wiryana, M. C. Ariesta and G. P. Kusuma, "Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input-Process-Output," Procedia computer science, pp. 441-448, 2017.
2. Suharjito, F. Wiryana, G. P. Kusuma and A. Zahra, "Feature Extraction Methods in Sign Language Recognition System: A Literature Review," 2018 Indonesian Association for Pattern Recognition International Conference (INAPR), pp. 11-15, 2018.
3. H. Wang, X. Chai, Y. Zhou and X. Chen, "Fast sign language recognition benefited from low rank approximation," 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2015.
4. P. Kumar, H. Gauba, P. P. Roy and D. P. Dogra, "Coupled HMM-based multi-sensor data fusion for sign language recognition," Pattern Recognition Letter, pp. 1-8, 2017.
5. J. L. Raheja, M. Minhas, D. Prashanth, T. Shah and A. Chaudary, "Robust gesture recognition using Kinect: A comparison between DTW and HMM," Optik, pp. 1098-1104, 2015.
6. T. Handhika, R. I. M. Zen, Murni, D. P. Lestari and I. Sari, "Gesture recognition for Indonesian Sign Language (BISINDO)," Journal of Physics: Conference Series, 2018.
7. R. Fatmi, S. Rashad, R. Integlia and G. Hutchison, "American Sign Language Recognition using Hidden Markov Models and Wearable Motion Sensors," Trans. MLDM, pp. 41-55, 2017.
8. M. M. Zaki and S. I. Shaheen, "Sign language recognition using a combination of new vision based features," Pattern Recognition Letters, pp. 572-577, 2011.
9. D. Tewari and S. K. Srivastava, "A visual recognition of static hand gestures in indian sign language based on kohonen self-organizing map algorithm," International Journal of Engineering and Advanced Technology (IJEAT), pp. 165-170, 2012.
10. A. Tharwat, T. Gaber, A. E. Hassanien, M. K. Shahin and B. Refaat, "SIFT-Based Arabic Sign Language Recognition System," Afro-european conference for industrial advancement, pp. 359-370, 2015.
11. Suharjito, H. Gunawan, N. Thiracitta and A. Nugroho, "Sign Language Recognition Using Modified Convolutional Neural Network Model," 2018 Indonesian Association for Pattern Recognition International Conference (INAPR), pp. 1-5, 2018.
12. Suharjito, N. Thiracitta, H. Gunawan and G. Witjaksono, "The Comparison of Some Hidden Markov Models for Sign Language Recognition," 2018 Indonesian Association for Pattern Recognition International Conference (INAPR), pp. 6-10, 2018.
13. K. Imagawa, S. Li and S. Igi, "Color-based hands tracking system for sign language recognition," Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition, pp. 462-467, 1998.
14. K. Y. Fok, N. Ganganath, C. T. Cheng and C. K. Tse, "A Real-Time ASL Recognition System Using Leap Motion Sensors," 2015 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, pp. 411-414, 2015.
15. E. Supriyati and M. Iqbal, "Recognition System of Indonesia Sign Language based on Sensor and Artificial Neural Network," Makara Journal of Technology, pp. 25-31, 2013.
16. R. Hartanto, A. Susanti and P. I. Santosa, "Real time static hand gesture recognition system prototype for Indonesian sign language," 2014 6th International Conference on Information Technology and Electrical Engineering (ICITEE), pp. 1-6, 2014.
17. K. B. Shaik, P. Ganesan, V. Kalist, B. S. Sathish and J. M. M. Jenitha, "Comparative Study of Skin Color Detection and Segmentation in HSV and YCbCr Color Space," Procedia Computer Science, pp. 41-48, 2015.

AUTHORS PROFILE

Suharjito, is the Head of Computer Science Department in Binus Online Learning Program of Bina Nusantara University. He received under graduated degree in mathematics from The Faculty of Mathematics and Natural Science in Gadjah Mada University, Yogyakarta, Indonesia in 1994. He received master degree in information technology engineering from Sepuluh November Institute of Technology, Surabaya, Indonesia in 2000. He received the PhD degree in system engineering from the Bogor Agricultural University (IPB), Bogor, Indonesia in 2011. His research interests are intelligent system, Fuzzy system, image processing and software engineering.

Herman Gunawan, is currently a student who studied Computer Science in Bina Nusantara University. He is currently working in Bina Nusantara University as laboratory assistant that focuses on Computer Science subject. His research interest includes artificial intelligence and database related subject.

Narada Thiracitta, is currently a student who studied Computer Science in Bina Nusantara University. He was a laboratory assistant on Bina Nusantara University that focuses on Computer Science subject. His research interest includes artificial intelligence and database related subject.