



# Sign Language Recognition using Principal Component Analysis and Support Vector Machine

Astri Novianty<sup>a,\*</sup>, Fairuz Azmi<sup>b</sup>

<sup>a,b</sup> Department of Computer Engineering, Telkom University, Indonesia

<sup>a</sup> [astrinov@telkomuniversity.ac.id](mailto:astrinov@telkomuniversity.ac.id), <sup>b</sup> [worldliner@telkomuniversity.ac.id](mailto:worldliner@telkomuniversity.ac.id)

## ARTICLE INFO

Received July 19<sup>th</sup>, 2020  
Revised October 16<sup>th</sup>, 2020  
Accepted October 20<sup>th</sup>, 2020  
Available online March 17<sup>th</sup>, 2021

### Keywords

sign language,  
recognition,  
dimensional reduction,  
principal component analysis,  
support vector machines

## ABSTRACT

The World Health Organization (WHO) estimates that over five percent of the world's population are hearing-impaired. One of the communication problems that often arise between deaf or speech impaired with normal people is the low level of knowledge and understanding of the deaf or speech impaired's normal sign language in their daily communication. To overcome this problem, we build a sign language recognition system, especially for the Indonesian language. The sign language system for Bahasa Indonesia, called Bisindo, is unique from the others. Our work utilizes two image processing algorithms for the pre-processing, namely the grayscale conversion and the histogram equalization. Subsequently, the principal component analysis (PCA) is employed for dimensional reduction and feature extraction. Finally, the support vector machine (SVM) is applied as the classifier. Results indicate that the use of the histogram equalization significantly enhances the accuracy of the recognition. Comprehensive experiments by applying different random seeds for testing data confirm that our method achieves 76.8% accuracy. Accordingly, a more robust method is still open to enhance the accuracy in sign language recognition.

## Acknowledgment

This work is supported by Telkom University Research Grant (Grant Number: KR-011/PNLT3/PPM/2015).

### \* Corresponding author at:

Department of Computer Engineering,  
Jl. Telekomunikasi No. 01, Terusan Buah Batu, Bandung, Jawa Barat 40257  
Indonesia  
Email: [astrinov@telkomuniversity.ac.id](mailto:astrinov@telkomuniversity.ac.id)

### ORCID ID:

First Author: 0000-0002-8109-5544

<https://doi.org/10.25124/ijait.v4i01.3015>

Paper\_reg\_number IJAIT000040106 2020 © The Authors. Published by School of Applied Science, Telkom University.  
This is an open access article under the CC BY-NC 4.0 license (<https://creativecommons.org/licenses/by-nc/4.0/>)

## 1. Introduction

The World Health Organisation (WHO) estimates that over five percent of the world's population are hearing-impaired, about 360 million people. About 91 percents of them are adults, and the remaining are children [1]. In Indonesia, the number of people with hearing loss is estimated at 9.6 million people [2]. It indicates that the deaf in Indonesia needs more serious attention, especially in education through verbal communication.

A common problem in daily communication between the deaf or mute and normal people is that the normal has a low level of sign language knowledge. To overcome this communication gap and improve communication skills, we build a sign language recognition system for the Indonesian language.

The sign language recognition system has been widely studied internationally in two approaches: visual-based and device-based approaches [3] [4]. In a visual-based approach, recognition of sign language is done by utilizing image processing techniques and machine learning algorithms, which have been used widely in both segmentation [5] [6] and classification, including gesture recognition [7] [8]. In this study, we also utilize visual-based technology for recognizing the Indonesian sign language automatically.

The sign language system for Bahasa Indonesia, called Bisindo, is a unique system that is different from the others, and the research is limited. We propose a method to build a sign language recognition system in Bisindo using the Principal Component Analysis as the dimensional reduction and feature extraction. The support vector machine is subsequently applied for the recognition. The integration can be used for researchers in the Bisindo sign language recognition system to consider open research.

A number of researches on Bisindo appear in the literature. Yunanda et al. used Microsoft Kinect, combined with the Hidden-Markov model, to recognize the sign [9]. Similarly, Handhika et al. utilized the Hidden-Markov model for the recognition [10]. Indra et al. employed chain code contour and similarity of Euclidean distance for the recognition [11]. Although some research on Bisindo starts arising, the number is still limited. In addition, the use of PCA accompanied by SVM has not been studied yet.

To elaborate, the sign of language recognition system consists of two sub-systems, feature extraction, and classification or recognition. This system is expected to be the initial phase of Sign Language Translator System development that should integrate the visual and audio aspects for the future.

This manuscript is organized as follows. The introduction described the background and the problem statement, and the contribution. The utilized method is described in section Method, and the experiments, including the dataset, results, and discussion, are presented in section Results and Discussions. Finally, the conclusion is presented in section Conclusion.

## 2. Methods

Our work utilizes algorithms of image processing and machine learning. We briefly describe these algorithms in the following subsections.

### 2.1. Image Processing Algorithms

Some image processing algorithms are utilized for pre-processing the data set. The algorithms are described as follows.

### 2.1.1. Image Conversion

All images are obtained in color images containing three layers: red, green, and blue layers (RGB). In this scheme, the images are converted from RGB form (three layers) to grayscale (single layer), for example, images in Figure 1 Image Conversion from 3 Channels (RGB) to A Single Channel (Grayscale). Changing the image layer is a basic pre-processing stage for all schemes in this study. In this scheme, the grayscale images are directly fed to the next sub-system, i.e., the sub-segmentation system.

The grayscale image contains only a single layer derived from the combination of channel red, green, and blue layers in the original image [12] [13]. The intensity (pixel value) in the grayscale can be expressed as:

$$I_g = \alpha R + \beta G + \gamma B$$

where  $I_g$  is the intensity of the grayscale image, R, G, and B are channel red, green, and blue respectively,  $\alpha, \beta, \gamma$  are constants.

$$\alpha + \beta + \gamma = 1$$

The constants must satisfy, and commonly, the green channel has the highest value among the constants.

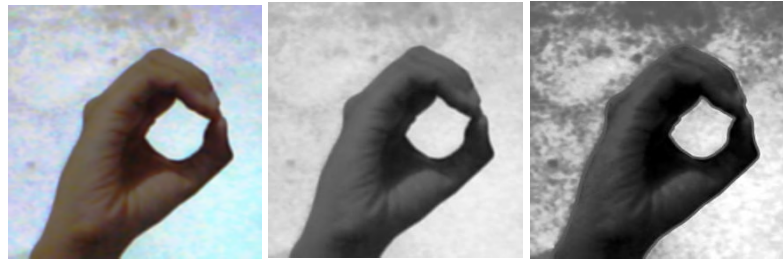


**Figure 1** Image Conversion from 3 Channels (RGB) to A Single Channel (Grayscale)

### 2.1.2. Histogram Equalization

Histogram equalization is a computer image processing technique used to increase contrast in an image [12] [13]. In this scheme, the pre-processing stages are as follows.

1. Convert the color images to grayscale
2. Apply thresholding process
3. Apply histogram equalization on the grayscale images
4. Multiplication between the image obtained from histogram equalization and the thresholded image, as shown in Figure 2 Image Conversion from 3 Channels (RGB) to A Single Channel (BW) and Applying Histogram Equalization. The difference between this scheme and the previous scheme is histogram equalization on the original image before the multiplication to get the global level of darkness over all images.



**Figure 2** Image Conversion from 3 Channels (RGB) to A Single Channel (BW) and Applying Histogram Equalization

## 2.2. The Principal Component Analysis (PCA)

PCA is a statistical technique that is very useful and widely used in object recognition and image compression. It is a common technique used to find patterns of data on high-dimensional data sets [14]. It can be used to transform the data from high dimensionality to a lower one.

Some fundamental stages in the PCA calculation are as follows [14] [15].

1. Preparation of input data in the form of a matrix,
2. Data normalization matrix,
3. Covariance matrix calculation,
4. Calculation of eigenvectors and eigenvalues of the covariance matrix,
5. Principal component selection to form characteristics vectors

In principle, the PCA is used for high-dimensional data reduction by selecting the given data's main components (principal component). Principal components are then utilized as a feature vector to find patterns in the data.

## 2.3. The Support Vector Machines (SVM)

One of the well-known algorithms for pattern recognition is the Support Vector Machines (SVM). This is reflected in the study results conducted by ICDM in 2006 as the top 10 in the Data Mining algorithms where the SVM goes to the 3<sup>rd</sup> [16].

The SVM is a discriminative classifier based on the separation of a hyperplane [17] [18]. In contrast to the neural network strategy that seeks class separating hyperplane, the SVM tries to find the input space's best hyperplane. The hyperplane can be expressed as

$$h(x) = \sum_{i=1}^N \alpha_i y_i K(x, x_i) + b_0$$

where  $\alpha_i$  is the estimated parameter for the SVM,  $y_i$  is the desired class for the corresponding  $x_i$ ,  $K$  as the kernel, and  $b_0$  as the bias.

The kernel used for the SVM during our experiments is linear. The kernel trick for the SVM in a high-dimensional workspace can be found in the references [19] [20].

## 3. Results and Discussions

Our experiments and results are as follows.

### 3.1. Dataset

The hand gesture datasets are obtained from online public data located in [21]. The datasets contain Bisindo alphabet signs. It consists of 24 letters, and the number of images is 585, which covers all letters in Bisindo fairly. Based on these images, the dataset is divided into two sub-datasets, namely the training set and testing set. The training set is taken 90% from the dataset, and the 10% remaining is treated as the testing set. To make the dataset is reproducible, we use seed when generating a random number to pick an image from the dataset. It is worth that for a fair experiment, we use seeds from 1 to 10 in Section 3.3. Some of the images are shown in Figure 3 Some Images in Datasets for Alphabet 'A'.



**Figure 3** Some Images in Datasets for Alphabet 'A'

The datasets' images are already cropped, but it still in color images, i.e., RGB channels.

### 3.2. Experimental Setup

There are two schemes used in our experiment, namely, experiment 1 and experiment 2. The difference between them is that the second experiment uses a histogram equalization algorithm after the image is grayscale. In contrast, the first experiment only uses a grayscale process for the pre-processing.

In general, the test scenario for experiments is carried out as follows.

1. All images in the original dataset are read and converted to grayscale images. In the first scheme of the experiment, we proceed directly to the next step. Whereas in the second experiment, the grayscale image will be

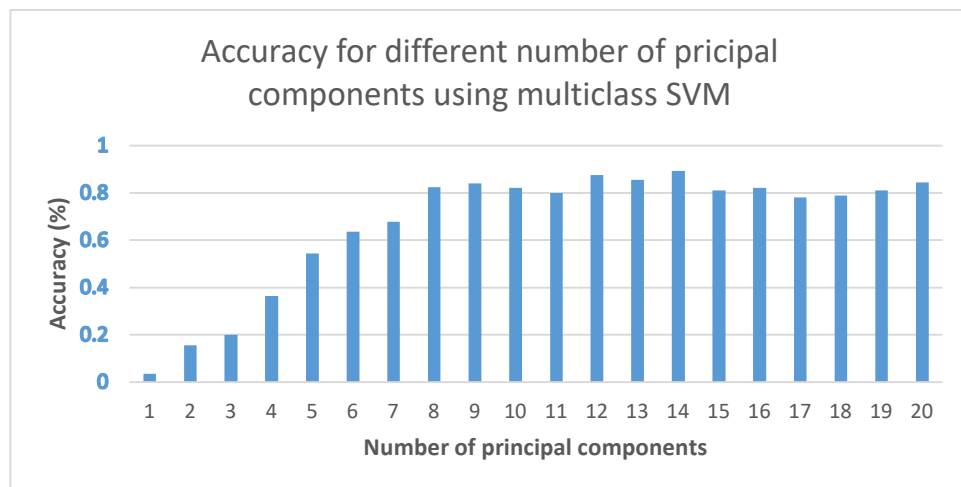
processed first by histogram equalization algorithm, then continue to the next step.

2. The pixels in each image are then converted to a feature vector.
3. All feature vectors are arranged into a single matrix.
4. The principal components are computed.
5. Some principal components are taken with covering the majority of the component.
6. Using the chosen component, the feature vector is converted to lower dimensionality.
7. SVM for multiclassification is calculated.
8. Cross-validation is used to compute the accuracy.

It is worth noting that the number of principal components is computed iteratively from 1 to 20. To carry on the cross validation, 10% of the dataset are picked randomly as testing.

### 3.3. Experimental Results and Discussions

The results of the experiments can be seen in the following chart.



**Figure 4** The Accuracy for Different Number of Principal Components (as features) using Multiclass SVM

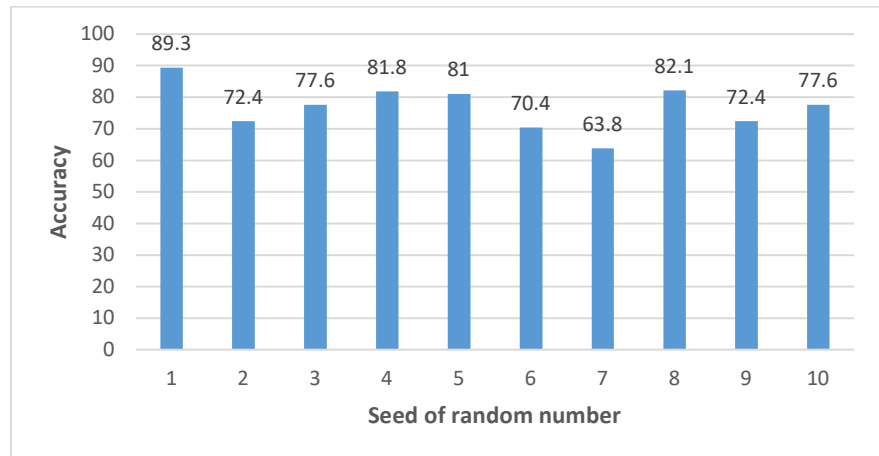
At the beginning of our experiments, the hand gesture dataset, which is in RGB format, is converted to grayscale images. In experiment 1, the grayscale images are utilized directly as a feature, while in experiment 2, each grayscale image is enhanced using histogram equalization. The contrast-enhanced images in experiment 2 are subsequently considered as a feature. The feature in experiments 1 and 2 are pixels from an image with the size 290x290. Then, principal component analysis is used to reduce the dimensionality. The original images are 290x290, which cover 84100 values. By applying PCA, the feature vector, if it is assumed that all pixels are a feature, is reduced to lower dimensionality, i.e., 1 to 20 only. The reduced features are the number of principal components and computed iteratively. Subsequently, the reduced components are utilized as a set of features for the support vector machines to carry on the classification. The accuracy for each number of principal components using multi-class SVM is shown in Figure 4.

As can be seen in Figure 4, experiment 2 demonstrates better performance compare to experiment 1. The histogram equalization is a common tool to enhance image contrast. In experiment 2, the use of the histogram equalization are also

effectively enhance the performance of classification. Starting from the number of principal components = 7, experiment 2 achieves higher accuracy.

Furthermore, the accuracy starts to be stable at the number of principal components = 8. In experiment 2, the classification consistently achieves more than 80% accuracy. The highest accuracy occurs when the number of principals is equal to 14.

Since the experiments above are sensitive to the dataset, it is important to see the random number generator's accuracy for different seeds. Experiments in Figure 4 are carried on using random seed = 1. By keeping the number of principal components equals 8, different random seeds are computed iteratively. Experimental results for different seeds are shown in Figure 5.



**Figure 5** The Accuracy for The Different Seed of Random Numbers to Create Different Data Training and Testing for The Multiclass SVM by Setting The Principal Component = 8

It can be seen that the accuracy changes accordingly to the seed of random number. The average accuracy using random seed from 1 to 10 is 76.8%. This accuracy can be considered as the final results of the experiments. The final result is lower than the preliminary results (89.%), which only utilizes a single seed only. Comprehensive experiments by applying different seeds confirm the final results.

#### 4. Conclusions

The obtained results show that increasing the number of principal components can initially enhance the classification performance; however, after a certain point, the performance becomes stable and not influenced by the number of principal components. Furthermore, the use of the histogram equalization for the grayscale datasets demonstrates better performance compared to the experiments that directly employ the grayscale without contrast enhancement. Preliminary results using the number of components = eight followed by a multi-class SVM classifier generate the best accuracy. Comprehensive experiments by applying different random seeds for testing to generate different training and testing data confirm that our method achieves 76.8% accuracy. Accordingly, a more robust method is still open to enhance the accuracy in sign language recognition.



## Bibliography

- [1] W. M. Centre, "Deafness and hearing loss," World Health Organization, March 2015. [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs300/en/>. [Accessed 30 August 2015].
- [2] A. Candra, "Jangan Abaikan Kesehatan Telinga," Kompas, 6 March 2013. [Online]. Available: <http://health.kompas.com/read/2013/03/06/15013243/>. [Accessed 30 August 2015].
- [3] R. Akmeliawati, M. P.-L. Ooi, and Y. C. Kuang, "Real-Time Malaysian Sign Language Translation using Colour Segmentation and Neural Network," in *IMTC 2007 - Instrumentation and Measurement Technology Conference*, Warsaw, 2007.
- [4] S.-H. Yu, C.-L. Huang, S.-C. Hsu, H.-W. Lin and a. H.-W. Wang, "Vision-Based Continuous Sign Language Recognition using Product HMM," in *IEEE First Asian Conference on* , 2011.
- [5] A. Pratondo, B. P. Nguyen, C.-K. Chui, and S.-H. Ong, "Vocal Cord Segmentation from CT Images Using Machine Learning," in *The 10th Anniversary Asian Conference on Computer-Aided Surgery*, Fukuoka, 2014.
- [6] A. Pratondo, C.-K. Chui and S.-H. Ong, "Robust Edge-Stop Functions for Edge-Based Active Contour Models in Medical Image Segmentation," *IEEE Signal Processing Letters*, vol. PP, no. 99, 2015.
- [7] V. Veeriah and P. Swaminathan, "Robust Hand Gesture Recognition Algorithm for Simple Mouse Control," *International Journal of Computer and Communication Engineering*, vol. 2, no. 2, pp. 219-221, 2013.
- [8] F.-S. Chen, C.-M. Fu and C.-L. Huang, "Hand gesture recognition using a real-time tracking method and hidden Markov models," *Image and Vision Computing*, vol. 21, pp. 745-758, 2003.
- [9] A. Yunanda, F. Mandita and A. Armin, "Pengenalan Bahasa Isyarat Indonesia (BISINDO) Untuk Karakter Huruf Dengan Menggunakan Microsoft Kinect.," *Fountain of Informatics Journal*, vol. 3, no. 2, pp. 41-45, 2018.
- [10] T. Handhika, R. I. M. Zen, D. P. Lestari, and I. Sari, "Gesture recognition for Indonesian Sign Language (Bisindo)," in *2nd International Conference on Statistics, Mathematics, Teaching, and Research*, Makasar, 2017.
- [11] D. Indra, S. Madenda and E. P. Wibowo, "Recognition of Bisindo Alphabets Based on Chain Code Contour and Similarity of Euclidean Distance," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 7, no. 5, pp. 1645-1652, 2017.
- [12] R. C. Gonzalez, *Digital Image Processing*, 4th Edition, Pearson India, 2018.
- [13] M. Nixon and A. Aguado, *Feature Extraction and Image Processing for Computer Vision* 4th Edition, London: Academic Press, 2019.
- [14] L. I. Smith, "A tutorial on Principal Components Analysis," 26 February 2002. [Online]. Available: [http://www.sccg.sk/~haladova/principal\\_components.pdf](http://www.sccg.sk/~haladova/principal_components.pdf). [Accessed 30 August 2015].
- [15] P. E. Hart, D. G. Stork, and R. O. Duda, *Pattern Classification*, Singapore: Wiley-Interscience, 2000.
- [16] H. Byun and S. Lee, "A Survey on Pattern Recognition Applications of Support Vector Machines," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 17, no. 3, p. 459-486, 2003.
- [17] S. O. Haykin, *Neural Networks and Learning Machines*, 3 ed., Prentice Hall, 2008.
- [18] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, p. 273-297, 1995.
- [19] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. Flannery, "Support Vector Machines," in *Numerical Recipes: The Art of Scientific Computing*, Cambridge, Cambridge University Press, 2007.
- [20] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge, UK: Cambridge University Press, 2000.
- [21] R. Maizul, "OpenCV-Bisindo-Alphabets-Sign-Detection," github.com, 2018.