Indonesian Sign Language Number Recognition using SIFT Algorithm

Isa Mahfudi*, Moechammad Sarosa, Rosa Andrie Asmara, M. Azrino Gustalika

Electrical Engineering Department, State Polytechnic of Malang, Malang, Indonesia

*isa_mahfudi@polinema.ac.id

Abstract. Indonesian sign language (ISL) is generally used for deaf individuals and poor people communication in communicating. They use sign language as their primary language which consists of 2 types of action: sign and finger spelling. However, not all people understand their sign language so that this becomes a problem for them to communicate with normal people. This problem also becomes a factor they are isolated feel from the social life. It needs a solution that can help them to be able to interacting with normal people. Many research that offers a variety of methods in solving the problem of sign language recognition based on image processing. SIFT (Scale Invariant Feature Transform) algorithm is one of the methods that can be used to identify an object. SIFT is claimed very resistant to scaling, rotation, illumination and noise. Using SIFT algorithm for Indonesian sign language recognition number result rate recognition to 82% with the use of a total of 100 samples image dataset consisting 50 sample for training data and 50 sample images for testing data. Change threshold value get affect the result of the recognition. The best value threshold is 0.45 with rate recognition of 94%.

1. Introduction
	Indonesian sign language (ISL) is generally used for deaf individuals and poor people communication in communicating. They use sign language as their primary language which consists of 2 types of action: sign and finger spelling. The system Indonesian sign language (SISL) dictionary as a formal standard in communicating their daily. The SISL dictionary is shown in figure 1. However, not all people understand their sign language so that this becomes a problem for them to communicate with normal people. This problem also becomes a factor they are isolated feel from the social life. There is a solution for them to be able to communicate with normal people. One of them is using the services of a sign language interpreter. this requires additional costs because of the need to pay the translator. Other solutions use a sign language recognition technique.

Techniques for sign language Recognition can be implemented with 3 methods[1]. Utilize the image processing technology, using glove data [1][2] and using sensors EMG (electromyography)[1]. Utilization technology of image processing to recognize sign language has its own challenges. Many research that offers a variety of methods in solving the problem of sign language recognition based on image processing [3][4][5][6][7][8][9]. Until now these existing methods still have shortcomings and will continue to do the repairs so that generated the best sign language recognition. SIFT (Scale Invariant Feature Transform) algorithm is one of the methods that can be used to identify an object.
SIFT is claimed very resistant to scaling, rotation, illumination, and noise [10][11]. In this paper, researchers proposed the SIFT method to use on the recognition of Indonesian sign language number.

Figure 1. The system Indonesian sign language (SISL) Dictionary

2. Related Work
Sign language recognition has been widely performed by the researcher, some of the methods proposed for produce the high accuracy. A method of HMM [9][12] has been used for sign language recognition. HMM used for classification of techniques. The using of HMM method resulted in accuracy 85%. The using of Edge Orientation Histogram for recognition of the 26 ASL characters use 2600 samples training datasets result in accuracy to 88.26% [8]. Researchers [6][7] used a Principle Component Analysis (PCA) algorithm as a technique of classification for sign language recognition of Vietnamese and evaluate its Euclidean and Mahalanobis distances in classification. This resulted in the success of the introduction to 90.4% and 91.5%. The use of the Kinect Sensor is also applied [5], this Kinect sensor used to detect the numbers and letters of the United Kingdom and integrated with the SVM method for classification process. This resulted in 100% accuracy for numbers and 70.59% for letters with a distance of 300 mm – Kinect sensor from 1000mm.

SIFT algorithm has been applied as an extraction and classification method. Researchers [15][16][17]. SIFT as a feature extraction method for an approach to batik pattern recognition. Using a collection of 120 batik images generated from 20 basic batik patterns, the proposed method shows an improvement over the original SIFT matching method with an equal error rate of 8.47% [15]. Integrating feature extraction between BOF and SIFT and using SVM for classification method for batik image classification results average accuracy of this method reaches 97.67%, 95.47% and 79% in normal image, rotated the image and scaled image, respectively[15]. Researcher [14] applied SIFT on Brain CT Images as Matching Algorithm. SIFT method combines with the gray feature. This experimental results accuracy from 89.83% to 92.78% with the time complexity basically unchanged, at the same time, the number of correct matches are not much reduced.

3. Proposed System
This research scheme refers to previous researchers who have proposed [7][13]. The Block diagram scheme proposed is shown figure 2. The workings of this system started taking pictures from the webcam as input data. the result image capture will do the process of segmentation to separate the object from the background. This segmentation using the HSV color model. This segmentation process results will be processed to minimize image cropping as well as do a filter for removing noise morphology and smooth image. next process will be applied extraction and classification method using SIFT algorithm[17][18].
3.1. Data Acquisition
This step is the stage for collect training dataset or testing dataset. This step started from capturing by webcam to take pictures with 1.3-megapixel resolution. When taking picture, the position of the camera facing straight with the position of the hands. The distance of the camera is set as far as 1 m from the hand. Captured hand gesture images describe Indonesian sign language number standard, starting from character number 1 to number 10. The total of the dataset is 100 sample pictures consists of 50 sample pictures for training data and 50 pictures for testing data. all sample datasets are conditioned with the same background and lighting. Indonesian sign language numbers dataset shown in Figure 3.

3.2. Preprocessing
Preprocessing consist image acquisition, segmentation and morphological filtering method. The first step preprocessing is segmentation, started by filtering skin color for getting the skin color. The input RGB Image is transformed into HSV color space. After the skin object has been separate from the background, applied morphology filters for reducing noise and smooth the object. morphology operations used are dilation, erosion, and opening. The block diagram preprocessing is shown in Figure 4.
3.3. Sift Based Image Matching
The SIFT algorithm has five major stages:

3.3.1. Scale-space construction. Suppose \( I(x,y) \) refers to the input image. In order to find reliable feature points in scale-space, filters are needed to process the images [17][18]. Gaussian transformation of images in scale-space can be expressed as:

\[
L(x, y, \sigma) = G(x, y, \sigma) \ast I(x, y)
\]

In which, \( G(x, y, \sigma) \) is a scale variable Gaussian function:

\[
G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}}
\]

\( \sigma \) is called scale space coordinate and decides image’s smoothness. Large-scale corresponds to image’s profile feature, and small-scale corresponds to minutiae. \( I(x,y) \) represents an image[18]. In order to compute reliable feature points in scale-space effectively, DOG (Difference-of-Gaussian) operator is applied to the approximate Laplacian-Gaussian operator. Suppose the approximate operator is \( D(x, y, \sigma) \) after processing the image we will have[17]:

\[
D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, k\sigma)) \ast I(x, y)
\]

\[
= L(x, y, k\sigma) - L(x, y, \sigma)
\]

3.3.2. Scale-space extrema detection. After approximating Laplacian-Gaussian operator with DOG operator, extrema detection is required in DOG scale space. The gray scale value of the point to be detected is compared with 18 pixels at the corresponding position of the front and rear images in the scale space, as well as the surrounding ones of the point itself. Altogether, a single point requires 26 comparisons. When an extremum (maximum or minimum) is detected, it is kept for further processing[17].

3.3.3. Orientation assignment. Local extrema detected in DOG scale-space are called keypoints after the operations of improving positioning accuracy and eliminating low-contrast points. To determine the keypoint orientation, a gradient orientation histogram is computed in the neighborhood of the
keypoint. The contribution of each neighboring pixel is weighted by the gradient magnitude and a Gaussian window with a value of \( \delta \) that is 1.5 times the scale of the keypoint. Peaks in the histogram correspond to dominant orientations. A separate keypoint is created for the direction corresponding to the histogram maximum and any other direction within 80% of the maximum value. All the properties of the keypoint are measured relative to the keypoint orientation, which provides invariance to rotation[17].

3.3.4. **Keypoint descriptor.** The descriptor is formed from a vector containing the values of all the orientation histogram entries. The best results are achieved with a 4×4 array of histograms with 8 orientation bins in each histogram[17]. this paper uses a 4×4×8 = 128 element feature vector for each keypoint.

3.3.5. **Feature vector matching.** As each feature point corresponds to a 128 element feature vector, searching for matching points from two images becomes searching for matching feature vectors. Euclidean distance is used for measuring the similarity degree between two feature vectors, the ratio between the closest Euclidean distance and that of the second closest are computed and if the ratio is lower than a certain threshold [17]. When the ratio of two distances is less than a given threshold, accept the match points. The Euclidian distance of two N dimension functions \( f_1(x) \) and \( f_2(x) \) is as follows[19]:

\[
d((f(x), g(x)) = \sqrt{(x_1 - g(x_1))^2 + f(x_1 - g(x_1))^2 + \cdots f(x_n - g(x_n))^2}
\]

\[
= \sqrt{\sum_{i=1}^{n} f(x_i - g(x_i))^2}
\]

In this paper using 0.65 the vectors are considered to be matched with each other. Example matching result with Threshold = 0.65 is shown figure 5.

![Matching results with T=0.65.](image)

**Figure 5.** Matching results with T=0.65.

4. **Experimental Results and Analysis**

This stage will be shown the result of experiments to the recognition of sign language numbers using SIFT algorithm with a variety of characters. Figure 6 shows a matching result with the value Threshold of 0.65.
The next stage will be shown the test results with 100 image dataset that consists 50 sample images testing data and 50 sample images training data. The result of the proposed system is shown in Table 1. There is 10 characters number Indonesian sign language with a total success rate up to 82%. The results show that the success rates of characters number “3”, “6”, “7”, “8” and “9” are low because they are shape-like. Thus the system makes some mistakes while recognition. For the characters such as “1”, “2”, “4”, “5” and “10”, the images of hand gesture are good quality after preprocessing stage. Thus the system achieves high success rate in recognition for these characters. Table 2 show test results matching by changing the value of the threshold.

| Character | Number of test image | Number of correct recognition | Success Rate |
|-----------|----------------------|--------------------------------|--------------|
| 1         | 5                    | 5                              | 100%         |
| 2         | 5                    | 5                              | 100%         |
| 3         | 5                    | 4                              | 80%          |
| 4         | 5                    | 5                              | 100%         |
| 5         | 5                    | 5                              | 100%         |
| 6         | 5                    | 5                              | 80%          |
| 7         | 5                    | 2                              | 40%          |
| 8         | 5                    | 2                              | 40%          |
| 9         | 5                    | 3                              | 80%          |
| 10        | 5                    | 5                              | 100%         |

Average 82%

| Value of Threshold | 0.45 | 0.55 | 0.65 | 0.75 | 0.85 | 0.95 |
|--------------------|------|------|------|------|------|------|
| Recognition Rate (%) | 94   | 90   | 82   | 78   | 74   | 72   |

Table 2 shows the recognition rate with a different threshold. We obtain the best result with T=0.45 that produces the value recognition rate to 94%. The illustrate of results sign language recognition number shown in Figure 6.
5. Conclusion
This paper proposes SIFT method for Indonesian sign language recognition number. on the value of the threshold of 0.65, the character numbers can be recognized very well are 1, 2, 4, 5 and 10 with the success rate of 100%. Change the value of the threshold obtained the best recognition rate at T = 0.45 which produces 94%.

6. Acknowledgment
The Authors would like to thank The State Polytechnic of Malang for supporting to attend this conference.

References
[1] M. E. Al-Ahdal and N. Md Tahir 2012 Review in Sign Language Recognition Systems IEEE Symposium on Computers & Informatics (ISCI) 52 – 57
[2] F. Husain, S. Gandhi, T. Nijhawan and Varsha Agarwal 2012 GESTURE RECOGNITION SYSTEM USING MATLAB: A LITERATURE REVIEW International Journal of Scientific Research and Management Studies (IJSRMS) 2 11 425-432.
[3] A. Izzah and N. Suciati 2014 Translation of Sign Language Using Generic Fourier Descriptor and Nearest Neighbour Int. J. Cybern. Informatics 3 1 31–41
[4] A. Kuznetsova, L. Leal-Taixe, and B. Rosenhahn 2013 Real-Time Sign Language Recognition Using a Consumer Depth Camera IEEE Int. Conf. Comput. Vis. Work 83–90
[5] W. Yi Yeh, T. Hui Tseng, J. Wei Hsieh and C. Ming Tsai 2016 Sign Language Recognition System Via Kinect: Number And English Alphabet International Conference on Machine Learning and Cybernetics (ICMLC) 2, 660 - 665
[6] S. N. Sawant 2014 Sign Language Recognition System to aid Deaf-dumb People Using PCA, Int. J. Comput. Sci. Eng. Technol 5 5 570–574
[7] T. N. T. Huong, T. V. Huu, T. Le Xuan and S. V. Van 2015 Static hand gesture recognition for Vietnamese sign language (VSL) using principle components analysis International Conference on Communications, Management and Telecommunications (ComManTel) 138-141
[8] J. R. Pansare and M. Ingle 2016 Vision-based approach for American Sign Language recognition using Edge Orientation Histogram IEEE International Conference on Image Vision, and Computing 86–90
[9] P. Kumar, H. Gauba, P. P. Roy and D. P. Dogra 2016 Coupled HMM-based Multi-Sensor Data Fusion for Sign Language Recognition, Pattern Recognition Letters 86 3–10,
[10] R. Azhar, D. Tuwohingide, D. Kamudi, Sarimuddin and N. Suciati 2015 Batik Image Classification Using SIFT Feature Extraction Bag of Features and Support Vector Machine Elsevier Procedia Computer Science 24-30.
[11] X. Zhou, K. Wang, J. Fu 2016 A Method of SIFT Simplifying and Matching Algorithm Improvement International Conference on Industrial Informatics - Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICI) 3-77

[12] M. Jebali, P. Dalle, and M. Jemni 2013 HMM-based method to overcome spatiotemporal sign language recognition issues IEEE Conference Publications 1-6

[13] K. S. Barai and S. Banoth 2016 Hand Gesture Recognition For Human Computer Interaction International Journal of Current Trends in Engineering & Research (IJCTER) 2 5, 241-245.

[14] Z. Xiao, L. Yu, Z. Qin, H. Ren, and Z. Geng 2016 A Point Matching Algorithm for Brain CT Images Based on SIFT and Gray Feature IEEE Conference Publications 1-6

[15] I. Nurhaida, A. Noviyanto, R. Manurung, A. M. Arymurthy 2015 automatic Indonesian’s Batik Pattern Recognition sing SIFT Approach, International Conference on Computer Science and Computational Intelligence (ICCSCI) 59 567-576.

[16] R. Azhar, D. Tuwohingide, D. Kamudi, Sarimuddin and N. Suciati 2015 “Batik Image Classification Using SIFT Feature Extraction, Bag of Features and Support Vector Machine”, The Third Information Systems International Conference 24-30

[17] W. Wei, H. Jun, and T. Yiping 2008 Image Matching for Geomorphic Measurement Based on SIFT and RANSAC Methods International Conference on Computer Science and Software Engineering 2 317-320

[18] X. Zhou, K. Wang and J. Fu 2016 A Method of SIFT Simplifying and Matching Algorithm Improvement International Conference on Industrial Informatics - Computing Technology, Intelligent Technology, Industrial Information Integration 73-77

[19] J. Zhu, S. Wang, and F. Meng 2011 SIFT Method for Paper Detection System, International Conference on Multimedia Technology 711-714