

# Basic Investigation for Improvement of Sign Language Recognition Using Classification Scheme

Hirotoishi Shibata, Hiromitsu Nishimura, and Hiroshi Tanaka<sup>(✉)</sup>

Kanagawa Institute of Technology, 1030 Shimo-Ogino,  
Atsugi, Kanagawa, Japan  
s1585009@cce.kanagawa-it.ac.jp,  
{nisimura, h\_tanaka}@ic.kanagawa-it.ac.jp

**Abstract.** Sign language is a commonly-used communication method for hearing-impaired or speech-impaired people. However, it is quite difficult to learn sign language. If automatic translation for sign language can be realized, it becomes very meaningful and convenient not only for impaired people but also physically unimpaired people. The cause of the difficulty in automatic translation is that there are so many variations in sign language motions, which degrades recognition performance. This paper presents a recognition method for maintaining the recognition performance for many sign language motions. A scheme is introduced to classification using a decision tree, which can decrease the number of words to be recognized at a time by dividing them into groups. The used hand, the characteristics of hand motion and the relative position between hands and face have been considered in creating the decision tree. It is confirmed by experiments that the recognition success rate increased from 41 % and 59 % to 59 % and 82 %, respectively, for a basic 17 words of sign language with four sign language operators.

**Keywords:** Sign language · Color gloves · Optical camera · Recognition · Classification · Decision tree

## 1 Introduction

Sign language is a commonly-used communication method for hearing-impaired or speech-impaired people. Written conversation is another method for communication with physically unimpaired people. However, this method takes some time and cannot transmit the intended message simultaneously to many people. The automated conversion from words to sign language has been investigating recently [1, 2], and it seems to be approaching a practical level. Evolution of animation technology and computer processing ability make this possible.

Although interpretation from sign language to words has also been studied for many years, the technologies have not yet matured to the level of practical use. Specifically, some methods that use a special sensor or device [3, 4] incur a high introduction cost or sensors must be attached to the body. The detection target is mainly limited to hand motions and the hand shape, and finger motion are not observed.

Therefore, it seems difficult to realize a high recognition performance. The number of recognized words is limited [5, 6]. These technological results are insufficient for practical use. In addition, although the limited scene for usage should be proposed, they have not investigated from aspect of this point in previous studies.

The authors also have been investigating a recognition method for sign language that uses colored gloves [7]. The main feature of our proposed method is to distinguish each finger, and the front and back of the palm. A recognition rate of 83.8 % for 24 words was obtained by the proposed method [8]. The 24 words in this investigation were selected from a Japanese official certification examination for sign language. To make the recognition method practical, the number of words, and the recognition performance must be increased.

This paper presents a recognition method that is appropriate for increasing the number of words with high recognition performance. Classification scheme for words to be recognized has not been included in a conventional investigation, therefore, the recognition performance is thought to degrade according to the increase in the number of words. The authors propose a classification that considers the features of the sign language motion for each word, for example, the size of hand motion and the used hands, that is, one or both for the sign language, etc. A decision tree is introduced for the classification in this study. The recognition process is applied to the words after classification using the decision tree. The number of words to be recognized can be decreased, therefore, a higher recognition performance is expected to be maintained compared with conventional method [8]. The effectiveness of the proposed method was confirmed by the experiment that was carried out by operators wearing colored gloves.

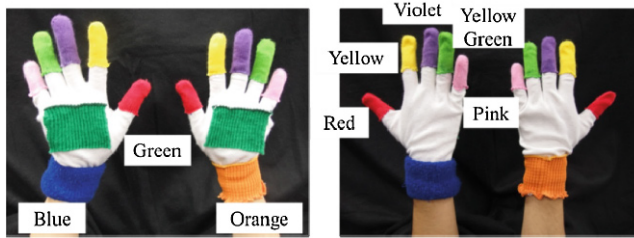
## 2 Motion Detection and Colored Gloves to Be Used

It is desirable to select commonly used equipment for sign language recognition. This makes it easier to disseminate its usage among many people. At present, smartphones have become consumer devices, and they are equipped with many sensors, including a camera. Their processing capability has also been increasing. The implemented camera sensor and the smartphone's high processing ability will enable the device to recognize sign language and output its translated meaning as text or voice communication.

The following items seem to be necessary to recognize sign language by considering its motion.

- (a) Identification of each finger and both hands
- (b) Motion detection of wrists/hands and fingers
- (c) Hand shape recognition
- (d) Discrimination between the two sides of the palm
- (e) Relative position of hands and face
- (f) Detection of mouth motion and recognition of facial expression

Identifying each finger is one of the important factors for hand shape recognition. Colored gloves are proposed for hand shape recognition [9]. The tip of each finger of the glove has a different color. This makes it easy to identify each finger and can lead to reliable recognition of hand shapes. If we use wrist bands for both hands, the right and



**Fig. 1.** Proposed colored gloves (Color figure online)

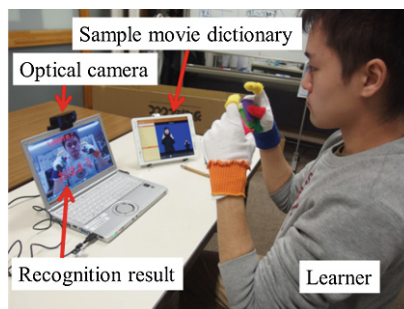
left hand are easily distinguished. In addition, the palms of the hands can be identified by the presence of colored regions. Wrist motions can be monitored by detecting the center of gravity of the regions traversed by the colored wrist bands, and we regard this motion as hand motion. The face position is detected by face recognition using image processing. Therefore, challenges (a)–(e) are considered to be met by using colored gloves.

The colored gloves we designed are shown in Fig. 1. Five colors are used so as to uniquely identify each finger, different additional colors distinguish each wrist, and green patches locate the palms of the hands. Thus a total of 8 colors are proposed to recognize sign language. Background subtraction is used on the camera image to extract 8 colored regions, each being identified by its hue and saturation values [7].

### 3 Classification by Decision Tree for Sign Language Words

#### 3.1 Application of This Study and Targeted Words

The authors now consider that this investigating method applies to learning tools for sign language. An image of using scenery is shown in Fig. 2. A user watches a sign language sample movie dictionary [10] in order to memorize their motions, and after memorizing he/she tries a motion in front of a camera without the sample movie to review learning. Using this research, the recognition system can output a recognition result from their motion, to be used as a learning tool.



**Fig. 2.** Application of this research

The 17 basic sign language words are included in this movie dictionary. Therefore, in first stage of this investigation, recognition technologies in this study are investigated for these 17 words. They are shown in Table 1.

**Table 1.** Sign language words

Word number	1	2	3	4	5	6
Word	You あなた	Thank you ありがとう	No いいえ	Good job お疲れ様	Good morning おはよう	Sorry ごめんなさい
Word number	7	8	9	10	11	12
Word	That's right 1 そうですね1	That's right 2 そうですね2	What? 何?	I see 1 なるほど1	I see 2 なるほど2	Drink 飲む
Word number	13	14	15	16	17	
Word	Nice to meet you 初めまして	Of course もちろん	Good 良い	Be nice よろしくお願いします	I 私	

### 3.2 Classification of Sign Language Words

There are countless sign language words. It seems to be quite difficult to maintain high recognition performance for a lot of words. Therefore, we propose a classification for sign language words by their characteristics of motion in order to maintain recognition performance. Better recognition performance can be expected by reducing the number of sign language words by classification before the recognition process. The motion characteristics of 17 recognition words are analyzed, and the following features are considered for classification in this investigation.

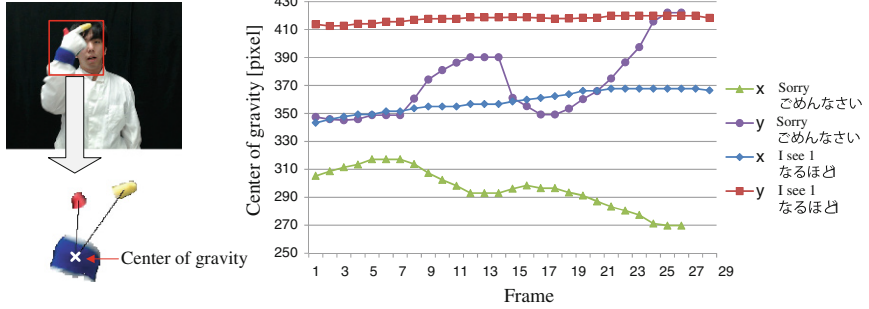
1. Hands used for sign language: Right hand/Both hands
2. Movement of each hand: Large motion/Small motion
3. Difference of distance between each hand and the face: Large/Small

### 3.3 Classification by Hand Detection

There are sign language words that are expressed only using the right hand. Therefore, as the first step for classification, the authors proposed classification by the used hands, that is, right hand or both hands. This is decided by the detection of the colored region of blue assigned for the right wrist and orange for the left wrist. Appearance of the wrist in more than one-fifth of the total frames of a sign language motion will indicate that the hand is used.

### 3.4 Classification by Hand Movement

The authors propose a second classification by the size of hand motion. Some sign language words have large hand motions, and others have little motion. As the same as previous section, the center of gravity of the colored region of each wrist is used for detection of hand movement. Figure 3 indicates an example of the movement of the center of gravity of two sign language motions: the words “Sorry” and “I see”. The unit of the y axis indicates a pixel.



**Fig. 3.** Movement of center of gravity of the colored region of the wrist

Movement measurement of the hand motion in pixels depends on the number of pixels of used camera and the distance between camera and the colored gloves. Therefore, experimental conditions are set in this investigation as follows. The distance between the camera (Logicool HD ProWebcam 920) and the colored gloves is about 1 m, the image frame size is 800 \* 600 pixels, and the frame rate is 30 fps. The hand motion and hand shape are measured based on the results of color detection for the colored regions of gloves. Detection is severely affected by the illumination conditions. Therefore, the experiment was carried out under a constant illumination condition of 230 lx.

Two types of sign language operators were kept in this investigation. One group consisted of operators who learned sign language using a movie dictionary (amateur), and the other group was sign language user in their daily lives (native signer). We consider that the diversity is important in creating a dictionary for sign language recognition. Therefore, the decision tree for classification and the dictionary for recognition were composed of samples generated by an operator pair with an amateur and a native signer. Recognition experiments were carried out by a pair with a different amateur and a different native signer. The operators list is shown in Table 2.

**Table 2.** Sign language operator list

Name	Role	Sign language skill
Operator A	Making decision tree and dictionary for recognition	Native signer
Operator B	B Same above	Amateur
Operator C	Recognition target	Native signer
Operator D	Same above	Amateur

The movement is evaluated from the average of the movement of the center of gravity of each wrist region, and its average was calculated from the following expression for classification.

Here,

n: the total number of frames of sign language motion data

i: i<sup>th</sup> frame of motion data

$W_x$ : x coordinate of the of center of gravity of colored region of the wrist  
 $W_y$ : y coordinate of the of center of gravity of colored region of the wrist

$$\frac{\sum_{i=1}^{n-1} \sqrt{(W_x(i+1) - W_x(i))^2 + (W_y(i+1) - W_y(i))^2}}{n - 1} \tag{1}$$

The 17 sign language words were examined by the operators A & B. The data from these two operators are shown in Fig. 4. It is natural that there is some variation between two operators and among each of the sign language words. Here, we introduce the idea of range for each classification criterion. The upper limit for classifying words as small motions, and the lower limit for large motions are defined. The “no decision” range to classify ambiguities in motion size is introduced. This method helps to avoid classification failure; as a result, it contributes to maintaining the recognition performance.

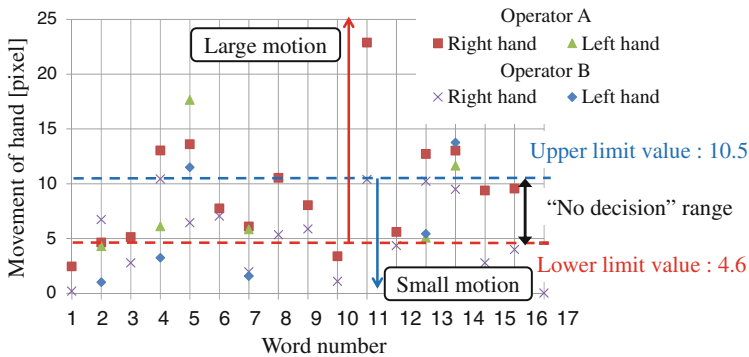


Fig. 4. Movement of the hand motion and classification criterion

### 3.5 Classification by Distance Between Face and Each Hand

There are differences in the distance between the face and each hand in each sign language word. An example is shown in Fig. 5. The photo on the left shows an example of a large difference in the distance between left and right hand and face ( $d_R$  and  $d_L$ ), and one on the right shows an example of a small difference.  $d_R$  is almost equal to  $d_L$  in the photo on the right. The authors use this feature as a third classification in our scheme.

The classification criterion is the average value of the difference of the distance between the face and each hand in motion. Its value is calculated by expression (2). Here, the meaning of  $i$  and  $n$  is the same as expression (1), and the others are as follows.

- $d_R$ : distance from the right wrist and the center of the gravity of the face
- $d_L$ : distance from the left wrist and the center of the gravity of the face
- $fs$ : size of face



Fig. 5. Distance between face and each hand

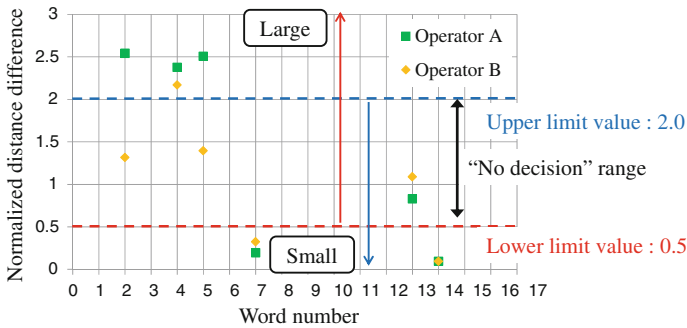


Fig. 6. Distance difference between face and each hand and classification criterion

The distance between the face and hands is normalized using the image size of the face. This prevents any effect from the differences in distance between the sign language operator and the camera. This is a necessary function when this method is applied for practical use. The position and the size of the face are detected using the face recognizer included in OpenCV [11].

$$\sum_{i=1}^n \frac{|d_R(i) - d_L(i)|}{fs(i)} \quad (2)$$

Figure 6 shows the distance difference between the face and each hand. This result was obtained by the same two operators obtained the result in Fig. 4. The authors propose the upper limit for classifying as a small distance difference and lower limit for large distance difference the same as Fig. 4. Although some words are included in a common area by the “no decision” range, this avoids discrimination failure at this stage of classification.

### 3.6 Decision Tree Obtained by Two Operators

The decision tree can be composed by using the results of Sects. 3.3, 3.4 and 3.5. Figure 7 shows a decision tree. The 17 words are each assigned to a group, that is, leaf node. The feature of this classification is that some words belong to multiple groups. This is because the “no decision” range is defined in the classification.

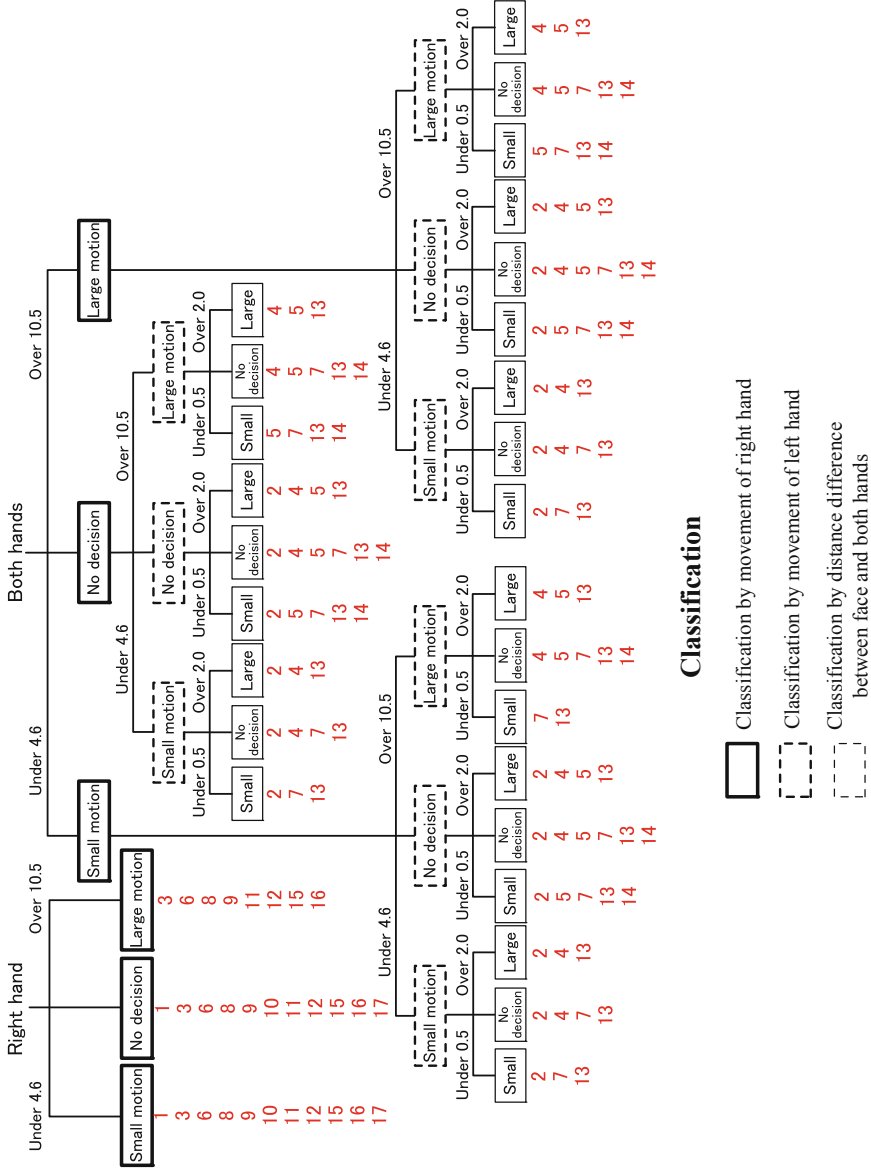
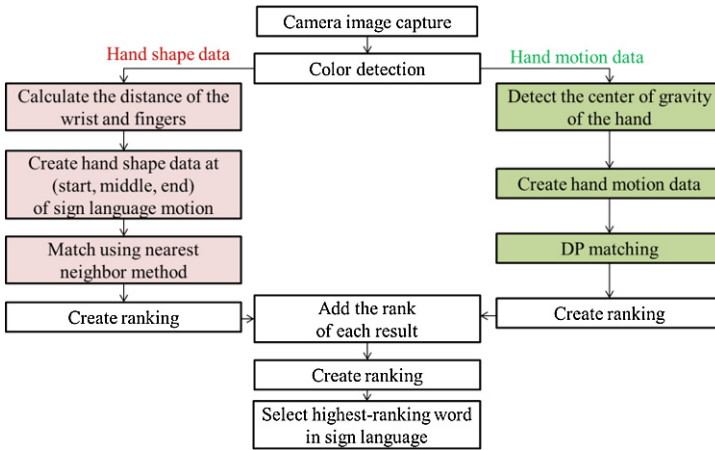


Fig. 7. Decision tree from features of motion and classification result





**Fig. 8.** Flow of recognition method

This helps avoid failure in classification; as a result, this can increase the performance of sign language recognition. The recognition method is applied to each word that is belonged to each group. Since the number of the words in the recognition process can be decreased, higher recognition performance is expected from the proposed method. When a word cannot be classified, recognition using the 17 words dictionary is carried out in this process, for example, in the case of no detection of the right hand.

## 4 Sign Language Recognition Method

A sign language recognition method after classification that uses hand shape and hand motion recognition is shown in Fig. 8 [8]. At this stage, the hand shape recognition process is applied at the beginning, middle, and the end of a sign language motion. The hand motion recognition process is applied over the span between the beginning and the end of the sign language motion. The identification of the word is based on these results.

Hand shape recognition is based on the hand shape feature vector whose magnitude is the distance between the center of the wrist and the tip of each finger. The magnitude for an invisible finger-tip due to a finger being bent and occlusion is set at zero. The shape recognition result is obtained by selecting the hand shape for which the distance between template feature vectors prepared in advance and the observed target feature vector is smallest [9]. Hand motion can be detected by obtaining the center of gravity of the colored region of the wrist band. The DP matching scheme is applied to motion recognition. Differences in motion sizes are taken into account by normalizing the motion ranges.

## 5 Recognition Experiments and Evaluation

### 5.1 Experiment Method

The recognition results after this classification are compared with the recognition results from conventional method, that is, without classification, in order to verify the effectiveness of the proposed method. If the classification is appropriate, it can be expected to the increase recognition success rate.

When the “no decision” range between the upper and lower limit value is smaller, it can be expected that classification errors appear. It is considered to exist appropriate range since the narrow range produces classification errors that directly affect sign language recognition performance. However, narrow range reduces the number of words to be recognized by the classification process, which leads to high recognition performance. Therefore, we carry on recognition experiment by changing this range. There are two ranges shown in Figs. 4 and 6. These ranges were changed simultaneously.

### 5.2 Experiment Result

First, the recognition results using the conventional method [8] without classification are shown in Table 3. Each number shows each candidate’s ranking of words as the recognition result. Namely, number 1 means a successful recognition result for that word. According to Table 3, the recognition success rate was 41 % and 59 % for two operators. The table on the left shows operator C, and the table on the right shows operator D.

**Table 3.** Recognition results using conventional method

		Dictionary																
		Wn1	Wn2	Wn3	Wn4	Wn5	Wn6	Wn7	Wn8	Wn9	Wn10	Wn11	Wn12	Wn13	Wn14	Wn15	Wn16	Wn17
Operator C	Wn1	3	14	11	11	14	7	13	7	7	5	7	4	16	17	2	5	1
	Wn2	4	14	11	5	12	10	15	3	6	8	9	11	17	15	6	12	2
	Wn3	9	17	3	12	15	2	14	4	1	3	4	10	16	13	6	7	11
	Wn4	10	15	14	16	7	12	8	6	11	8	12	15	17	3	5	3	3
	Wn5	9	17	11	12	3	6	14	5	1	6	8	12	16	15	1	3	10
	Wn6	10	15	13	14	11	12	5	2	8	4	11	17	15	5	3	8	8
	Wn7	11	13	13	3	15	11	8	6	5	11	7	7	16	19	3	17	2
	Wn8	6	17	5	12	14	7	13	2	4	9	3	11	16	15	1	8	10
	Wn9	9	17	11	12	13	6	15	2	4	8	2	11	16	14	4	6	10
	Wn10	3	17	2	12	14	9	13	7	1	4	5	3	15	16	5	9	8
	Wn11	8	17	7	12	13	6	14	10	2	4	1	10	15	15	3	4	9
	Wn12	7	13	6	12	15	11	13	9	5	3	4	1	16	17	2	10	7
	Wn13	2	12	14	3	15	11	16	9	9	7	6	7	11	17	3	13	4
	Wn14	12	15	10	14	3	12	16	8	7	17	6	4	11	3	1	2	8
	Wn15	9	17	10	13	12	4	15	4	7	8	3	11	14	16	1	2	6
	Wn16	9	17	7	12	13	11	14	4	4	9	2	8	16	15	1	2	6
	Wn17	5	17	11	10	13	9	14	1	8	6	8	12	15	16	3	3	2
Operator D	Wn1	1	16	11	11	15	8	13	2	3	9	6	7	17	14	5	9	3
	Wn2	5	2	11	7	4	16	16	8	3	15	11	6	9	14	1	13	9
	Wn3	9	17	1	14	12	4	15	3	2	7	10	8	16	12	4	6	11
	Wn4	10	12	14	7	15	5	3	8	8	11	6	12	18	17	4	7	2
	Wn5	1	16	7	13	2	11	14	2	4	8	6	9	17	14	5	12	10
	Wn6	6	17	11	12	14	4	13	8	3	9	8	10	15	16	1	1	4
	Wn7	3	14	14	1	13	10	5	6	6	11	7	12	18	16	2	8	4
	Wn8	6	17	11	10	14	4	13	4	2	8	9	12	15	16	7	2	1
	Wn9	5	17	4	13	15	7	14	2	7	9	6	11	16	12	2	9	8
	Wn10	4	17	6	11	14	6	13	10	7	11	5	12	15	16	6	9	2
	Wn11	6	16	3	12	13	9	13	11	1	6	1	10	15	17	4	5	8
	Wn12	2	12	9	13	14	11	14	5	2	7	10	1	16	17	4	8	6
	Wn13	13	15	16	3	4	14	11	10	5	7	9	8	17	12	12	5	3
	Wn14	7	15	10	14	11	8	16	3	11	11	6	11	17	4	11	6	5
	Wn15	5	17	12	10	15	1	13	9	6	8	6	11	14	16	4	2	2
	Wn16	8	17	10	10	14	5	13	7	3	6	8	12	15	16	2	1	4
	Wn17	3	16	12	11	13	10	14	4	8	8	6	4	15	17	2	6	8

Operator C Recognition success rate: 41%

Operator D Recognition success rate: 59%

Next, the recognition results using the proposed method are shown in Table 4. The “no decision” ranges shown in Figs. 4 and 6 were used in this experiment. The black cells indicate that these were eliminated by classification process. In this experiment, classification error wasn’t occurred. The classification results are different between operator C and D. However, the number of words to be recognized has been decreased by classification. According to Table 4, the recognition success rate was 59 % and 82 % respectively. In this result, it was confirmed that the proposed method has the effect of improving recognition performance.

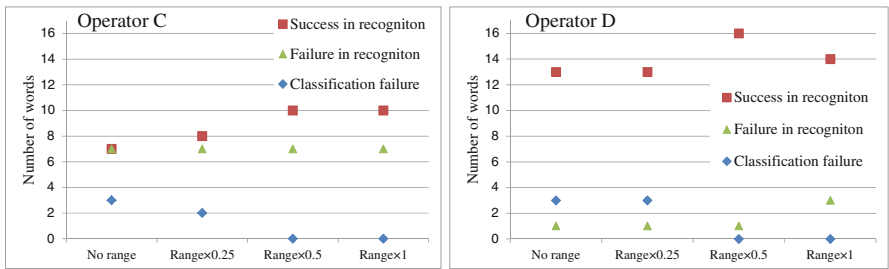
**Table 4.** Recognition results using proposed method

		Dictionary																	
		Wn1	Wn2	Wn3	Wn4	Wn5	Wn6	Wn7	Wn8	Wn9	Wn10	Wn11	Wn12	Wn13	Wn14	Wn15	Wn16	Wn17	
Operator C	Wn1	3	14	11	14	7	13	7	5	7	4	16	17	2	5	1			
	Wn2		2		1														
	Wn3	10				1	4	1	7	1	7	6	7	11					
	Wn4	4		1	5	2							3	6					
	Wn5			2	1														
	Wn6	10		5		1		5	2	8	3	11			3		5	3	9
	Wn7		2				1							3					
	Wn8	6		5			7		4	10	1	11			1	8	9		
	Wn9	8		1		6		2	5	6	2	11			2	9	10		
	Wn10	3		2		11	7	1	4	9	10	6	8	8					
	Wn11			6		4	7	2		1	7				2	4			
	Wn12	6		8		11		6	4	2	4		1	5	2	10	6		
	Wn13		2	3	5	4							1	5					
	Wn14				2		4						3	1					
	Wn15	9		9		4		5	7	5	3	9			1	2	8		
	Wn16	9		3		9		5	4	9	2	7			1	2	5		
	Wn17	4		10		8		1	8	7	4	10			1	3	4		

Operator C Recognition success rate: 59%

		Dictionary																	
		Wn1	Wn2	Wn3	Wn4	Wn5	Wn6	Wn7	Wn8	Wn9	Wn10	Wn11	Wn12	Wn13	Wn14	Wn15	Wn16	Wn17	
Operator D	Wn1	1	11	3			8		3	3	8	5	7		5	10	2		
	Wn2																		
	Wn3	10		1			3		4	2	7	9	7		2				
	Wn4			2		1													
	Wn5			8		2		1		3	5	3		6	7		3		
	Wn6																		
	Wn7		6		1	3		1							5	4			
	Wn8	5		10			1		1	3	8	8	11			5	3	5	
	Wn9			4			6		3	11	4	7				2	8		
	Wn10	4		5			7	10	2	1	5	11			8	9	3		
	Wn11			2			6	7	1		2	8			4	4	4		
	Wn12	2		3	9		11		2	5	2	7	9		1	4	7	6	
	Wn13																		
	Wn14								2		3				4	1			
	Wn15	7		11			1		9	7	5	5	10			2	2	2	
	Wn16	9		3			9		3	5	3	5	7			2	1		
	Wn17	3		11			9		6	9	8	5	4			2	8	1	

Operator D Recognition success rate: 82%



**Fig. 9.** Change of the recognition performance by “no decision” range

### 5.3 Change of the Recognition Performance by “no Decision” Range

The “no decision” range affects the recognition performance. The recognition performance was investigated by changing this range and the results are shown in Fig. 9. In this figure, the results on the left were obtained by operator C, and the results on right were operator D. Range  $\times 1$  indicates the recognition result from the previous section, that is, the original range. When the original range was changed to range  $\times 0.25$ , classification failures appeared in each recognition result. It was found that a narrow range causes classification failure and a lower success rate. It was confirmed that it is necessary to define an appropriate range in this proposed method to achieve recognition performance.

## 6 Conclusion

The authors have proposed a classification method for realizing high performance recognition for sign language. The number of recognition targeted motion can be decreased by classification, since multiple motions can be divided into several groups. The recognition method is applied to the motions that belong to each group. The classification process has been defined by considering the features of each sign language motion. In this study, three features are taken into account, that is, the used hand, the range of hand motion and the distance relation between the face and each hand.

The decision tree was created based on these results. The feature of the proposed decision tree is a “no decision” range in order to avoid classification failures. Experiments were carried out to evaluate the recognition success rate using the conventional method and the proposed method for a basic 17 words in sign language. The success rate was confirmed to be increased from 41 % and 59 % to 59 % and 82 %, respectively. The effectiveness of the proposed method has been confirmed by experiments carried out with sign language operators.

## References

1. Jen, T., Adamo-Villani, N.: the effect of rendering style on perception of sign language animations. In: Antona, M., Stephanidis, C. (eds.) UAHCI 2015. LNCS, vol. 9176, pp. 383–392. Springer, Heidelberg (2015)
2. Adamo-Villani, N., Wilbur, R.B.: ASL-Pro: American sign language animation with prosodic elements. In: Antona, M., Stephanidis, C. (eds.) UAHCI 2015. LNCS, vol. 9176, pp. 307–318. Springer, Heidelberg (2015)
3. Baatar, B., Tanaka, J.: Comparing sensor based and vision based techniques for dynamic gesture recognition. In: The 10th Asia Pacific Conference on Computer Human Interaction (APCHI), Poster 2P-21 (2012)
4. Matsuda, Y., Sakuma, I., Jimbo, Y., Kobayashi, E., Arafune, T., Isomura, T.: Development of finger braille recognition system. *J. Biom. Sci. Eng.* **5**(1), 54–65 (2010)
5. Humphries, T., Padden, C., O’Rourke, T.: Basic Course in American Sign Language. T. J. Publishers Inc., Silver Spring (1994)
6. Murakami, K., Taguchi, H.: Gesture recognition using recurrent neural networks. In: CHI 1991 Conference Proceedings, pp. 237–242 (1991)
7. Sugaya, T., Nishimura, H., Tanaka, H.: Enhancement of accuracy of hand shape recognition using color calibration by clustering scheme and majority voting method. In: Yamamoto, S. (ed.) HCI 2014, Part I. LNCS, vol. 8521, pp. 251–260. Springer, Heidelberg (2014)
8. Sugaya, T., Tsuchiya, H., Iwasawa, H., Nishimura, H., Tanaka, H.: Fundamental study on sign language recognition using color detection with an optical camera. In: International Conference on Imaging and Printing Technologies (ICIPT), Mandarin Hotel, Bangkok, Thailand, pp. 8–13 (2014)
9. Sugaya, T., Suzuki, T., Nishimura, H., Tanaka, H.: Basic investigation into hand shape recognition using colored gloves taking account of the peripheral environment. In: Yamamoto, S. (ed.) HCI 2013, Part I. LNCS, vol. 8016, pp. 133–142. Springer, Heidelberg (2013)
10. SmartDeaf: <http://www.smartdeaf.com/>
11. OpenCV: <http://opencv.org/>