Investigation of Sign Language Recognition Performance by Integration of Multiple Feature Elements and Classifiers

Tatsunori Ozawa[1], Yuna Okayasu[1,2], Maitai Dahlan[3], Hiromitsu Nishimura[1,2], and Hiroshi Tanaka[1,2(✉)]

[1] Course of Information and Computer Sciences, Graduate School of Kanagawa Institute of Technology, 1030 Shimo-ogino, Atsugi-shi, Kanagawa, Japan
{sl785014, sl421150}@cce.kanagawa-it.ac.jp,
nisimura@ic.kanagawa-it.ac.jp

[2] Department of Information and Computer Sciences, Kanagawa Institute of Technology, Atsugi, Japan
h_tanaka@ic.kanagawa-it.ac.jp

[3] Course of Mechanical Engineering, Graduate School of Chulalongkorn University, Bangkok, Thailand
maitai8town@gmail.com

Abstract. Sign languages are used by healthy individuals when communicating with those who are hearing or speech impaired as well by those with hearing or speech impediments. It is quite difficult to acquire sign language skills since there are vast number of sign language words and some signing motions are very complex. Several attempts at machine translation have been investigated for a limited number of sign language motions by using KINECT and a data glove, which is equipped with a strain gauge to monitor the angles at which fingers are bent, to detect hand motions and hand shapes.

One of the key features of our proposed method is using an optical camera and colored gloves for detection of sign language motion. The optical camera is implemented in a smartphone. This makes it possible to remove the limitation of using area and occasion as a machine translation tool.

The authors propose two new schemes. One is to add the two feature elements, that is, hand direction obtained from the angle between the wrist and fingertips, and hand rotation calculated from the visible size of the palm and wrist incorporating the four conventional elements comprising motion trajectory, motion velocity, hand position and hand shape. The other is integrating the results which is obtained by each classifier to enhance the recognition performance. The six kinds of classifiers have been applied to 35 sign language motions.

A total of 3150 pieces of motion data, that is, 2100 pieces of motion data as training data and 1050 pieces as evaluation data, were used to evaluate the proposed method. The recognition results were examined by integrating the feature elements and classifier. The success rate for 35 words was respectively 76.2% and 94.2%, for the selection of the first ranked answer, and the selection of the first, second or third ranked answers. These values suggest that the proposed method could be used as a review tool for assessing how well learner have mastered sign language motions.

S. Yamamoto and H. Mori (Eds.): HIMI 2018, LNCS 10904, pp. 291–305, 2018.
https://doi.org/10.1007/978-3-319-92043-6_25

Keywords: Sign language · Color gloves · Optical camera · Classifiers
Feature element · Ensemble learning

1 Introduction

Sign language recognition is indispensable in communication with and between hearing impaired people. It is quite difficult to acquire sign language skills since there are vast number of sign language words and some signing motions are very complicated. Furthermore, even if one person has learned sign languages, communication is impossible unless these signs can be recognized. Several gesture recognition systems [1] have been proposed for a limited number of sign language motions by using KINECT [2] and a data glove [3], which is equipped with a strain gauge to monitor the angles at which fingers are bent, to detect hand motions and hand shapes.

One of the key features of our proposed method is the use of an optical camera and colored gloves for detection of sign language motions. The optical camera is implemented in a smartphone. In motion recognition that is not limited to sign language, the widely available Toolkit [4] has also been released, but technology to recognize the movement of fingertips like sign language using only optical camera images has yet to be developed. Studies that did not use colored gloves but used different means to recognize the shape of static hands have been reported [5]. However, we are aiming to develop a sign language recognition technique using optical cameras and colored gloves, giving priority to ease of introduction and the capacity to detect high-speed movements.

We proposed a method for recognizing sign language motion from hand position information acquired by an optical camera using the hidden Markov model (HMM) at HCII'2017 conference [6]. To achieve better recognition performance, we propose two new feature elements to accurately describe the sign language motions. We also propose several different recognition methods and examine a scheme to integrate the recognition results in this paper. It is known that superior recognition performance can be achieved by ensemble learning [7] that combines multiple recognition methods. Our proposal is based on the idea that classifiers work in a complimentary fashion in relation to each other since each classifier is designed based on different classification criteria.

2 Motion Detection and Data Creation

2.1 Colored Gloves

Because identifying each finger is one of the crucial factors for hand shape recognition, colored gloves are proposed for hand shape recognition [8]. The tip of each finger of the glove has a distinct color. This makes it easy to discriminate each finger and results in reliable recognition of hand shapes. If we use wrist bands for both hands, the right and left hand are easily distinguished. The entire hand motion can be detected by the movement of colored wrist bands. In addition, the palms of the hands can be identified

by the presence of colored regions. Therefore, the issues outlined in (1)–(4) below can be overcome by using colored gloves.

The colored gloves we designed are shown in Fig. 1. Five colors are used so as to uniquely discriminate each finger, different additional colors distinguish each wrist, and green patches locate the palms of the hands. Thus, a total of eight colors are proposed to facilitate sign language recognition.

(1) Identification of each finger and both hands
(2) Motion detection of wrists/hands and fingers
(3) Hand shape recognition
(4) Discrimination between the palm and the back of the hand.

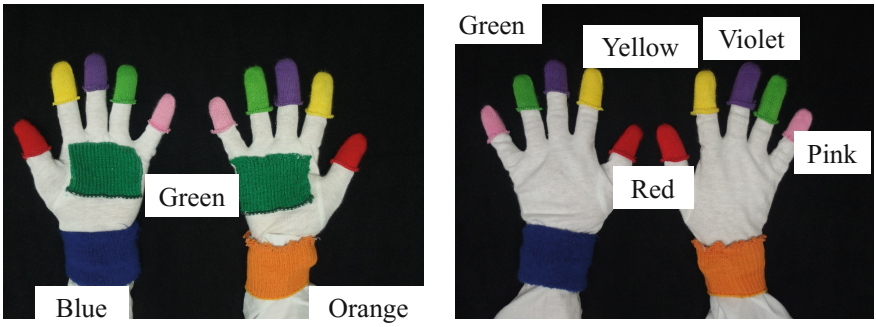


Fig. 1. Colored gloves (Color figure online)

2.2 Current Application Target

While automatic translation remains our final goal, it is too difficult to realize with current technology. Therefore, the authors are now trying to produce a kind of learning tool for sign language. Video data relevant to sign language can be obtained from a web site [9, 10]. Figure 2 shows an image from an instruction video demonstrating sign language motion. A learner memorizes the motions of each sign from this video.

However, it is quite difficult not only to memorize the motion but also to confirm the validity of the motion memorized from the video. The authors are investigating a sign language recognition method that incorporates a tool for checking the learned motion. After memorizing the sign language motion, the learner displays the same motion in front of a web camera connected to a PC. If the PC recognizes his/her motion, the result shows on the display. The learner can evaluate his/her hand and finger movements meaning this system can be used as a review tool for sign language. Although it would be ideal for the recognition success rate to be relatively high for learner's review, this cannot be achieved at this stage. Therefore, our current goal for review tool is to achieve an about 80% success rate.

Fig. 2. Instructional video for sign language motions

2.3 Motion Data Acquisition

It is important to gather accurate motion data. Therefore, the authors asked for the cooperation of the person in charge of making the motion video SmartDeaf [11] to compile the set of motion data used in this investigation. Figure 3 shows an actual motion data capturing scene. To shorten the duration, the motions of two signers were captured simultaneously. The supervisor checked the motions while the signers were performing and confirmed the accuracy of the recorded motion data after the signer had completed the motion.

The conditions under which the motion data were captured are as follows [6].

(1) Camera image resolution of 800 × 600 pixels.
(2) Illumination is set at about 200 lx for both the camera side and signer side.
(3) Frame rate is 30 fps (frames per second). This is the maximum rate for a standard Web camera and smartphone.
(4) The distance between the camera and signer is one meter, as this distance is considered to coincide with a real-life situations.
(5) The color of the signer's clothes and the background wall is black to facilitate easy detection of the colored region of colored gloves.
(6) The height in the field of view of the camera is set at a position such that the wrists of the signer cannot be detected when he/she lowers his/her arm in order to make clear the beginning and the end of a sign language motion.

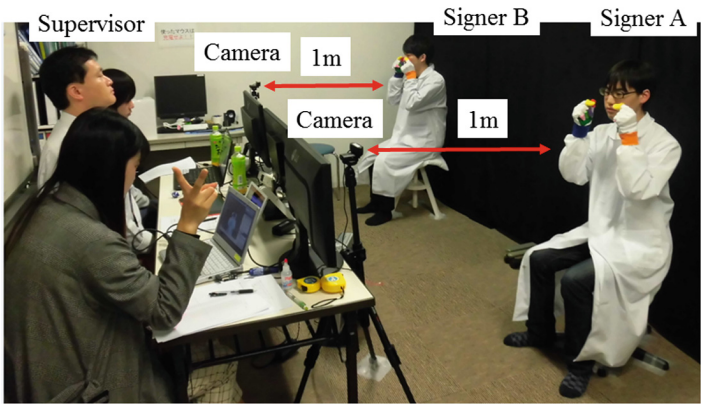


Fig. 3. Scene of data acquisition

2.4 Recognition Data Creation

2.4.1 Color Extraction and Feature Element

Each color region is detected by using colored gloves and an optical camera. An example of the detection of patterns made by color extraction is shown in Fig. 4. Their extraction is based on the hue and saturation values set in the calibration. The center of gravity of each colored region is used to specify the location of each part. The colored region size can be obtained from this image, and this size is also used for creating feature elements. The motion of the blue region assigned to the wrist can be interpreted as the motion of the entire hand.

We have tried to extract many kinds of information from sign language motion as feature elements in order to maintain a high recognition level. The feature data that identify each sign language motion is one of the crucial factors for determining recognition performance. We obtain the following feature data from the position of each colored region and the number of pixels, that is, the region size. The features and their elements are summarized in Table 1.

(1) Hand trajectory, i.e. the shape of the motion
(2) Hand position
(3) Hand velocity
(4) Hand shape, i.e. relative finger location
(5) Hand direction, i.e. hand angle
(6) Hand rotation, i.e. whether from the palm to the back of the hand or vice versa.

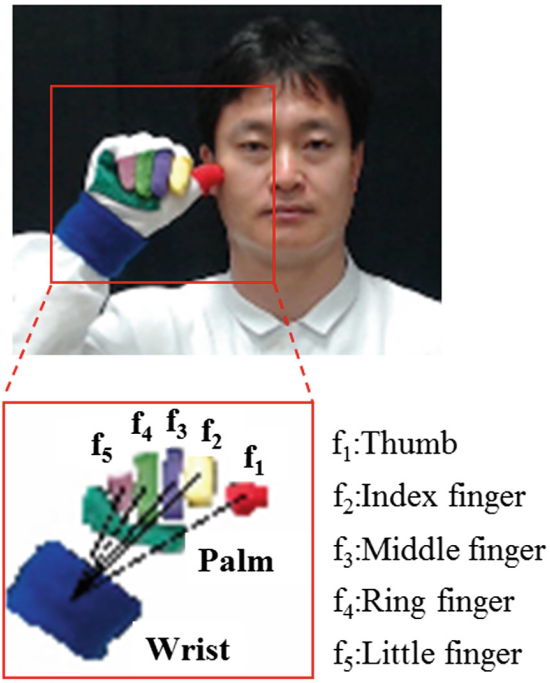


Fig. 4. Example of color extraction (Color figure online)

Table 1. Feature elements

Features	Calculation method	Number of dimensions
Trajectory	$tx_i = (x_i - \bar{x})/A$ $ty_i = (y_i - \bar{y})/A$ $A = \sqrt{\frac{1}{n}\sum_{i-1}^{n}\left((x_i - \bar{x})^2 + (y_i - \bar{y})^2\right)}$	2
Position	$px_i = x_i/800$ $py_i = y_i/600$	2
Velocity	$dx_i = x_i - x_{i-1}$ $dy_i = y_i - y_{i-1}$	2
Shape	$d_{ji} = \sqrt{(fx_{ji} - x_i)^2 + (fy_{ji} - y_i)^2}$ $j = 1, 2, 3, 4, 5$	5
Direction	$\alpha_i = \text{a tan } 2(fy_{ji} - y_i, fx_{ij} - x_i)$ j is priority order of {Index finger, Middle finger, Ring finger, Little finger, Thumb}	1
Rotation	$\beta_i = \left((Area\ of\ wrist)_i, (Area\ of\ palm)_i\right)$	2

(x, y): Wrist position (f_x, f_y): Finger position
i: Frame number, j: Finger number (1–5)

2.4.2 Preprocessing

(1) Interpolation

The center of gravity of the wrist is obtained by detecting the blue color of the wrist band. The position information of the hand is regarded as being equivalent to this wrist position. However, the position information sometimes cannot be obtained due to changes in illumination conditions resulting from movement during sign language motion, occlusion, and so on. We encountered situations where blue could not be extracted.

This problem is not merely a matter of the wrist but also affects the feature elements to be calculated using hand position. Feature elements with the exception of rotation need this information. Linear interpolation is applied when the position of the hand cannot be obtained. An example of linear interpolation is shown in Fig. 5. The hand position in x and y coordinates is shown in this figure. The interpolated values in each frame were used to calculate the feature elements.

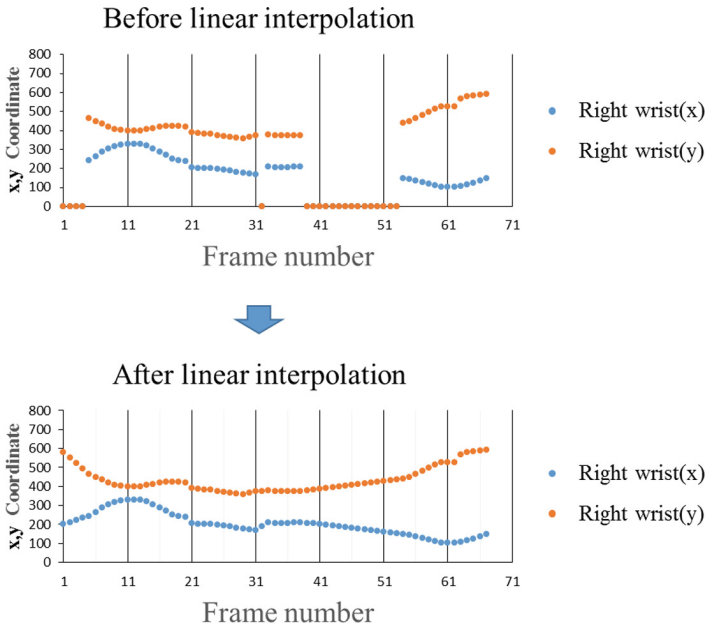


Fig. 5. Example of interpolation

(2) Element length adjustment

Besides classification by HMM, it is necessary to equalize the number of elements of the feature elements vector that is data input into the classifier. The time required for

each sign language motion is different. Moreover, even if it is the same word, the same number of vector elements is not realistic from the viewpoint of individual differences and repeatability.

The authors adjusted the number of vector elements according to the method shown in Fig. 6. Some elements in the vector were divided into groups, and their average was calculated to represent this group. This becomes a new data set having the same element length. Based on the results of a preliminary examination to ascertain recognition performance, the number of vector elements was set to ten.

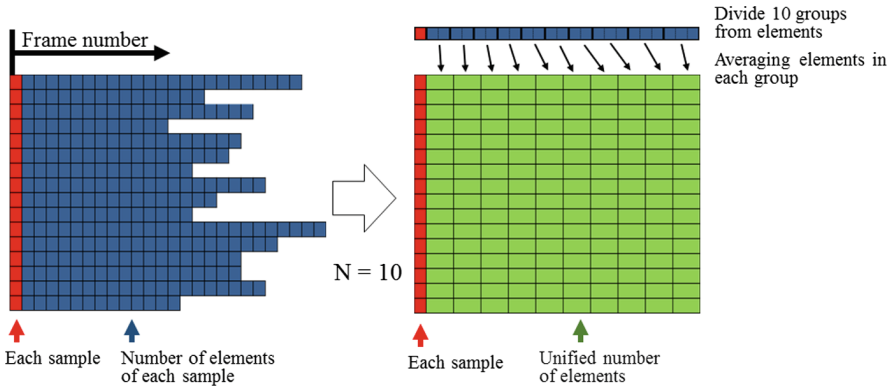


Fig. 6. Adjustment of element length

3 Recognition Method and Experiments

3.1 Recognition Method

One of our new proposals is adding two new feature elements, namely, hand rotation and hand direction, as described in Sect. 2.4.1. The other is to use the plural classifiers to enhance the discrimination performance for each sign language motion. This is based on the idea of ensemble learning [7] that classifiers work in a complementary fashion in relation to each other since each classifier is designed based on different classification criteria. If the results obtained by each classifier are integrated, the classification performance is improved by compensating for deficits in individual classifiers.

Six classifiers were selected in our proposed method. The hidden Markov model toolkit (HTK) [12] which is based on the hidden Markov model (HMM), and support vector machine (SVM) implementing LIBSVM [13], were used as classification tools. In addition, the function provided in MATLAB was used for applying other classifiers. Decision tree (DT), discriminant analysis (DA), linear classification method (LCM) and k-NN (k nearest neighbor) method were applied by using MATLAB function "fitcecoc" [14].

Figure 7 shows the recognition method we propose in this paper. Our proposal is to integrate the classification results by each feature element and each classifier. The

likelihood or probability for each sign language motion (hereinafter referred to as a "word") is obtained as the output of each classifier by using each feature element. We use six classifiers and six feature elements, therefore 36 results, that is, likelihood or probability is obtained by each classifier and feature element. The ranking is created based on this value. Integration means that the ranking (order of the candidate answer by classification results) is added up, and the recognition result is the word that obtained the smallest value.

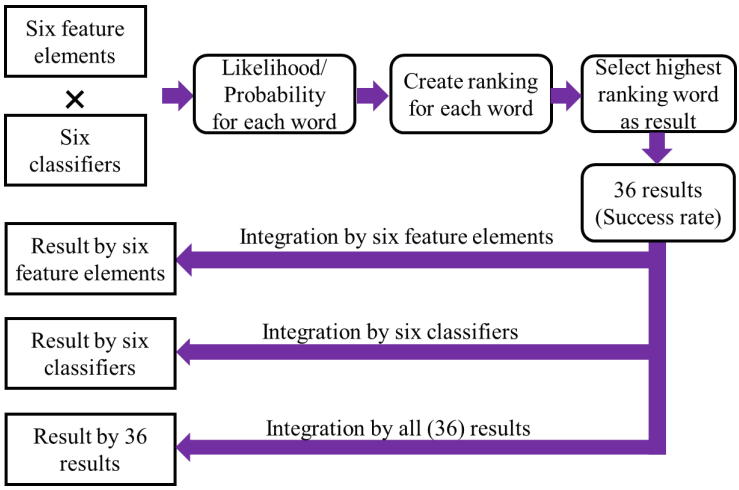


Fig. 7. Recognition method by feature elements and classifiers

3.2 Experimental Results

Thirty-five sign language words were selected for evaluation using the proposed method. Selected words are shown in Table 2. The learning video content in the SmartDeaf is divided into each category based on the usage area and occasion. Nearly 100 words are included in each category. The category of "Health and Diseases" is extremely important to hearing impaired people, therefore we selected this category. In this category, 35 words require right-hand motions only, and all these words were selected.

The data used for training and evaluation are shown in Table 3. The total of 3150 pieces of motion data, that is, 2100 pieces of motion data as training data and 1050 pieces as evaluation data, were used to evaluate the proposed method. There were 11 signers people, with 10 or 20 samples per signer and word. The training was carried out by using 20 samples signed by three signers, that is, 60 samples for each word. Ten samples from three signers were selected for evaluation. The data for training and data for evaluation were obtained from different signers. Here, in the HMM scheme, the appropriate number of the state and the initial parameters of models for each word were clarified [15] before evaluation, and these values are used in this investigation.

Table 2. Target sign language words (35 words)

1. アトピー 1. Atopic	2. おしっこ 2. Urinary	3. ガン 3. Cancer	4. コンタクト 4. Contact lens
5. 喘息 5. Asthma	6. 体調 6. Physical condition	7. ハゲ 7. Bald	8. 発熱 8. Fever
9. 病気 9. Sickness	10. 盲腸 10. Cecum	11. 顔が赤い 11. Blushing	12. カテーテル 12. Catheter
13. 禁煙 13. No smoking	14. 喫煙 14. Smoking	15. 薬を飲む 15. Take medicine	16. 呼吸 16. Breath
17. 耳鼻科 17. Otolaryngology	18. 頭痛 18. Headache	19. 摘出 19. Remove	20. 糖尿病 20. Diabetes
21. 脳卒中 21. Stroke	22. 吐き気 22. Nausea	23. 鼻水 23. Runny nose	24. 昼寝 24. Nap
25. 虫歯 25. Tooth decay	26. 命 26. Life	27. インフルエンザ 27. Influenza	28. 風邪 28. Cold
29. ギリギリ痛む 29. “Giri Giri” Painful	30. 検出 30. Detection	31. 腰 31. Waist	32. ズキンズキン痛む 32. “Zukin Zukin” Painful
33. 毒 33. Poison	34. 捻挫 34. Sprain	35. 冷や汗 35. Cold sweat	

Table 3. Data for training and evaluation

Signer	Words	Number of samples	
		Training	Evaluation
A	#1–#35	20	-
B	#1–#25	20	-
C	#1–#25	20	-
D	#1–#25	-	10
E	#1–#25	-	10
F	#1–#25	-	10
G	#26–#35	20	-
H	#26–#35	20	-
I	#26–#35	-	10
J	#26–#35	-	10
K	#26–#35	-	10
	Total of samples	2,100	1,050

The results are summarized in Table 4. The number indicates the success rate by percentage. Each number in this table was derived by each feature element and classifier. The success rate by one feature element and one classifier ranged from 17.3 to 52.8% for 35 words. The bottom row is the integration results obtained by six features, and the right-hand column shows the results obtained by six classifiers. Integration means that the ranking (the order of candidate answer) is added up, and the recognition

result is the word that obtained the smallest value. The integrated result by six feature elements and each classier ranged from 53.6 to 73.1%, the result by six classifiers and each feature element ranged from 33.6 to 51.8%, and the result for total integration, that is, the result obtained by six classifiers and six feature elements was 76.2%. It was verified that integrating each feature element contributes significantly to raising recognition performance. A 3.1% performance enhancement was achieved, which was the difference between the integrated result (76.2%) and the result by SVM (73.1%), which was the best result for a single classifier.

For all classifiers, the integrated result by feature elements improved classification performance markedly, demonstrating that this method can effectively enhance performance. Although an outstanding outcome cannot be obtained by combining each classifier compared with that of integrating feature elements, enhancement can be brought about by integrating classifiers.

Table 4. Recognition result by integration of highest ranking results

Training data A, B, C, J, K & evaluation data D, E, F, L, M, N							
Feature elements	Classifiers						
	HMM	SVM	DT	DA	LCM	KNN	Classifier integrated results
Trajectory	17.3	46.4	34.2	48.5	42.8	43.5	42.7
Position	25.4	50.5	42.5	52.8	44.4	47.1	51.4
Velocity	24.0	46.8	36.6	51.7	46.1	47.0	51.8
Shape	26.8	46.2	38.4	45.5	40.4	42.0	50.6
Direction	18.8	29.4	27.6	29.9	28.1	33.3	33.6
Rotation	14.4	42.3	35.1	42.4	42.5	38.4	46.0
Element integrated results	53.6	73.1	70.1	72.5	70.8	71.8	**76.2**

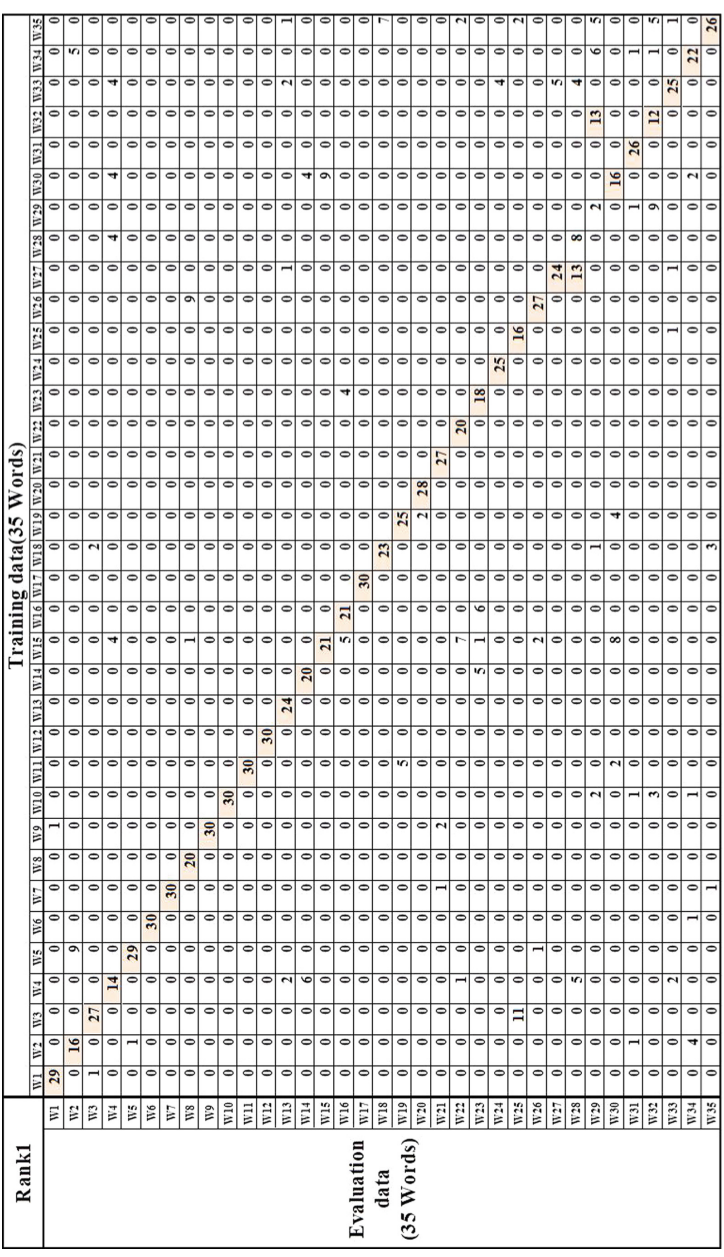
4 Discussions for Experimental Results

The confusion matrix for the integrated six elements and six classifiers is shown in Table 5. In the case that ranking one is not unique after integration, summing up was not conducted in making this table. Therefore, the summation value of some rows, that is, Words 24, 25, 27 and 29 is not 30. Each word has the highest probability as a recognition result except Words 28 and 29. Good results could not be obtained for these words, the number of correct answers was 8 and 2 for 30 samples, that is, three signers and 10 samples respectively.

Word 28 was mistakenly taken for Words 27 and 33. The representative scenes of these words are shown in Fig. 8. The differences among these three words involve small motions conducted in front of the face. The little finger is bent just in Word 27.

Word 29 was mistakenly taken for Word 32. The scenes where these words were demonstrated are shown in Fig. 9. The difference in how these two words are signed is

Table 5. Confusion matrix for the 35 targeted words



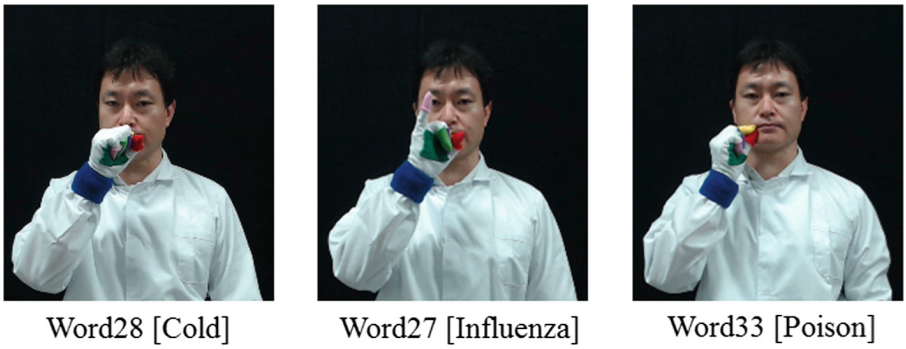


Fig. 8. Representative shots of words where performance was poor (1)

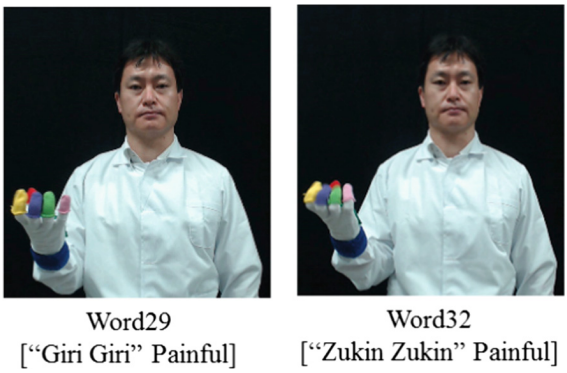


Fig. 9. Representative shots of words where performance was poor (2)

only the motion speed of the fingers. Since these motions closely resemble each other, the proposed six feature element cannot classify these words. A new feature element seems to be necessary to discriminate these words.

If we assume that the correct answer is included in the top three rankings for the 35 words, in other words, if we include the third lowest value as a correct result, the result as shown in Table 6 was obtained. The integrated result was raised from 76.2%, shown in Table 5, to 94.1%. This result demonstrates that each recognition result by each feature element and classifier is a good one even though it cannot get the top ranking. This value suggests that the proposed method could be used as a review tool for assessing how well learners have mastered sign language motions.

Table 6. Recognition result by integration of top 3 results

Training data A, B, C, J, K & evaluation data D, E, F, L, M, N							
Feature elements	Classifiers						
	HMM	SVM	DT	DA	LCM	KNN	Classifier integrated results
Trajectory	31.0	69.3	61.5	69.6	67.3	65.1	66.9
Position	52.0	77.1	64.3	80.6	68.9	74.5	76.8
Velocity	48.3	70.6	58.9	74.2	71.2	73.1	76.0
Shape	56.8	69.6	66.0	67.6	63.3	66.6	74.5
Direction	39.1	48.4	57.3	46.8	46.3	57.7	57.9
Rotation	35.5	70.8	64.6	70.2	70.2	65.1	72.9
Element integrated results	79.0	90.8	90.1	90.4	88.1	91.5	**94.1**

5 Conclusion

The enhancement of recognition performance for the motions used in sign language is described in this paper. The main feature of the proposed method is integrating the results of six feature elements and six classifiers in order to accurately characterize and discriminate among the sign language motions. The six feature elements were obtained by color extraction from colored gloves and a wrist band. The elements of trajectory, position, and velocity are obtained from the center of gravity of the blue regions of the wrist band. New feature elements were added, and the hand direction was obtained from the angle between each fingertip and the wrist. The hand rotation is calculated from the region size of wrist and palm. Each element is applied to six classifiers to discriminate each motion. The integrated result of six feature elements and six classifiers was 76.2%. In the current investigation, the classification limitation was that six feature elements cannot express the difference in the motions of groups 27, 28 and 33 and groups 29 and 32. However, if we take the top three rankings as a correct result, the integrated success rate for 35 words was increased to 94.2%. This value suggests that the proposed method is a feasible review tool for learners to validate the accuracy of their sign language movements. However, there were four words for which the success rates were approximately 30% or less. The low performance for these words must be resolved if overall performance is to be improved.

References

1. 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)
2. Zafrulla, Z., Brashear, H., Starner, T., Hamilton, H., Presti, P.: American sign language recognition with the Kinect. In: Proceedings of the 13th International Conference on Multimodal Interfaces, pp. 276–286 (2011)

3. Jitcharoenpory, R., Senechakr, P., Dahlan, M., Suchato, A., Chuangsuwanich, E., Punyabukkana, P.: Recognizing words in Thai Sign Language using flex sensors and gyroscopes. In: i-CREATe2017, 4 p. (2017)
4. Channaiah Chandana, K., Nikhita, K., Nikitha, P., Bhavani, N.K., Sudeep, J.: Hand gestures recognition system for deaf, dumb and blind people. IJIRCCE **5**(5), 10058–10062 (2017)
5. Singha, J., Das, K.: Hand gesture recognition based on Karhunen-Loeve transform. In: Mobile & Embedded Technology International Conference 2013, pp. 365–371 (2013)
6. Ozawa, T., Shibata, H., Nishimura, H., Tanaka, H.: Investigation of feature elements and performance improvement for sign language recognition by hidden Markov model. In: Antona, M., Stephanidis, C. (eds.) UAHCI 2017 Part II. LNCS, vol. 10278, pp. 76–88. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58703-5_6
7. Dietterich, T.G.: Ensemble methods in machine learning. In: Kittler, J., Roli, F. (eds.) MCS 2000. LNCS, vol. 1857, pp. 1–15. Springer, Heidelberg (2000). https://doi.org/10.1007/3-540-45014-9_1
8. 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.) HIMI 2013 Part I. LNCS, vol. 8016, pp. 133–142. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39209-2_16
9. NHK (Japan Broadcasting Corporation), NHK Sign Language CG. http://cgi2.nhk.or.jp/signlanguage/
10. Signing Savvy | ASL Sign Language Video Dictionary. https://www.signingsavvy.com/
11. KCC Corporation, Smart Deaf. http://www.smartdeaf.com/
12. HTK version 3.4.1. http://htk.eng.cam.ac.uk/
13. LIBSVM. https://jp.mathworks.com/help/stats/fitcecoc.html/
14. MATLAB. https://jp.mathworks.com/help/stats/fitcecoc.html/
15. Okayasu, Y., Ozawa, T., Dahlan, M., Nishimura, H., Tanaka, H.: Performance enhancement by combining visual clues to identify sing language motions. IEEE Pacific Rim Conference, 4 p. (2017)