The Seventeenth International Symposium on Artificial Life and Robotics 2012 (AROB 17th ’12),
B-Con Plaza, Beppu, Oita, Japan, January 19-21, 2012

Japanese Finger-spelling Recognition Using a Chest-mounted Camera

Akira Nagasue, Joo Kooi Tan, Hyoungseop Kim and Seiji Ishikawa

Kyushu Institute of Technology, Kitakyushu 804-8550, Japan
Tel: 093-884-3191, Email: nagasue@ss10.cntl.kyutech.ac.jp

Abstract: This paper proposes a technique for recognizing Japanese Finger-spelling using a sign language user’s chest-mounted camera. Unlike existent systems, the technique employs a sign language user’s chest-mounted camera attached to the sign language user himself and recognizes his/her sign language through the captured images of Japanese Finger-spelling. We use a silhouette picture of a hand, and MHI (Motion History Image) for Japanese finger-spelling recognition. The recognition method uses the eigenspace method. Furthermore, in order to recognize Japanese finger-spelling from an animation (image sequence), a character segmentation technique is also proposed. Finally, finger characters are recognized and the performance is shown experimentally.

Keywords: Sign language recognition, Chest-mounted camera, Motion history images, Eigenspace method.

1. INTRODUCTION

Recently, requests on computer-based welfare assistance systems have been increasing. In order to develop these systems, a great number of researches aim the development of automatic Japanese sign language recognition systems using image processing. However, conventional recognition systems deal with those input images obtained from a camera facing a disabled person in hearing and speaking [1] or using many sensors [2, 3]. But, normally, we do not possess this kind of system including a camera in our daily life, since it is hardly the case that we happen to meet such disabled people who wish to talk to us. Thus this kind of system would not spread for practical use in our daily life. On the contrary, if a Japanese sign language user is equipped with such a system, it will definitely improve his/her quality of life (QOL), as he/she can communicate to everyone to whom he/she wishes to do so.

In this paper, we propose a system that is held by a sign language user for recognizing his/her sign language. Unlike existent systems, the proposed system employs a chest-mounted camera shown in **Fig. 1** which is held by a sign language user and recognizes his/her Japanese finger-spelling as shown in **Fig. 2** through the captured images of his/her finger shapes. We think this system can be employed as a new human interface for not only Japanese sign language users but also healthy persons.

Fig. 1. Location of the chest-mounted camera

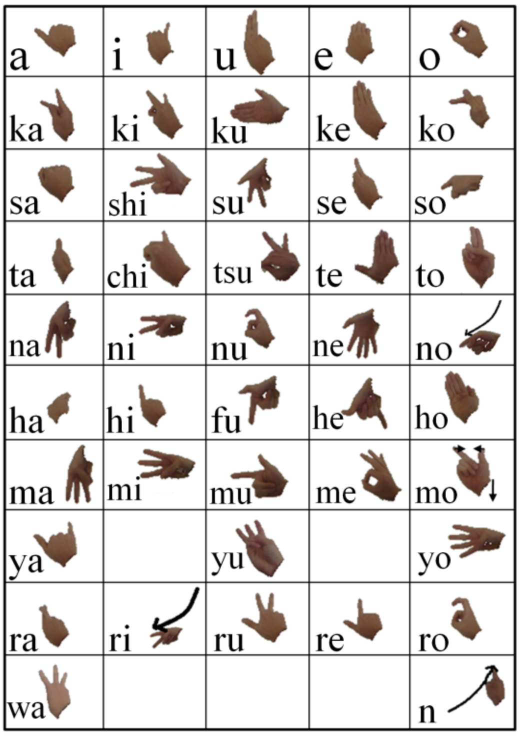


Fig. 2. Japanese finger-spelling list

2. PROPOSED METHOD

In the first step, we extract a hand area from an image taken from a chest-mounted camera. Next, we recognize the Japanese finger-spelling performed by the user employing the extracted hand area as shown in **Fig. 3**.

(a) (b)

Fig. 3. Skin color extraction:
(a) Input image, (b) extracted hand.

2.1 Character recognition

Recognition methods for Japanese finger-spelling are usually composed of a nearest neighbor search in the eigenspace using a hand silhouette for non-moving finger-spelling and nearest neighbor search in eigenspace using a Motion History Image (MHI) [4] and hand silhouettes for moving finger-spelling.

MHI is defined by the following formula.

$$H_\tau(x,y,t)=\begin{cases}\tau & \text{if } D(x,y,t)=1 \\ \max(0,H_\tau(x,y,t-1)-1) & \text{otherwise}\end{cases} \quad (1)$$

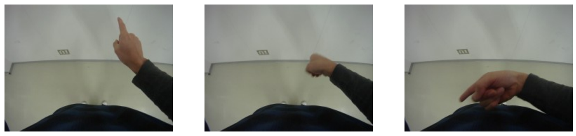
Here x is the abscissa of an image, y is the ordinate of the image, t expresses time, and τ is a parameter which adjusts time that has passed in the sequence. An example of the expression of a sign language letter by a MHI is shown in **Fig. 4**. Examples of the images used for recognition is shown as well in **Fig. 5**.

Based on this strategy, the recognition of a finger character can be performed regardless of whether motion exists or not.

2.2 Character segmentation

Finger character segmentation in an image sequence is also performed by considering inter-frame difference using hand silhouette images. When the number of moving pixels in the inter-frame difference exceeds a threshold value of motion (Th_M), it is recognized as the frame in which the hand has moved, and when the number of pixels does not exceed the threshold value, it is recognized as the frame which has a stopped (static) hand, i.e., the frame which shows the finger character.

This finger character representation frame is counted. If the number of the frames exceed the threshold value (Th_F), the finger character is recognized as that without a motion. Otherwise it is recognized as that with a motion.



(a)

(b)

Fig. 4. Examples of MHI: (a) Input images, (b) MHIs.

(a) (b)

(c) (d)

Fig. 5. Examples of database images: (a) Hand silhouette image "a", (b) MHI and hand silhouette image "n", (c) database image "a", (d) database image "n".

3. EXPERIMENTAL RESULTS

3.1 Experiment of Japanese finger-spelling recognition

In the first place, in order to perform the experiment on Japanese finger-spelling recognition, we took the videos of 45 Japanese finger-spelling characters five times. We used four sets for learning, and one set for recognition. This process was repeated five times for one person, and it was repeated with respect to five persons. The recognition rate is computed from the following formula.

$$Recognition\ rate = \frac{CR}{ALL} \quad (2)$$

Here CR is the number of correctly recognized characters, whereas ALL is the number of all characters.

Finally the proposed method achieved 89.1% (1002/1125) of the recognition rate. A detailed result is shown in **Table 1** and **Fig. 6**.

It is thought that incorrect recognition arose in the characters in which the shape or motion are similar with each other, e.g. "no" and "ri". It is shown in **Fig. 7**.

Table 1. Individual recognition rates

	Person A	Person B	Person C	Person D	Person E
Data 1	91.1%	93.3%	75.6%	88.9%	82.2%
Data 2	95.6%	84.4%	91.1%	82.2%	88.9%
Data 3	95.6%	93.3%	88.9%	86.7%	77.8%
Data 4	100.0%	95.6%	95.6%	84.4%	86.7%
Data 5	86.7%	93.3%	91.1%	84.4%	93.3%
Average	93.8%	92.0%	88.4%	85.3%	85.8%

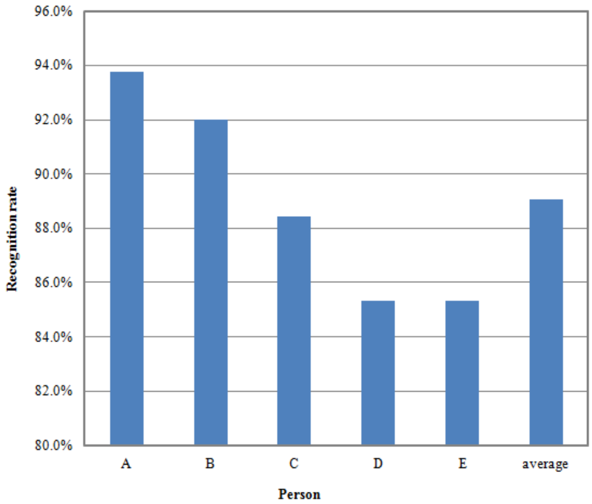


Fig. 6. Average recognition rates of five persons.

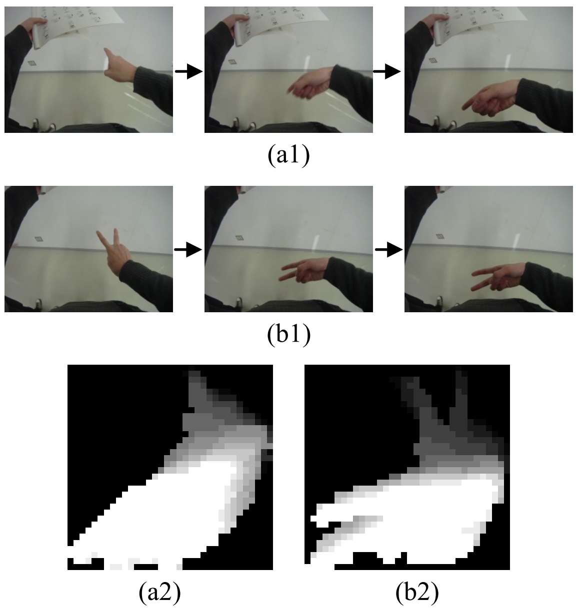


Fig. 7. Examples of similar character:(a1) Input images "no", (b1) input images "ri", (a2) database image "no", (b2) database image "ri".

3.2. Experiment of character segmentation

In the second place, we performed the experiment on character segmentation. We took videos of 6 sentences using Japanese finger-spelling characters four times. They are 'kon-nichiwa', 'ohayo', 'kore ikura', 'itsu kitano' and 'toki wa kanenari'. An evaluation method is described in the following.

First, the correct answer key-frame which should certainly be contained is set to the character representation section by manual for every character of a text. When the correct answer key-frame is contained in the detected character representation section, it is considered as success (TP: True Positive), and when more than one are contained with the case where it is not contained, it is considered as failure (FP: False Positive). Moreover, the finger character accompanied by a motion or a finger character without a motion is distinguished among success (TP), and success is defined as R (Right). It evaluates using right display rate ($success$) and distinction rate ($decision$). The following formula defines $success$ and $decision$.

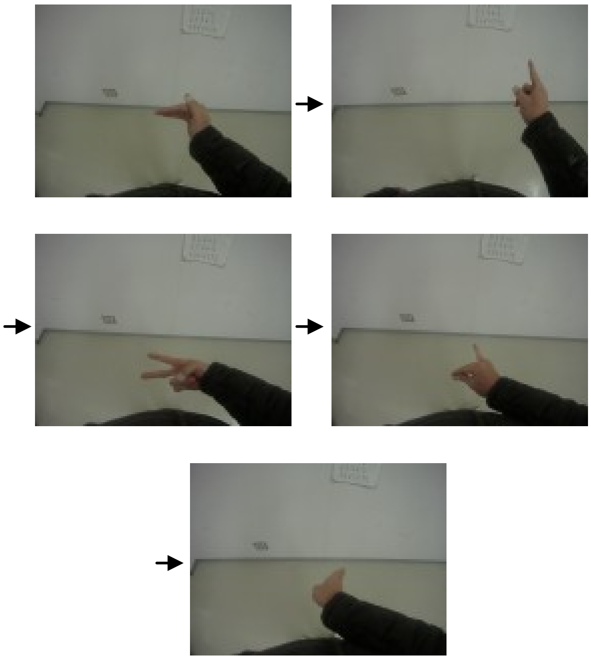
$$success = \frac{TP}{NUM} \tag{3}$$

$$decision = \frac{R}{TP} \tag{4}$$

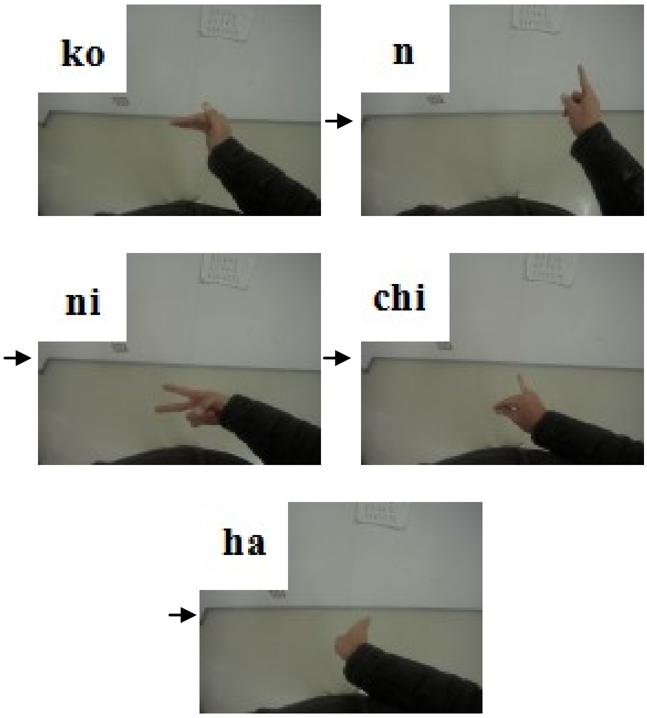
Here NUM is the number of characters in a text.

The proposed method segmented the sentences at a rate of 90.8% ($success$). Among these, 77.8% of the letters were correctly recognized concerning the presence or absence of movement ($decision$).

Since the time to display a character differs from person to person, the existence of a motion in a character has become less discriminative feature. Also, since the motion was distinguished using the number of change pixels, it is thought that the size of the hand being reflected has caused errors. The result of the text recognition using both Japanese finger-spelling recognition method and character segmentation method is shown in **Fig. 8**.



(a)



(b)

Fig. 8. The recognition result of "kon-nichiwa": (a) Part of input animation, (b) part of output animation.

4. CONCLUSIONS

We proposed a Japanese finger-spelling recognition method from the chest-mounted camera images of a Japanese sign language user. The performance of the method was satisfactory. However, there still exists a problem in the accuracy of character segmentation. As future work, we are going to improve the character segmentation method. Application to real time recognition is planned and a voice sound and P-sound recognition method will be investigated as well.

REFERENCES

[1] Liwicki, S., Everingham, M. (2009), Automatic recognition of fingerspelled words in British Sign Language. 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Workshops, pp.50-57.

[2] Brashear, H., Henderson, V., Park K. H., et al (2006), American sign language recognition in game development for deaf children. Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility, pp.79-86.

[3] Brashear, H., Starner, T., Lukowicz, P., et al (2003), Using multiple sensors for mobile sign language recognition. Seventh IEEE International Symposium, pp.45-52.

[4] Bobick, A. F., Davis, J. W. (2001), The recognition of human movement using temporal templates. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.23, No.3, pp.257-267.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI (22650022) which is greatly acknowledged.

