Automatic Synthesis of Training Data for Sign Language Recognition using HMM

Satoko SHINJO , Yutaka YAMADA , Tadashi MATSUO , Yoshiaki SHIRAI , Nobutaka SHIMADA
Ritsumeikan University, Shiga, JAPAN

***Abstract*- For recognition of a sign language, many samples are required for learning by Hidden Markov Model (HMM). However, acquisition of many samples is difficult. This paper proposes a method to synthesize learning data from a combination of real data. We make a database of appearance of each hand extracted from video images. New data is created by exchanging a hand shape with another similar shape in the database. Recognition experiments show the effectiveness of the proposed method.**

I. INTRODUCTION

Sign language is a communication tool for hearing-impaired people. Because few hearing people can use a sign language, a system is needed to translate the sign language into natural language.

In our method, hand shape features are extracted from video images because use of special equipments such as data gloves [1] is troublesome.

Sign language words are trained and recognized using Hidden Markov Model (HMM) [2]. In the training phase of HMM, image sequences are automatically segmented into states based on features of hand motion and probability distributions of features for each state is computed. For building reliable HMM, many training samples are necessary. However, collecting many samples from different subjects is difficult. Therefore we propose a method of synthesizing training data automatically from a limited number of actual samples.

Here, we focus on the hand shape. Fukuda [3] et al. proposed a sign language classification based on the hand shape. Because the 3-D hand shape is not easily extracted, the hand contour in the image is used as features. Then one hand shape may correspond to multiple contour shapes.

First, the database of sign language word and its constituent contour shapes is generated from sign language images. Then the training data of each word is synthesized by replacing the contour shapes with the same shape in other words.

II. THE RECOGNITION SYSTEM FOR SIGN LANGUAGE

Fig. 1 is summary of the recognition system using HMM for sign language. First face and hands regions are detected using skin color from video images. Since the colors of the face and hands change depending on the environment and the subject, the range of the skin color is determined from the initial image, where the person region is extracted by background subtraction and skin regions are extracted at fixed positions of the person region. Next, features of sign language are extracted. Features of sign language consist of the position, velocity, and shape of hand regions [4]. Then generate the initial HMM and training HMM. Here we use a left-to-right HMM, where each state has a normally-distributed output probability. The image sequence is segmented into several states on the basis of the motion of hands [4]. Those states are used to generate the states of the initial HMM. Initial HMM is used to generate the training model. Then test data is recognized. We select one as result with the highest likelihood.

III. CLASSIFICATION OF HAND SHAPE LANGUAGE

Here we subdivide sign language by hands shape using expand [3], collect hands shape data and create database (Fig. 2).

A. *The classes by finger signs indicating phones*

Fukuda [3] et al. proposed a classification of the hand shape used in sign language. This classification is based on the finger alphabet of the Japanese syllabary. In order to deal with words which are not expressed by those alphabets, we add three hand shapes.

B. *The classes subdivided by the appearance of hands in image*

Since the view of a hand in images changes depending on a hand position and pose, we further extend the classification to include typical variations of the same hand shape. Fig. 3 shows the extended classification.

C. *Making a hand shape database*

The shape is important when a hand is not moving. In a static period, therefore, if the hand shape is almost constant, the hand shape features are saved in the database.

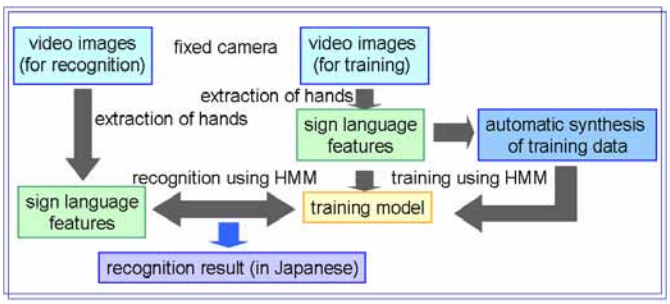

Figure1. The recognition system for sign language.

It is difficult to detect the hand shape change from the number of protrusions because it changes depending on the small change of the hand direction or on the slight change of the finger directions. Therefore, the degree of circularity is used: if it is constant in the static period, the hand shape is considered to be constant. The degree of circularity, the direction of the principal axis, and the number of protrusions are saved as the features of the corresponding hand shape.

When there is a frame that segments the previous and the next moving period (see Ⓐ in Fig4. (b) for example), if the degree of circularity is constant in the vicinity of the frame, the features about hand shape are also saved.

Collecting data of each hand shape in this way, database is generated. For example, the word in Fig. 4 (a) and the word in Fig. 4 (b) have the same class of the hand shape. Therefore, they can exchange hand shape in the database.

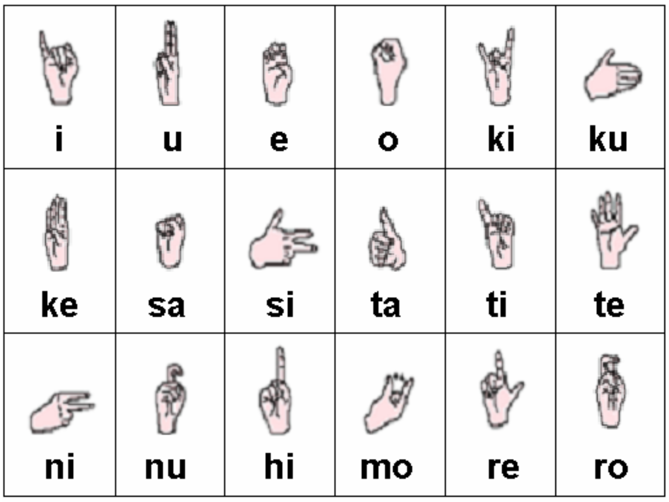

Figure2. The classes by finger signs indicating phones.

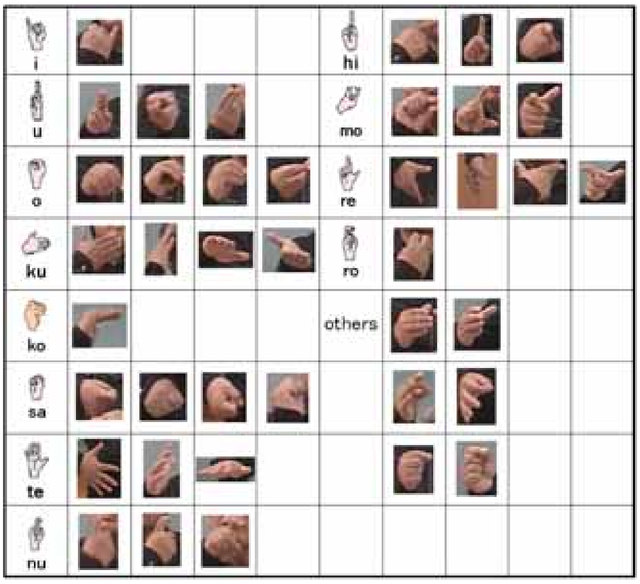

Figure3. The classes subdivided by the appearance of hands in image.

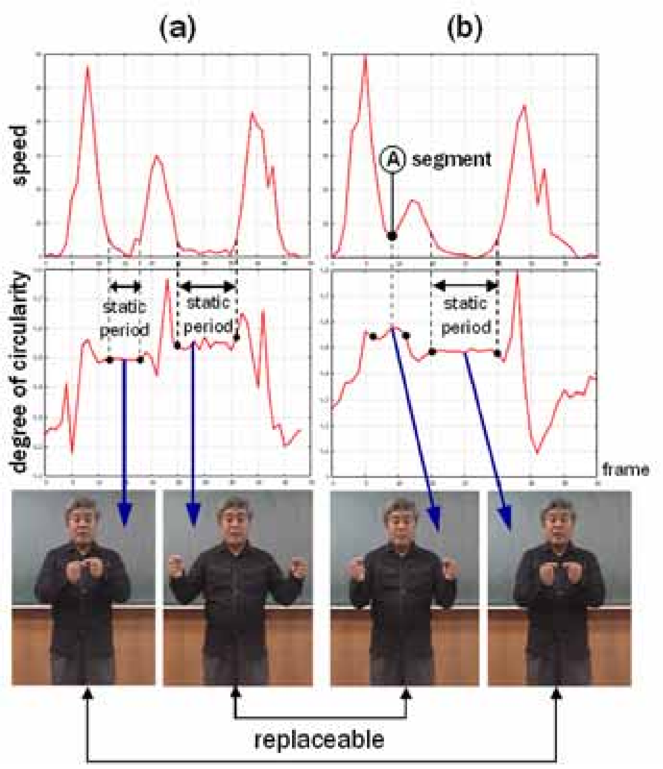

Figure4. The extraction of intervals with stable shapes by circularity.

IV. Automatic Synthesis of Training Data

In order to synthesize a training data of a word, the degree of circularity is checked to find constant periods. Then, each hand shape the database is searched for the similar hand shape. If it is found, the shape features replace the original features. The other features such as the position and the motion are kept unchanged. Outline of this process is shown in Fig. 5.

However, even if the shape and the view of a hand are the same for different words, the direction of the principal axis may be different. In this case, the shape is not exchanged: the shape exchange is allowed only when the hand shape and direction are the same.

V. Experimental Results

Experiments of recognition are performed using the HMM generated from original image sequences and the synthesized data. In the experiments, images of two persons are used. For each word, three sequences of images are obtained and the total of six samples is obtained. Excluding words whose images are not successfully processed, 14 words expressed by both hands and 21 words expressed by one hand are used. One sequence is used for recognition. For training two sequences one selected from each person excluding the one for recognition the sequence for recognition (in total 4 sequences). By changing the sequence for the recognition, six experiments are performed.

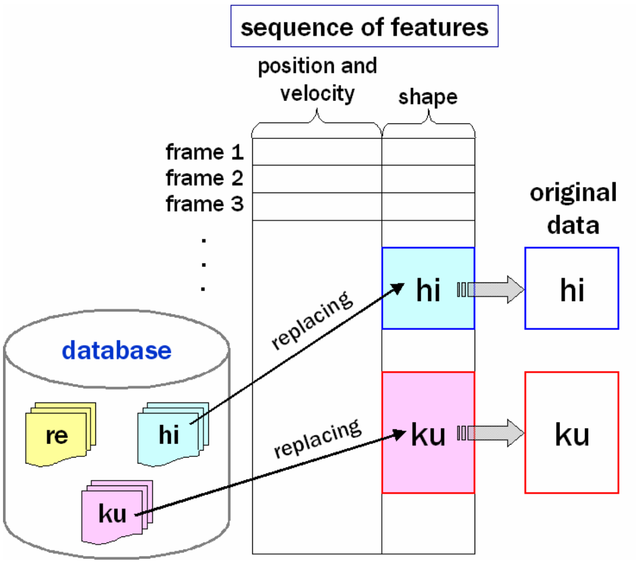

Figure5. The replacement of shape features.

A. *The recognition result by using at most one synthesized data in learning*

First, at most one synthesized data is used for training. Synthesized data selected at random. Experimental results compare only original data used for training with at most one synthesized data are shown in Table. 1.

Table 1 shows recognition rates of the experiments using only actual data and the experiments using synthesized data have little differences.

TABLE I The recognition result by using at most one synthesized data in learning.

(A) BOTH HANDS (14WORDS X 3RECOGNITION DATA X 4 INITIAL MODEL=168)

Synthesized data	use		Not use	
Recognition data	A	B	A	B
The number of success	163	158	148	146
Recognition rate	97.0%	94.0%	88.1%	86.9%

(B)ONE HAND (21WORDS X 3RECOGNITION DATA X 4 INITIAL MODEL=252)

Synthesized data	use		Not use	
Recognition data	A	B	A	B
The number of success	230	222	229	219
Recognition rate	91.3%	88.1%	90.9%	86.9%

VI. CONCLUSION

In this research, we proposed a method of automatically synthesizing of training data using a hand shape database for sign language recognition. Hand shapes in sign language words are classified and features of each hand shape are saved in the database. Training data are synthesized by replacing hand shape features in the database.

Although experiments with a small number of sign language words, the effect of using synthesized data is not prominent, the method seems promising for larger data set.

The future problems include extension to a larger data set, and to collecting more images of different persons. In addition to the use of the hand shape, the hand position and motion should be included.

REFERENCES

[1] H.Sagawa, M.Takeuchi, ``A Method for Recognizing a Sequence of Sign language Words Represented in Japanese Sign Language Sentence", Proc. Int. Conf. Automatic Face and Gesture Recognition (FG2000), pp.~434--439, 2000.

[2] J. Kinscher, H. Trebbe, ``The Munster Taging Project - Mathematical Background", Arbeitabereich Linguistik, University of Munster, D-58149 Munster, 1995.

[3] Y. Fukuda, H. Kimura et al. ``Expression through hand and finger actions used in `words in continuous signing': A study in reference to the dialogue in Japanese Sign Language by the deaf persons", Technical Report of IEICE, Vol. 97, No. 586, pp. 81--85, 1998 (in Japanese).

[4] K. Kawahigashi, Y. Shirai, J. Miura, and N. Shimada: Automatic Synthesis of Training Data for Sign Language Recognition Using HMM, Proc. 10th International Conference on Computers Helping People with Special Needs, pp.623-626, 2006