SIGN LANGUAGE TRANSLATION SYSTEM USING CONTINUOUS DP MATCHING

Hirohiko Sagawa, Hiroshi Sakou, Masahiro Abe
Central Research Laboratory, Hitachi, Ltd.

1-280, Higashi-Koigakubo, Kokubunji-shi, Tokyo 185, Japan

ABSTRACT

We developed a prototype sign language translation system which translates sign language into spoken language. Sign language is input to a computer using a DataGlove. Words in the sign language sentence are recognized by continuous dynamic programming (DP) matching. Continuous DP matching has been improved to decrease recognition time by matching compressed patterns. Using this method, the word recognition rate was 97.3% and the recognition is approximately 16 times faster than without compression. The system then translates the words to sentences of spoken language by adding particles and inflections.

1 INTRODUCTION

Sign language is the usual method of communication for deaf people[1]. Deaf people and hearing people generally communicate through a sign-language interpreter. However, when they wish to have privacy, such as when in hospital or in a police station, they need to communicate directly with hearing people without a sign-language interpreter. Furthermore, sometimes they cannot speak when they wish, since the number of interpreters is limited and their services are in great demand. Therefore, there is an increasing need for automatic translation of sign language.

There has already been research into sign-language recognition[2, 3, 4, 5]. The objects to be recognized are words represented by static patterns such as different combinations of finger positions, and simple linear motions such as up, down, left, and right. However, continuous recognition of sign language is not yet possible. In real conversations, many things are represented by a continuous combination of dynamic patterns, and these are vital for maintaining the conversation flow. It is, therefore, important to regard sign language as dynamic patterns.

In this paper, we present a sign language translation system that uses continuous DP matching. We also describe a method of compressing sign language patterns and matching compressed patterns in order to decrease recognition time.

2 STRUCTURE OF SIGN LANGUAGE TRANSLATION SYSTEM

The structure of the system is shown in Fig. 1.

The object of this system is Japanese sign language. There are three types of Japanese sign language as follows[6].

1. Traditional Sign Language

   This type is quite different from the Japanese language.

2. Japanese-based Sign Language

   This type corresponds exactly to the Japanese language.

3. Composite Sign Language.

   This type borrows its words from the traditional sign language and the order of words is similar to the Japanese language.

Our system translates the composite sign language. The pattern of the sign language is input to a computer using a DataGlove[7]. The computer represents the sign language as a continuous pattern of 32-dimensional vectors. Words in sign language are recognized by continuous dynamic programming (DP) matching[8]. This is discussed in the next section in detail.

After matching, the words in the input pattern are detected and combined according to their overlap to generate the candidates of a sentence.

If the candidate is a correct sentence, the sequence of words translates to a sentence. To generate the sentence, first, the attributes (Fig. 2) and case-frame (Fig. 3) of each word in the candidate are read. Next, the relations of the words are analyzed based on the attributes and case-frame. Finally, the particles and inflections are added to words based on their relations and supplement rule (Fig. 4).

Translated sentences are displayed on the screen.

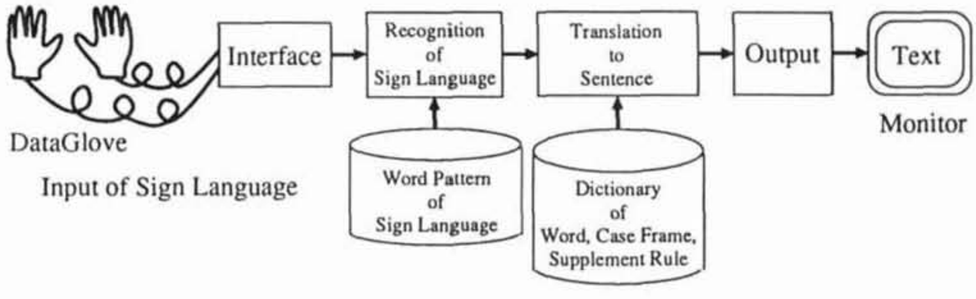

Figure 1: Structure of sign language translation system

Word	Part of Speech	Meaning
stomach	Noun	<Body>
weight	Adjective	<Status>
feel	Verb	<Motion>

Figure 2: Example of word dictionary

Verb	Relation	Meaning	Particle
feel	objective	<Status, Motion, Existence, Change>	"wo"
open	objective	<Thing, Body>	"wo"
	subjective	<Man>	"ha"

Figure 3: Example of case-frame dictionary

Condition	Part of Speech of Source = Noun, Pronoun, Proper Noun
	Part of Speech of Destination = Verb
	Relation = Subjective, Time, Place, Objective
Supplement	Word + Particle

Figure 4: Example of supplement rule

3 SIGN LANGUAGE RECOGNITION METHOD

3.1 Continuous DP Matching

We used continuous DP matching to recognize the sign-language words in the input pattern. DP matching is usually used in the field of speech recognition. We chose this method, since the data from the DataGlove is similar to the spectrum pattern of speech.

In DP matching, constraints such as in Fig. 5 are used to find the corresponding points of patterns. This constraint is called the DP path. The distance between patterns is calculated according to the DP path. In continuous DP matching, especially, the distance is sequentially calculated along the time axis of the input pattern (Fig. 6). This gives sufficient recognition accuracy, but the recognition time increases in proportion to both the number of words and the length of input pattern when conventional continuous DP matching is used. To decrease the recognition time, we examined (1) the compression of the sign-language patterns based on their dynamic features, and (2) a matching method based on comparing the compressed patterns.

3.2 Compression of Sign Language

To compress patterns, two kinds of features are detected as follows.

1. the position when velocity is minimum
2. the position when the change of vector direction exceeds a threshold (Eq. 1)

$$\mathbf{v}(t) = \mathbf{p}(t) - \mathbf{p}(t-1)$$
$$\sum_{t=T_i}^{T_j} \theta(\mathbf{v}(t), \mathbf{v}(t+1)) \geq \Theta \quad (1)$$

$$\begin{aligned} \mathbf{p}(t) &= \text{pattern vector at } t \\ \mathbf{v}(t) &= \text{velocity vector at } t \\ \theta(\mathbf{v}(t_1), \mathbf{v}(t_2)) &= \text{angle between } \mathbf{v}(t_1) \text{ and } \mathbf{v}(t_2) \\ \Theta &= \text{threshold} \end{aligned}$$

Patterns are stored as detected points and the times to move between them. By this method, patterns were closely modeled by a few points in the original trajectory of the patterns.

3.3 Compressed Continuous DP Matching

To match compressed patterns, the constraints for the DP path are set loosely and the time between feature points is used as the weight for each path (Fig. 7). The distance between corresponding feature points in the input pattern and the word pattern is defined as the sum of the usual vector distance (Euclidean distance) and a value based on the ratio of the time intervals between feature points (Eq. 2).

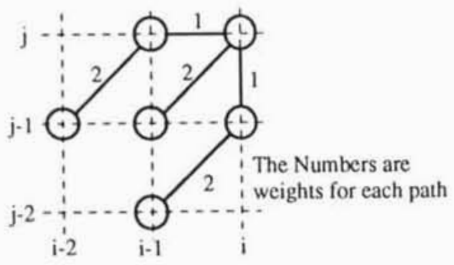

Figure 5: Example of DP path

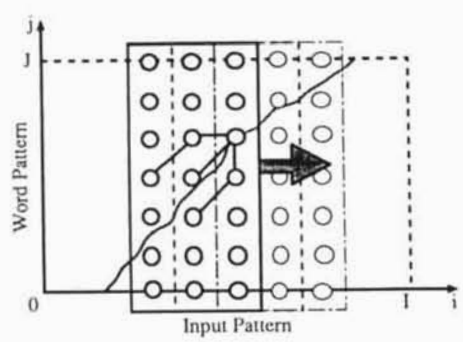

Figure 6: Principle of continuous DP matching

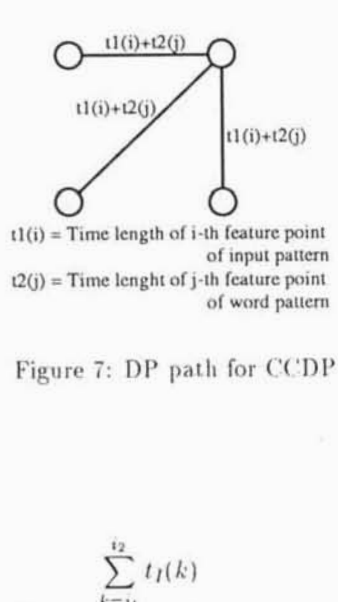

Figure 7: DP path for CCDP

$$
\begin{aligned}
r &= \frac{\sum_{k=i_1}^{i_2} t_I(k)}{\sum_{k=j_1}^{j_2} t_J(k)} \\
d_1(i,j) &= \text{(distance between elements)} \\
d_2 &= \begin{cases} (r-1.0)^2 & ;\text{if } r \geq 1.0 \\ (\frac{1}{r}-1.0)^2 & ;\text{if } r < 1.0 \end{cases} \\
d(i,j) &= w_1 \times d_1(i,j) + w_2 \times d_2 \qquad (2)
\end{aligned}
$$

w_1, w_2 = weight

i_1, i_2 = top and end of input corresponding to j

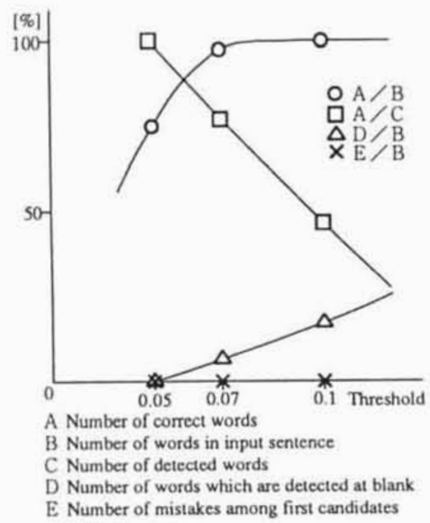

Figure 8: Relation between threshold and recognition rate (CDP)

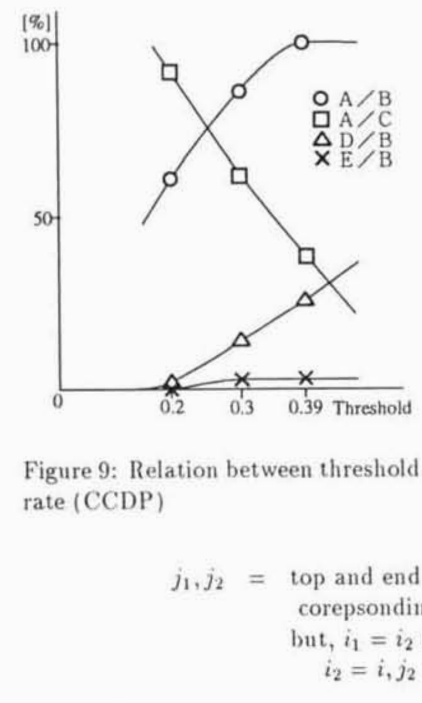

Figure 9: Relation between threshold and recognition rate (CCDP)

j_1, j_2 = top and end of word corepsonding to i but, $i_1 = i_2$ or $j_1 = j_2$, $i_2 = i, j_2 = j$

We call this type of continuous DP matching "compressed continuous DP matching (CCDP)".

4 EXPERIMENTS

We carried out experiments on recognition of words from sign language sentences using CDP and CCDP.

We prepared the data of 17 kinds of words (20 data for each word) as word pattern, and the data of 5 kinds of sentence (5 data for each sentence) as test data. All the data was from one person.

To make the word pattern, part of the word was taken from data by hand. Then, the data were aver-

Method	CDP	CCDP
Calculation	MN	C^2MN
Comparison	$M(N-1)$	$CM(CN-1)$
Memory	$11PN$	$8PCN$
Compression	0.0%	72.7%
Matching time	30.68 sec	1.91 sec
E1	100.0%	97.3%
E2	45.2%	38.3%

M length of input pattern
N length of word pattern
P number of words
C 1-CR (CR:compression rate)

Data

number of words	17 words
length of words	38.7 frame (average)
number of sentences	25 sentences
length of sentences	123.2 frame (average)

E1 the rate that correct words are detected as first candidate.
E2 the rate between correct and and all candidates

Matching time is CPU time on HP9000/model 720

Table 1: Comparison between 2 types of continuous DP matching

aged after DP matching for each word.

The results are shown in Figs. 8 and 9.

5 DISCUSSION

In this section we discuss the performance of CDP and CCDP. Tab. 1 shows the amount of calculation, memory for matching, compression rate, matching time and accuracy of CDP and CCDP.

In CCDP, the number of calculations of distance is proportional to C^2, since the length of pattern is C times the original length. In the experiments, the compression rate was 72.7% on average.

The memory for CCDP is less than for CDP because of its DP path. Furthermore, the memory reduces in proportion to C. Therefore the quantity of word patterns in the system decreases and many word patterns can be stored.

To estimate the accuracy of recognition, two kinds of values were defined as follows.

E1 is the rate that correct words are first candidates when all of the correct words are detected.

E2 is the rate of correct words and all of the detected words when all of the correct words are detected.

The greater these values, the higher is the accuracy of recognition. E1 and E2 of CCDP are lower than those of CDP. This is because CCDP is an approximation of CDP. But the difference is not so large. And considering the matching time, CCDP is enough for practical use. From this result, it is clear that the compression based on dynamic features captures sufficient features of the sign language.

6 CONCLUSION

We presented a sign language translation system that recognizes sign language words from continuous sign-language patterns and translates the words to sentences.

To recognize sign-language words, we used continuous DP matching as used in the field of speech recognition. Further we improved it in order to decrease recognition time.

With these methods, the word detection rate was 97.3% using word patterns of 17 words and 25 sentences as test data. The recognition was approximately 16 times faster than without compression.

In the future, we will increase the words which this system can recognize, and develop a practical system.

REFERENCES

[1] S. Ishihara, "Sign Language for Everyone," NHK series, 1990 (in Japanese)

[2] S.S. Fels, "Building Adaptive Interface with Neural Networks : the Glove-Talk Pilot Study," Technical Report CRG-TR-90-1, University of Toronto, 1990

[3] T. Takahashi, F. Kishino, "A Hand Gesture Recognition Method and Its Application," The Transactions of IEICE, Vol.J73-D-II, No.12, pp.1985-1992, 1990 (in Japanese)

[4] H. Taguchi, S. Nishiyama, M. Kakimoto, K. Murakami, "Gesture Recognition using Neural Networks -Interaction with Virtual Creatures in Artificial Reality-," Proceedings of The 1991 IPSJ Fall Conference, Vol.2, pp.247-248 (in Japanese)

[5] S. Aoshima, K. Hanari, K. Yoshida, T. Yabuta, "Recognition of Finger Spelling by A Neural Type Sign Language Translation Device," Proceedings of The 1991 IEICE Fall Conference , Vol.1, p.129 (in Japanese)

[6] Y. Nakano, "A Study of Sign Language," Fukumura Publisher, 1981 (in Japanese)

[7] "DATAGLOVE MODEL 2 Operation Manual," VPL Research Inc., 1989

[8] R. Oka, "Continuous Words Recognition by Use of Continuous Dynamic Programming for Pattern Matching," Acous. Soc. J., SIG-S, S78-20, pp.145-152, 1978 (in Japanese)