Improvement of Early Recognition of Gesture Patterns based on Self-Organizing Map

Atsushi Shimada, Manabu Kawashima and Rin-ichiro Taniguchi

Department of Advanced Information Technology, Kyushu University
744 Motooka, Nishi-ku Fukuoka 819-0395 Japan
(Tel : +81-92-802-3580; Fax : +81-92-802-3579)
(Email: {atsushi, kawashima, rin}@limu.ait.kyushu-u.ac.jp)

Abstract: We propose an approach to achieve early recognition of gesture patterns. Early recognition is a method to recognize sequential patterns at their beginning parts. Therefore, in the case of gesture recognition, we can get a recognition result of human gestures before the gestures have finished. The most difficult problem of early recognition is that when the system determines the recognition result. Most traditional approaches suffer from this problem since the gestures comprehend ambiguity. Especially at the beginning part of them, it is very difficult to determinate the recognition result since enough input data has not been observed yet. Therefore, we have improved traditional approach by using Self-Organizing Map.

Keywords: Gesture Recognition, Early Recognition, Self-Organizing Map

I. INTRODUCTION

A man-machine seamless interaction is an important tool for various interactive systems such as virtual reality systems, video game consoles, human-robot communication, and so on[5,6]. To realize such an interaction, the system has to estimate human gestures in real-time. Generally, a gesture recognition result is acquired after the gesture has finished. Therefore, if a long gesture is observed, we have to wait for the response until the recognition result is determined. This is a problem to realize a "real-time" man-machine interaction.

Recent years, a new approach called "early recognition" has been proposed for gesture recognition[4,8,1]. The early recognition means that a system outputs a recognition result before a gesture has finished. It is a very useful technique to realize a real-time interaction. The most difficult problem of early recognition is that when the system determines the recognition result. In other words, the system has to ensure the recognition result before the observing gesture has finished. Most traditional approaches suffer from this problem since the gestures comprehend ambiguity. Especially at the beginning part of them, it is very difficult to determinate the recognition result since enough input data has not been observed yet.

In this paper, we propose a new strategy to achieve early recognition of gesture patterns. In the training phase, Self-Organizing Map (SOM)[2,3] is used to learn postures, which are elements of all gestures. The advantages of using SOM are 1) to reduce dimensionality of gesture patterns, 2) to reduce some redundant postures, 3) to represent a gesture pattern by a combination (set) of smaller number of neurons; we call it "Sparse Code". In the test phase, we introduce the Hausdorff distance to measure the similarity to measure the similarity between sparse code of unknown input pattern and sparse code of training patterns. The Hausdorff distance is more effective criterion for judgment than Euclidean distance or so on, since it can measure the distance between not two elements but sets of pattern.

II. DEFINITION OF EARLY RECOGNITION

1. Typical Gesture Recognition

Let $\boldsymbol{C}^i = \{c_1^i, \dots, c_n^i\}$ be a training gesture pattern which belongs to gesture class $i \in \boldsymbol{L}$. The $\boldsymbol{L}$ is a set of class labels. A gesture can be represented in a sequential n-long posture patterns. Therefore, c_n^i means the n-th posture of the gesture. When an unknown gesture $\boldsymbol{X} = \{x_1, \dots, x_l\}$ is observed, the typical gesture recognition problem is to find the most similar gesture from training patterns by

$$p = \underset{i}{\operatorname{argmin}}\{f(\boldsymbol{X}, \boldsymbol{C}^i)\}$$

where p is the class label and $f()$ is a distance function which evaluates the similarity between gesture pattern $\boldsymbol{X}$ and $\boldsymbol{C}^i$.

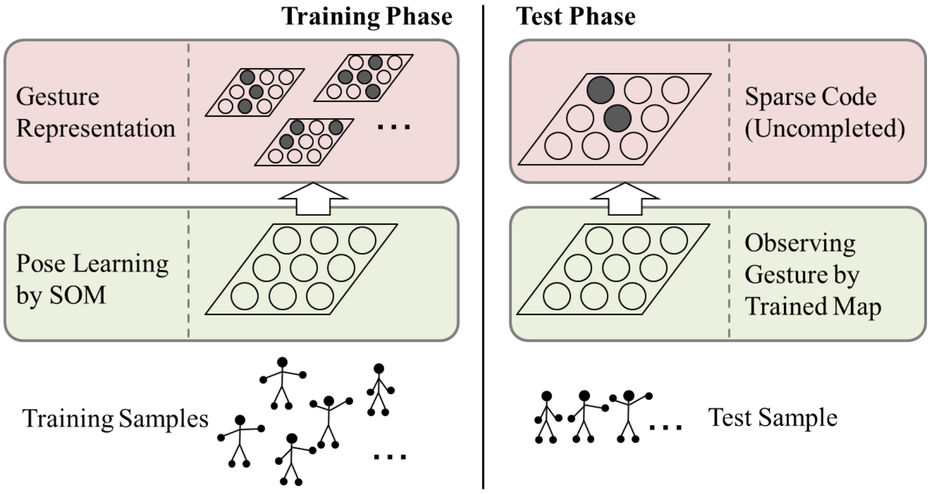

Fig. 1 Processing flow of training/recognition

2. Early Recognition of Gesture Patterns

The key issue of early recognition is to output a recognition result before acquiring complete input pattern. In the case of gesture recognition, it corresponds to the following problem. When a part of gesture pattern (unfinished gesture) $\boldsymbol{X}' = \{x_1, \dots, x_k\}$, $(k < l)$ is observed, the recognition result is determined by

$$p = \underset{i}{\operatorname{argmin}}\{f(\boldsymbol{X}', \boldsymbol{C}^i) < TH\}$$

where TH is a threshold of distance which adjusts the timing of recognition result. If the threshold is not introduced, a recognition result will be output without concrete proof. Therefore, we set a threshold to ensure reliability for the recognition result.

III. STRATEGY OF EARLY RECOGNITION

1. System Overview

First of all, we show the system overview in Fig. 1. The process can be divided into two phases; training phase and test phase. In the training phase, Self-Organizing Map (SOM) is used to learn postures, which are elements of all gestures. The advantages of using SOM are 1) to reduce dimensionality of gesture patterns, 2) to reduce some redundant postures, 3) to represent a gesture pattern by combination of smaller number of neurons and so on. Due to space limitation, we skip the detailed explanation about SOM and how to learn the postures (refer to the literature [7]). After the training of all postures, element postures of each gesture are input to the map again. And then, we can get a "Sparse Code" which represents a gesture pattern on the SOM.

In the test phase, the system observes person's gesture. Then, a parse code corresponding to the observing gesture is generated/updated immediately whenever a new observation is acquired. Finally, the sparse code is examined whether or not the system outputs the recognition result. Actually, the examination is achieved by measuring the distance between sparse codes based on Hausdorff distance.

1. Sparse Coding

When a posture x_k is input to the SOM, one neuron will be selected as winner. When a set of postures which consist of a gesture is sequentially input to the SOM, some neurons will be activated. We regard such an activation pattern as "Sparse Code", which represents an input gesture. Here, we define the notation of a sparse code. In the training phase, all training gestures are represented by using sparse code. For simplicity, we use the same notation $\boldsymbol{C}^i$, which was used for representation of training samples in section II, as the sparse code of them. Meanwhile, in the test phase, a sparse code of observing gesture is represented by $\boldsymbol{X}'$.

Note that the sparse code described here has not an ability to distinguish the gesture patterns whose elements are the same but the sequences are different. However, we can easily improve introduce temporal information into the system by our previous study[7].

2. Similarity Evaluation with Hausdorff distance

The number of elements in sparse code is different from each other since the number of activated neurons

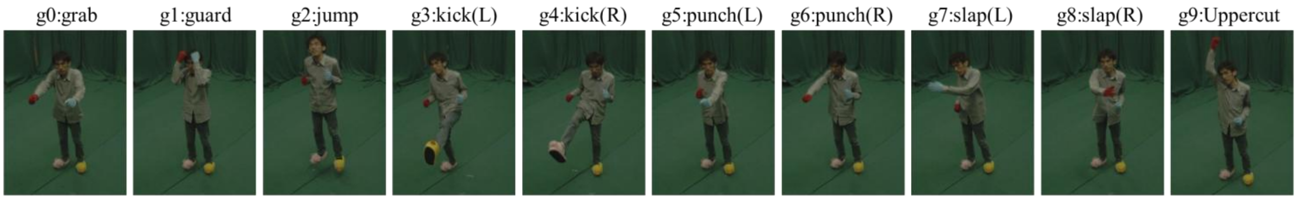

Fig. 2 10 kinds of gesture patterns

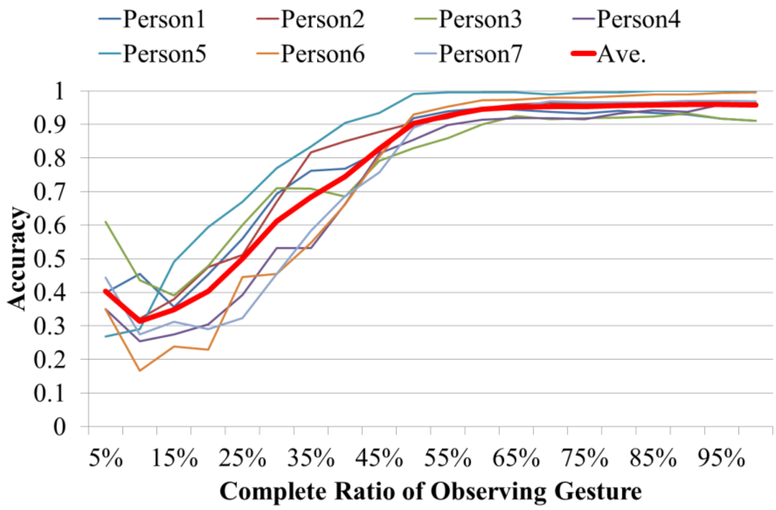

Fig. 3 Early Recognition Accuracy vs. Complete Ratilof Observing Gesture

depends on the gesture length and the gesture pattern. Therefore, we introduce the Hausdorff distance to measure the similarity between two sets of sparse code.

Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two non-empty subsets of a metric space. The Hausdorff distance $f(\boldsymbol{X},\boldsymbol{Y})$, which corresponds to the distance function explained in section II, is defined by

$$f(\boldsymbol{X},\boldsymbol{Y}) = \max_i\{\sup_{x\in X}\inf_{y\in Y} d(x,y),\ \inf_{y\in Y}\sup_{x\in X} d(x,y)\}$$

where $d(x,y)$ is the distance function. In our research, we use L2-distance between the coordinates of activated neurons.

IV. EXPERIMENTAL RESULT

1. Preparation

We demonstrate proposed early recognition of gesture patterns using motion data prepared by ourselves. Each gesture consists of a sequence of postures, and each posture is represented by 5 measured markers. Each marker is composed of data of (x, y, z)-axis. We prepared 10 kinds of gesture patterns ($\boldsymbol{L} = 10$, see Fig. 2) from 7 examinees. Each person did each gesture 40 times. We used 20 patterns for training and other 20 patterns for test. We conducted the experiment through cross-validation among examinees.

2. Recognition Result

Fig. 3 shows the result of early recognition The horizontal axis denotes the complete ratio of observing gesture pattern, and the vertical axis denotes the recognition accuracy. The bold curve indicates the average ratio of accuracy. For example, the recognition ratio exceeded 90% when more than 50% long gestures had been observed on average.

Fig. shows the recognition accuracy of each gesture class. And the detailed analysis results are shown in Table 1. Totally, we got good results for whole gesture classes. However, some cells in Table 1 have lower values. We investigated the matter why such results came out. There are some similar gesture patterns between gesture classes. For example, in the case of gesture class 5 and 7, the person moves his/her right hand forward from in front of his/her body. The difference between these gestures is the trajectory of the hand. However, the trajectory depended on examinee's habit. As the result, some trajectory was very similar

Table 1 Recognition Accuracy of Each Gesture Class and Each Person

	g0	g1	g2	g3	g4	g5	g6	g7	g8	g9
Person1	0.84	0.83	1.00	0.82	0.53	0.75	0.85	0.70	0.81	1.00
Person2	0.85	0.70	1.00	0.83	1.00	0.95	0.75	0.54	1.00	0.70
Person3	0.50	0.91	0.85	0.77	0.99	0.71	0.63	0.89	0.65	0.95
Person4	0.84	0.89	0.72	0.85	0.63	0.89	0.83	0.53	0.68	0.74
Person5	1.00	0.82	1.00	0.90	0.67	0.83	0.75	1.00	0.62	0.93
Person6	0.86	1.00	1.00	0.61	0.75	0.89	0.74	0.73	0.53	0.61
Person7	0.74	0.79	0.79	0.72	0.61	0.64	1.00	0.74	0.55	0.83
Average	0.81	0.85	0.91	0.79	0.74	0.81	0.79	0.73	0.69	0.82

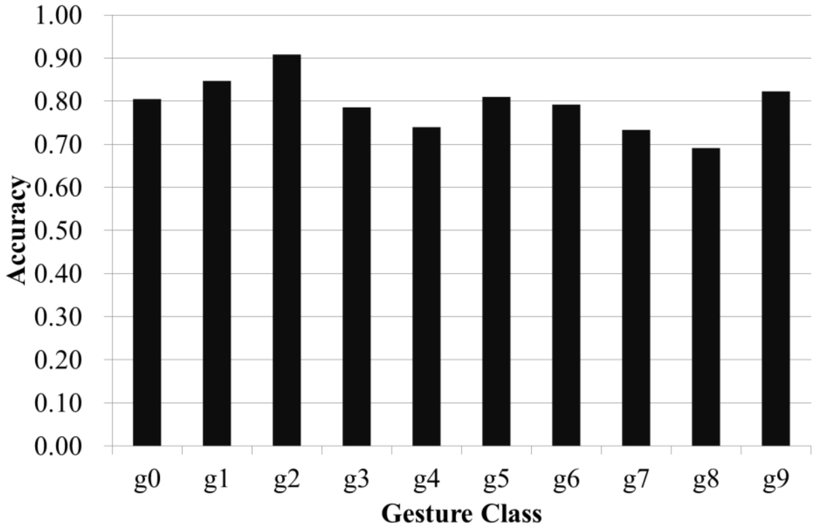

Fig. 4 Recognition Accuracy of Each Gesture Class

with the trajectory of the other gesture class. This is one of the factors which brought not good results.

V. CONCLUSION

We have proposed a new framework of early recognition of human gestures. We have used Self-Organizing Map (SOM) to learn human gestures. The SOM outputs sparse codes for each gesture. We estimated a human gesture based on Bayesian estimation using the sparse code. We got positive results of early recognition in the experiment. We are now researching to tackle the problem of common postures which are included in some gestures.

REFERENCES

[1] M. Kawashima, A. Shimada, and R. Taniguchi. Early recognition of gesture patterns using sparse code of self-organizing map. In 7th International Workshop On Self-Organizing Maps, pages 116–123, 06 2009.

[2] T. Kohonen. "Self-Organization and Associative Memory". Springer-Verlag, 1989.

[3] T. Kohonen. "Self-Organizing Maps". Springer Series in Information Science, 1995.

[4] A. Mori, S. Uchida, R. Kurazume, R. Taniguchi, T. Hasegawa, and H. Sakoe. Early recognition and prediction of gestures. In Proc. of International Conference on Pattern Recognition, volume 3, pages 560–563, 2006.

[5] H. S. Park, D. J. Jung, and H. J. Kim. Vision-based game interface using human gesture. In Advances in Image and Video Technology, volume 4319, pages 662–671, 2006.

[6] J. Park and J. Yi. Gesture recognition based interactive boxing game. In International Journal of Information Technology, volume 12, pages 36–44, 2006.

[7] A. Shimada and R. Taniguchi. Gesture recognition using sparse code of hierarchical som. In Proc. of International Conference on Pattern Recognition, 2008.

[8] S. Uchida and K. Amamoto. Early recognition of sequential patterns by classifier combination. In Proc. of International Conference on Pattern Recognition, 2008.