The Fifteenth International Symposium on Artificial Life and Robotics 2010 (AROB 15th '10),
B-Con Plaza, Beppu,Oita, Japan, February 4-6, 2010

Consideration on Gesture Recognition Based on Multilayer Neural Network by Using Input Device of Home Gaming Console

Ichiro Iimura⋆ Takuroh Fujiki⋆ Hidenobu Tsurusawa† Shigeru Nakayama‡

⋆*Department of Administration, Prefectural University of Kumamoto*
3-1-100 Tsukide, Kumamoto, 862-8502 Japan
{iiimura, g0530214}@pu-kumamoto.ac.jp

†*Department of Computer and Control Engineering, Oita National College of Technology*
1666 Maki, Oita, 870-0152 Japan
htsuru30@oita-ct.ac.jp

‡*Department of Information and Computer Science, Kagoshima University*
1-21-40 Korimoto, Kagoshima, 890-0065 Japan
shignaka@ibe.kagoshima-u.ac.jp

Abstract: We consider that the human intention and feelings conspicuously appear at the force added to a body rather than the positions of hands in the gesture that is nonverbal communication. In this paper, we have tried gesture recognition based on multilayer neural network by using acceleration sensor because the force to act during motion can be detected by acceleration. From the experimental result, we have proven that our gesture recognition method can recognize the gesture patterns such as numerical characters and graphic symbols. As an input device in this study, we used the Wii remote attached to the home gaming console Wii instead of a special device. Furthermore, we have obtained high recognition rate in the case of not only specific performers but also unspecified performers by using the permutation of the simple normalized acceleration for input values of multilayer neural network. Therefore, we could develop versatile gesture recognition engine.

Keywords: Gesture recognition, Neural network, Input device, Home gaming console, Wii remote

I. INTRODUCTION

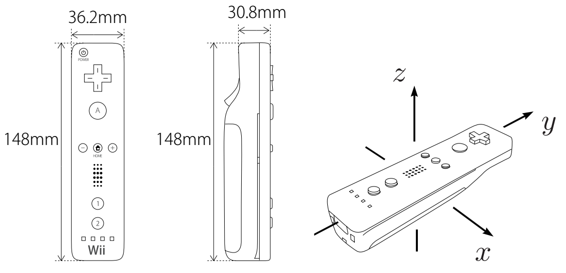
Recently, the operation mode of computers has diversified in the mobile environments, information appliances and so on, from miniaturization and high performance. Computers are fusing to our real world. Therefore, a study of new user interface technology becomes popular to realize that a human operates a machine naturally. In the communication of intention and feelings between humans, it is generally said that nonverbal means by the physical media such as gesture or hand-gesture, *nonverbal communication*, is one of the important communication as well as verbal means by words [1, 2]. If a machine/computer can understand this gesture, it can become the natural and intuitive new input means.

In conventional studies, Sawada et al [3]. proposed gesture recognition methods using acceleration data or features patterns of acceleration and developed their practical systems. In particular, three-axes acceleration sensor have been built into some of the recent mobile phones, and the input by intuitive operation that uses it is also possible. As the studies of Sawada et al. about the gesture recognition using acceleration after Reference [3], they progressed their studies to the application to the sign-language recognition [4] and the data retrieval by gesture [5].

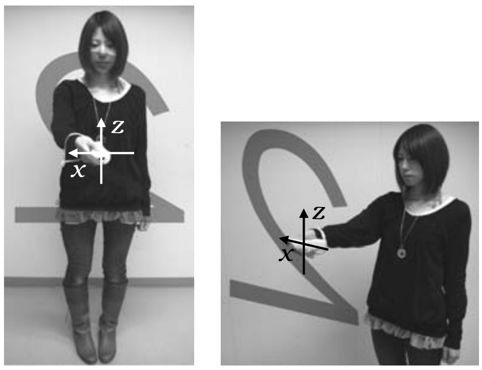
In this study, we have accepted the consideration of Sawada et al [3]. in which the human intention and feelings conspicuously appear at the force added to a body rather than the positions of hands in the gesture that is nonverbal communication. We have tried gesture recognition based on multilayer neural network by using acceleration sensor because the force to act during motion can be detected by acceleration. In the acceleration measurement, we used the Wii remote attached to the home gaming console Wii of Nintendo Co., Ltd. as an input device instead of a special device. Wii remote is equipped with Bluetooth, a three-axes acceleration sensor, and a CMOS sensor detecting infrared rays. In this study, we decided the adoption of the Wii remote as an input device, because of both the convenience of the no-cabling by using Bluetooth and the easiness of getting acceleration by using the three-axes acceleration sensor. The development language we used was Java language and used WiiRemoteJ 1.5 for controlling the Wii remote that is an input device. Besides, as applications to the entertainment, we are plan to study the recognition of flag signaling, the estimate of emotion [6, 7], and the development of application systems such as the instruction of a dance or an exercise. These studies is based on biological motion data [8] of the human movement which we get with three-axes acceleration sensor. Therefore, this study deserves a preliminary experiment.

II. KINETIC FEATURE QUANTITY FOR GESTURE RECOGNITION

We show the appearance of the input device, which we used for an acceleration measurement in this study, in Figs. 1(a), 1(b) and 1(c). Among x, y and z axial acceleration $a_x(t)$, $a_y(t)$ and $a_z(t)$ measured with this input device, we adopted the acceleration of x and z axis directions, which we showed in



(a) Top (b) Side (c) Perspective

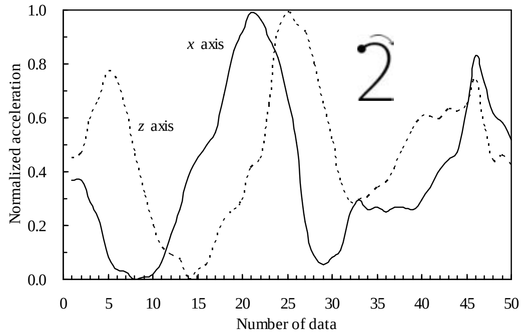


(d) Front (e) Diagonal

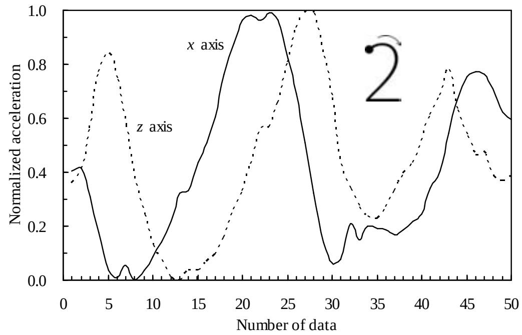
Figure 1: Input device with three-axes acceleration sensor and directions of two axes in which acceleration is measured.

Figs. 1(d) and 1(e), as kinetic feature quantity and used them for gesture recognition. In addition, the start and end of the gesture were judged from pressing Ⓐ-button of the input device shown in Fig. 1. Furthermore, as for the time and the force intensity of one gesture, it is thought that the individual variation occurs by presenters. Therefore, in this study, we normalized the number of acceleration data per one axis, e.g. $a_x(t)$ or $a_z(t)$, and the value of acceleration per one axis to N_a and the range of $[0, 1]$, respectively. We used the permutation which simply connected the normalized acceleration of the directions of x and z axes in the recognition experiment described later.

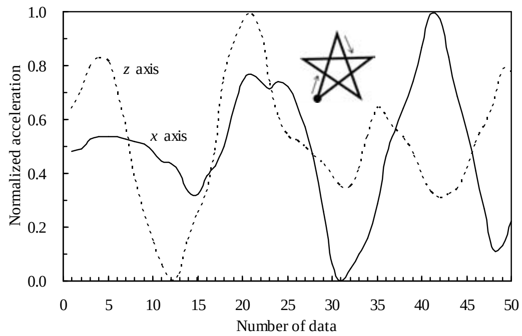
As an example, we show the transitions of the normalized acceleration in the directions of x and z axes in Fig. 2, concerning numerical character "2" and graphic symbol "☆" of two examinees A and B. In these figures, a solid line and a dotted line are the transitions of the x axial and z axial normalized acceleration, respectively. The number of each axial normalized data, N_a, is 50. Figs. 2(a) and 2(b) show the acceleration transitions of the gesture "2" in the different examinees. We can confirm that these transitions tend to be almost similar when we compare these transitions. We can confirm a similar tendency in Figs. 2(c) and 2(d), too. However, the acceleration transitions are clearly different in different gestures of a same examinee. From these results, we have considered that the recognizing gestures by the simple acceleration transitions is possible.



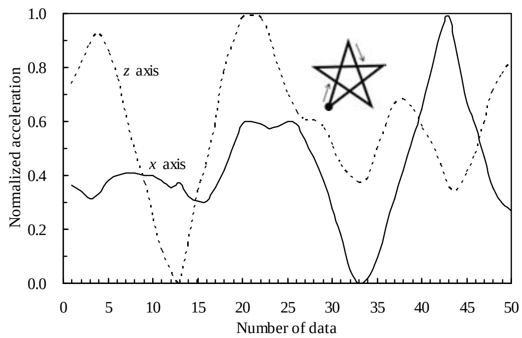
(a) "2" (Examinee A)



(b) "2" (Examinee B)



(c) "☆" (Examinee A)



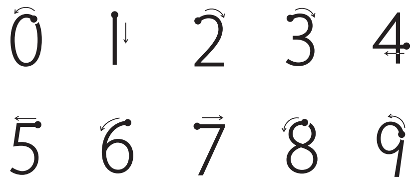
(d) "☆" (Examinee B)

Figure 2: Transitions of the normalized acceleration in the directions of x and z axes.

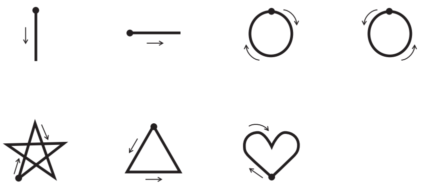
III. GESTURE RECOGNITION

1. Recognition experiment

Neural Network (NN) is one of the information processing mechanisms which modelled the structure



(a) Ten numerical-character-gesture patterns



(b) Seven graphic-symbol-gesture patterns

Figure 3: Gesture patterns used in the experiment.

of the nerve cell in the human brain. In this study, we adopted hierarchical NN of the three layers, whose usefulness has been confirmed by applications to the problems such as the pattern recognition, for recognizing gesture patterns. In addition, we used Back-Propagation (BP) method [9], which is the supervised learning method, for learning.

We used the ten numerical-character-gesture patterns and the seven graphic-symbol-gesture patterns shown in Fig. 3 for the recognition experiment of this study. We constructed two NNs for the numerical-character-gesture patterns and the graphic-symbol-gesture patterns, and separately experimented in the two kinds of gesture patterns. As described in Section II, the data of one pattern to use in the recognition experiment was the permutation which simply connected the normalized acceleration of the directions of x and z axes. N_a of each axis was 50. The data used in the experiment was collected from 20 examinees, and we got ten samples per one gesture pattern. Here, we randomly divided 20 examinees into two groups, Group A and Group B, in which each group consists of ten examinees. The collected data was classified as follows according to a use.

DL *Data for learning.* The data was gotten from ten examinees who belonged to Group A.

DR$^+$ *Data for recognizing of examinees whose data was used for learning.* The data was gotten from ten examinees who belonged to Group A again besides data for learning, DL.

DR$^-$ *Data for recognizing of examinees whose data was not used for learning.* The data was gotten from ten examinees who belonged to Group B.

We used DL for learning of NN, and used DR^+ and DR^- for recognition experiment.

The NN we constructed consists of the input layer of 100 units and the hidden layer of 150 units. As for the output layer, we had one unit correspond to one gesture pattern. That is, we decided the number of units in the output layer to ten units and seven units for recognizing the numerical-character-gesture patterns and the graphic-symbol-gesture patterns, respectively. The parameter values for learning by BP method are desided by the pretest. As a result, the learning rate η was 0.25 and the stabilization constant α, which determines the effect of past weight changes on the current direction of movement in weight space, was 0.5 [1] . The at-end condition for learning by NN was when the squared-sum of the error between output value of the output layer and teacher signal became less than 1.8×10^{-3}.

2. Experimental result and consideration

We constructed the NN for recognizing the numerical-character-gesture patterns by using DL of the numerical-character-gesture patterns. In the same way, we constructed the another NN for recognizing the graphic-symbol-gesture patterns by using DL of the graphic-symbol-gesture patterns, too. We experimented on recognizing gesture patterns by using each NN we constructed. The recognition results concerning the numerical-character-gesture patterns and the graphic-symbol-gesture patterns are shown in Figs. 4(a) and 4(b), respectively. In both figures, the axis of abscissas is each gesture pattern, and the axis of ordinates is recognition rate [%]. Because we gotten ten samples per one pattern from one examinee, the recognition rate concerning DR^+ and DR^- is 100% if the NN we constructed succeeds in recognition for 100 samples (= 10[samples/examinee] $\times$ 10[examinees]) per one pattern.

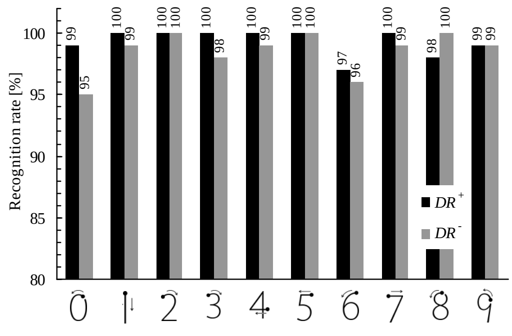
From these experimental results, we have confirmed that the recognition rate of DR^+ was almost 100%. We have confirmed that the recognition rate of DR^- was satisfactory performance, too. That is, we have obtained high recognition rate in the case of not only specific performers but also unspecified performers in spite of using the permutation of the simple normalized acceleration for input values of NN.

IV. COMPARISON WITH CONVENTIONAL STUDIES

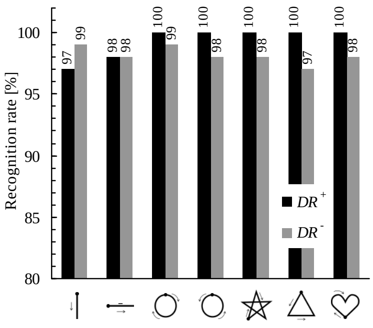
In order to confirm the effectiveness of the gesture recognition engine constructed in this study, we compared the proposed method with the conventional study of Sawada et al [3]. related to the gesture recognition. Table 1 shows the comparison result of the recognition rate concerning examinees whose data was not learned. Sawada et al. dealt with ten gesture patterns in Reference [3], but we dealt with seven gesture patterns except three patterns as shown in Fig. 3(b) because the specific motion of these three patterns was not clear in Reference [3].

In the conventional method [3], Sawada et al. recognized the gestures by pattern matching. In other words, we cannot simply compare the recognition performance because the conventional method is different from the proposed method in the recognition

[1] As for the details about the learning rate η and the stabilization constant α, please refer to Reference [9].



(a) Numerical-character gestures



(b) Graphic-symbol gestures

Figure 4: Recognition rate [%] in terms of ten numerical-character-gesture and seven graphic-symbol-gesture patterns.

Table 1: Comparison of recognition rate [%] in terms of seven graphic-symbol-gesture patterns.

Gesture patterns	Conventional method	Proposed method (DR^-)
↓	100.0	99.0
⇀	100.0	98.0
↻	100.0	99.0
↺	80.0	98.0
☆	100.0	98.0
△	80.0	97.0
♡	100.0	98.0
Average	94.3	98.1

method. However, in ↺ and △ which were misrecognized by the conventional method because of gesture patterns with similar kinetic feature quantity, the proposed method have shown high recognition rate as well as other gesture patterns. Furthermore, the proposed method has not shown the decrease of recognition rate in specific gestures. As for the average recognition rate in seven patterns, the proposed method is about 98.1%, whereas the conventional method is about 94.3%. In addition, the recognition rate of each pattern in the proposed method is the result of 100 samples by 10 examinees, whereas the recognition rate of each pattern shown in Reference [3] is the result of only ten samples by one examinee. From these analysis, we have confirmed the higher recognition stability of the proposed method compared with the conventional method.

V. CONCLUSIONS

In this study, we have tried gesture recognition based on multilayer neural network by using acceleration sensor because the force to act during motion can be detected by acceleration. From the experimental result, we have proven that our gesture recognition method can recognize the gesture patterns such as numerical symbols and graphic symbols. As an input device in this study, we used Wii remote attached to the home gaming console Wii instead of a special device. Furthermore, we have obtained high recognition rate in the case of not only specific performers but also unspecified performers by using the permutation of the simple normalized acceleration for input values of multilayer neural network. Therefore, we could develop versatile gesture recognition engine. In the near future, we are plan to try the recognition of more complicated gestures by using acceleration of three axes, though the acceleration of only two axes was used in this study, and to study the human-emotion estimate.

References

[1] Kurokawa, T. (1994), Nonverbal interface (in Japanese). Ohmsha

[2] Kanayama, N. (1983), Sekai 20-kakoku non babaru jiten (in Japanese). Kenkyusha

[3] Sawada, H. and Hashimoto, S. (1996), Gesture recognition using acceleration sensor and its application for musical performance control (in Japanese). IEICE Trans. Fundamentals J79-A(2):452–459

[4] Sawada, H., Hashimoto, S. and Matsushima, T. (1998), A study of gesture recognition based on motion and hand figure primitives and its application to sign language recognition (in Japanese). IPSJ Journal 39(5):1325–1333

[5] Notsu, T., Hashimoto, S. and Sawada, H. (2002), Multi-modal sign language database and its retrieval by gesture inputs (in Japanese). Trans. HIS 4(1):51–58

[6] Ishii, M. and Maruta, H. (2005), Percept of emotion from biological motion (in Japanese). IEICE Technical Report 105(165):123–126

[7] Maruta, H. and Ishii, M. (2006), Feature extraction with ICA for human-emotion discrimination from biological motion data (in Japanese). IEICE Technical Report 105(534):135–139

[8] Johansson, G. (1973), Visual perception of biological motion and a model for its analysis. Perception & Psychophysics 14:201–211

[9] Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986), Learning internal representations by error propagation. Parallel Distributed Processing, vol. 1, MIT Press, pp. 318–362