Gesture Recognition Based on the Detection of Periodic Motion

Masahito Takahashi and Kota Irie
Course of Precision Engineering
School of Science and Engineering
Chuo University
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

Kenji Terabayashi and Kazunori Umeda
Department of Precision Mechanics
Faculty of Science and Engineering
Chuo University
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan
Email: {terabayashi,umeda}@mech.chuo-u.ac.jp

Abstract—**In this paper, we propose a method to recognize periodic gestures from images. The proposed method uses a amplitude spectrum and a phase spectrum that are obtained by applying Fast Fourier Transform (FFT) to a time series of intensity images. FFT is applied to each pixel of low-resolution images. The method consists of 2 steps. First, the method detects periodic motion regions from the amplitude spectrum. Secondly, the method uses the phase spectrum in the detected periodic motion region to classify the gestures. The proposed method is robust to lighting conditions and individual differences in skin color because it does not rely on color information. Several experiments are performed to demonstrate the effectiveness of the proposed method.**

I. Introduction

In order to realize a more improved quality of life, a human machine interface is required so as to perform natural interaction between a human and a robot. Gestures, which we use frequently and intuitively in our everyday communication, can be one such human-machine interface.

Until now, many studies that recognize gestures have been reported [1], [2]. In the field of virtual reality, the three-dimensional position and motion tracking of a hand are detected and gesture is recognized using a contact-type device that carries a magnet sensor and a marker [3], [4]. However, the use of contact-type device is unsuitable in the pursuit of natural interaction because the user must carry the contact-type device, which limits his range of motion. For this reason, many studies based on image processing, which does not require a contact-type device, have been conducted [5], [6]. Gesture recognition occurs from motion and change, such as that which occurs when a hand is withdrawn from an input image [7]. Some techniques are used to extract a motion from an input image. One involves extracting image regions corresponding to human skin [8]. However, the extraction of stable hand regions is difficult because this technique is sensitive to the lighting conditions and individual differences in skin color. Other technique relies on background image subtraction [9]. The environments in which this technique can be utilized are limited to the places which background is stationary. However, the background is often cluttered in the real world. There is also a technique of action detection in complex scenes with cluttered backgrounds, heavy crowds, occluded bodies [10]. The technique of recognizing a gesture without preprocessing has also been studied, which extracts image regions corresponding to a user's hand, using a contact-type device [11]. This study recognizes gestures utilizing the features extracted from the video data of a time series.

In the studies referenced above, various hand gestures and complex motions were recognized. On the other hand, gestures with simple motion are enough for simple operations of a machine or robot. In this operation, the robustness to illumination change and individual differences is more important than complexity of various gestures. We focused on a hand waving that is an intuitive gesture and can be a strong cue, and have proposed a simple method that does not rely on color information for the identification of waving hands in images [12]. However, this method is limited to the recognition of periodic repeated motions. In addition, only one gesture can be used to express intention.

In this paper, we propose a new method for the recognition of periodic gestures that enables the expression of two or more simple intentions. The proposed method can be used to identify a periodic gesture utilizing the phase spectrum obtained by applying FFT [13]. The method is robust to lighting conditions and individual differences in skin color because it does not rely on color information. Additionally, the method is robust for hand speed by utilizing the phase spectrum differences acquired from a moving hand region.

This paper is organized as follows. In Section II, we describe our proposed method for recognizing periodic gestures. In Section III, we report experimental results for evaluating the performance of the proposed method. Finally, in Section IV, we present our conclusions.

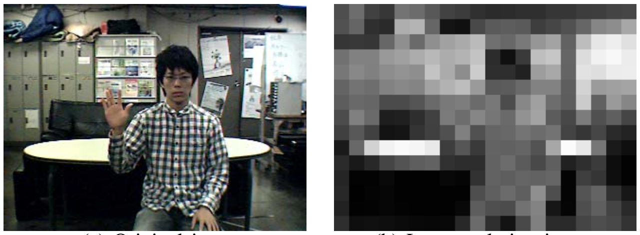
II. Periodic Gesture Recognition

In this section, we describe our proposed method. First, a periodic motion region is detected from a time series of intensity images (Sec.II-A). Next, the moving hand region is extracted from a periodic motion region (Sec.II-B). Finally, the kind of gesture is identified (Sec.II-C).

A. Detection of Periodic Motion Regions

When a hand is waved periodically, the intensity value of a pixel corresponding to the hand region vibrates between the hand region and the background. We focus on this vibration.

978-1-4244-7682-4/10/$26.00 ©2010 IEEE



(a) Original image (b) Low-resolution image

Fig. 1. Conversion of image to low resolution

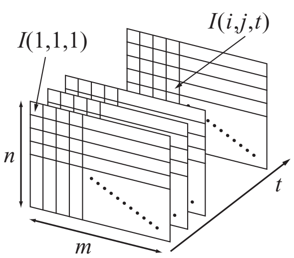


Fig. 2. Time series of low-resolution images

As this change of intensity value is periodic with a constant cycle, we can take advantage of FFT for quantifying. By this process, the periodic motion regions are detected.

1) Application of FFT to Time Series of Intensity Values: Features of periodic gestures are extracted from the amplitude and phase spectrums obtained from the time series of intensity values.

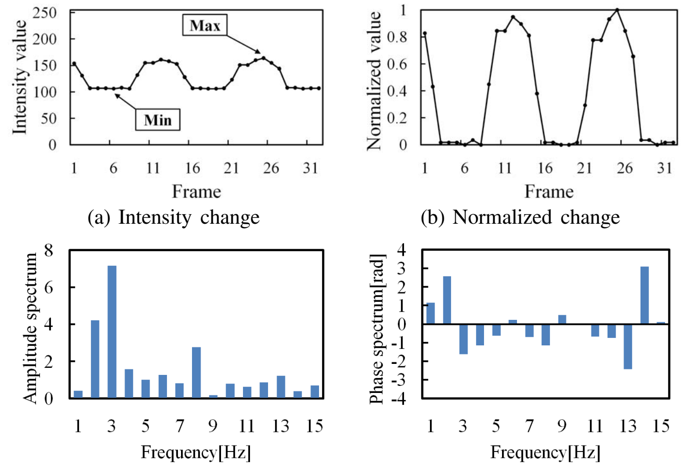
Each image is converted to low resolution, and time series of low-resolution images are obtained. Fig.1 shows an example of the reduction of image resolution. The original image Fig.1(a) is converted to the low-resolution image Fig.1(b). Let us assume that the number of pixels of the images is $m \times n$, and $I(i,j,t)$ is the intensity value of (i,j) pixel ($i=1,2,\cdots,m$, j=$1,2,\cdots,n$) of t-th frame, as shown in Fig.2. By reducing the resolution of the image, the pattern of the vibration is smoothed. In addition, robustness against noise is acquired, and the calculation cost is reduced.

The intensity value $I(i,j,t)$ of a pixel corresponds to the region of a waving hand changes as illustrated in Fig.3(a), since the rate of the hand and the background change periodically according to the waving of the hand. We apply the normalization process to intensity values of every pixel that have variation for the last u frames, as illustrated in Fig.3(b). Moreover, we apply FFT to the normalized intensity values of the pixel and obtain the amplitude and phase spectrums, as illustrated in Fig.3(c) and Fig.3(d). To remove the effect of noises and reduce the calculation cost, FFT is applied to the pixels that satisfies

$$\max(I_{t-u+1},\cdots,I_t)-\min(I_{t-u+1},\cdots,I_t)\geq I_{dif} \quad (1)$$

Amplitude spectrum A_n and phase spectrum $\arg F_n$ are given by

$$F_n=\sum_{k=0}^{N-1} I_k \mathbf{W}_N^{nk}, \quad (2)$$



(a) Intensity change (b) Normalized change

(c) Amplitude spectrum (d) Phase spectrum

Fig. 3. Application of FFT to time series of intensity values

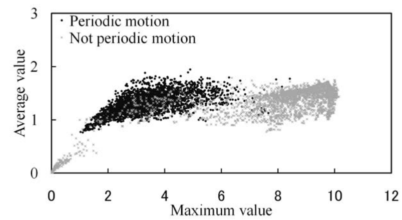


Fig. 4. Distribution of features of the amplitude spectrum

$$A_n=|F_n|, \quad (3)$$

$$\arg F_n=\tan^{-1}\frac{\mathrm{Im}(F_n)}{\mathrm{Re}(F_n)}. \quad (4)$$

N is the number of sampling, and $\mathbf{W}$ is the twiddle factor of the Discrete Fourier Transform (DFT). $\mathrm{Re}(F_n)$ and $\mathrm{Im}(F_n)$ show the real part and imaginary part of F_n.

2) Detection of a Periodic Motion Region on the Basis of SVM: A periodic motion region is detected for every pixel by applying SVM [14] to the amplitude spectrum.

SVMs have been successfully applied to pedestrian detection and motion/face recognition [15], [16]. An SVM is a set of related supervised learning methods used for 2 class classifications fundamentally. The problem of periodic motion detection results in a combination of binary classifications. Then, it is possible to apply an SVM.

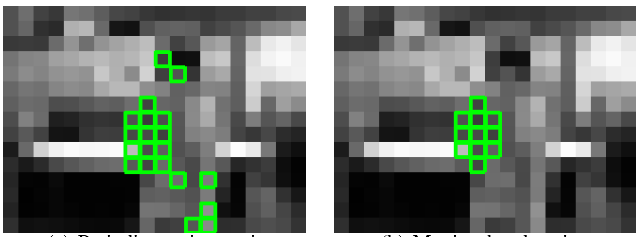
We utilize the features of the summary statistics of the amplitude spectrum A_n. For example, Fig.4 shows the distribution of the maximum value A_{max} and the mean values Ave of an amplitude spectrum, which is obtained by performing periodic motion and other types of motion. A_{max} and Ave are give by (5) and (6) respectively.

$$A_{\max}=\max(A_n). \quad (5)$$

$$Ave=\frac{2}{N}\sum_{n=1}^{N/2} A_n. \quad (6)$$

(a) Original image (b) Low-resolution image

Fig. 5. Detection of periodic motion



(a) Periodic motion region (b) Moving hand region

Fig. 6. Extraction of moving hand region

In the case of the problem of periodic motion detection, periodic motion data and other motion data are not separated by a linear separating hyperplane. An SVM creates a soft margin that permits some misclassifications. A hyperplane, which separates the feature space, is determined by the solution of the following optimization problem.

$$\text{minimize} \quad \boldsymbol{L}(\boldsymbol{w},\zeta) = \frac{1}{2}||\boldsymbol{w}||^2 + \gamma \sum_{a=1}^{M} (\zeta_a) \tag{7}$$

$$\text{subject to} \quad \begin{array}{l} \zeta_a \geq 0,\ s_a(\boldsymbol{w}^T \phi(\boldsymbol{x}_a) + b) \geq 1 - \zeta_a. \\ (a = 1, 2, \cdots, M) \end{array} \tag{8}$$

The training vectors $\boldsymbol{x}_a$ are mapped into a higher dimensional space by the function ϕ. $\boldsymbol{x}_a$ are summary statistics of the amplitude spectrum A_n, i.e., A_{max} and Ave. The class label is s_a. $\boldsymbol{w}$ and b are parameters called weight vector and bias respectively. γ is the penalty parameter of the error term. ζ_a is a slack variable. Furthermore, $K(x_a, x_b) \equiv \phi(x_a)^T \phi(x_b)$ is called the kernel function. The SVM used in our method applies the kernel function. Fig.5 shows an example of the detection of periodic motion by utilizing an SVM, when a hand is waved vertically. In Fig.5(b), framed pixels indicate regions in which periodic motion was detected. The region of the gesture of operation and the region that is vibrating periodically in connection with it are detected.

B. Extraction of Moving Hand Region

Periodic motion regions also contain non-hand regions, e.g., the vibrating region of the head, arm, shoulder or clothes, as illustrated in Fig.5(b). For this reason, the discrimination of a gesture is difficult. Therefore, the extracted periodic motion region is corrected. A morphological operation is applied to all the pixels in which periodic motion is detected. Furthermore,

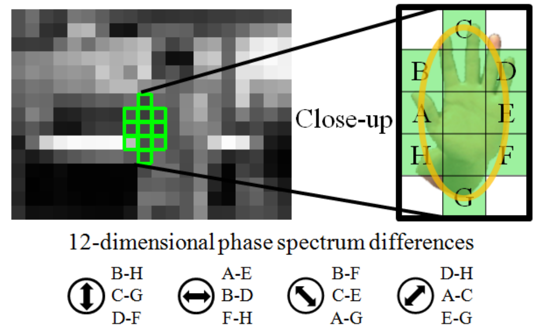
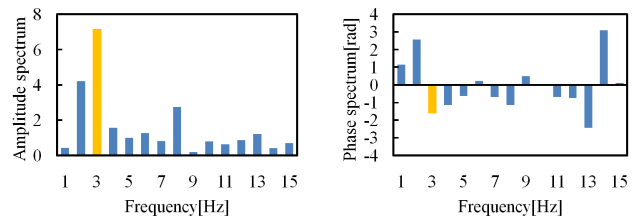


Fig. 7. Regions in which the component of the phase spectrum can be obtained



(a) Component of amplitude spectrum (b) Selected component of phase spectrum

Fig. 8. Selection of a component of phase spectrum

only the maximum connected region is extracted by applying labeling processing, and it is regarded as a moving hand region. The result of having applied a Morphological operation and labeling processing to Fig.6(a) is shown in Fig.6(b).

C. Recognition of Periodic Gesture

A rectangular region where all the pixels of a moving hand region are included is made. The phase spectrum of the pixel of eight points included in the rectangular region is acquired. The eight points are pixels that divide an ellipse circumference inscribed in the rectangular region into eight equal parts. As features, the 12-dimensional phase spectrum differences between the pixel of eight points are acquired. Fig.7 shows the regions where the component of the phase spectrum can be obtained. The component of the phase spectrum of the peak frequency is obtained when the amplitude spectrum is at its maximum, as illustrated in Fig8.

Support vector machines for multi-class classification are applied for the discrimination of periodic gestures. Multi-class problems with more than two classes have typically been solved combining independently produced binary classifiers. We use the one-against-one approach in which $k(k-1)/2$ classifiers are constructed and each one trains data from two different classes [17]. The SVM is calculated for the rectangular region and the periodic gesture with the maximum number of votes is discriminated as the specific gesture. To make the recognition more robust, the recognition of periodic gestures is performed when this condition continues for l frames. Fig.9 shows the flow of the recognition of periodic gestures described above.

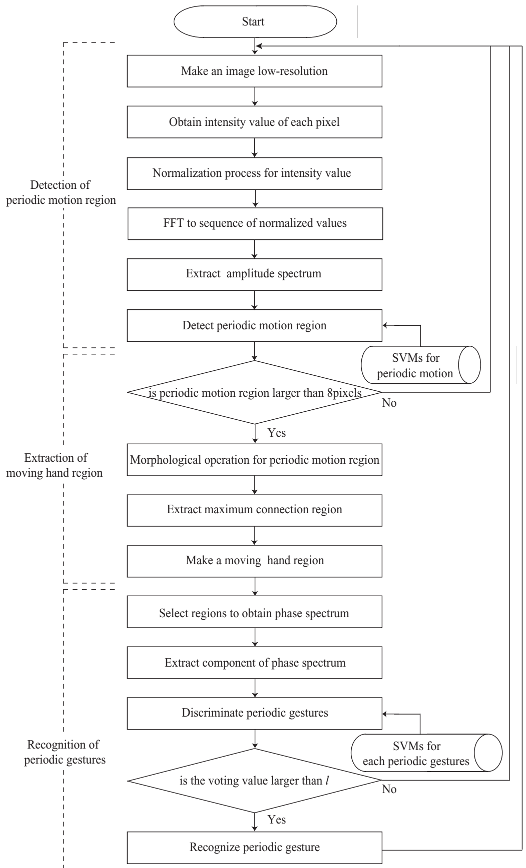


Fig. 9. Flow of recognition of a periodic gesture

III. Experiments

In this section, we show some experimental results to demonstrate the effectiveness of the proposed method to recognize periodic gestures. The following three points are evaluated in the recognition of the periodic gestures.

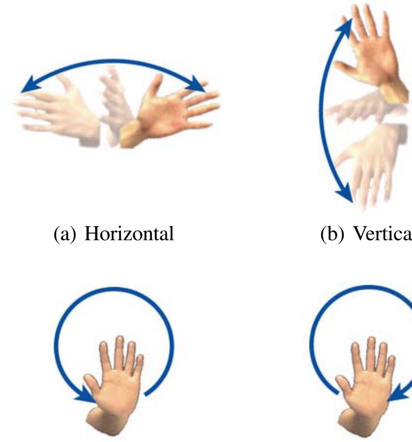
· Robustness for lighting conditions
· Robustness for a different subjects
· Robustness for hand speed

A. Four Periodic Gestures

The experiments were performed for four kinds of gestures: horizontal, vertical, clockwise, and counterclockwise repeated motions of hand. Fig.10 shows the four periodic gestures. They are typical examples of periodic gestures, that can be interpreted to have a certain meaning, such as decision, stopping, or changing some quantities. They can be applied to give a command in the Intelligent Room [18], for example.

B. Experimental Conditions and Parameters

Every calculation, including FFT for every pixel and recognition, is performed by a PC with Core2Quad(2.4[GHz]). For inputting images and converting the images to low resolution, we used Microsoft Directshow software and the image processing software Intel OpenCV. LIBSVM [19] was utilized as a SVM software. For a web camera, we used ELECOM UCAM-E130.



(a) Horizontal (b) Vertical

(c) Clockwise (d) Counterclockwise

Fig. 10. Four periodic gestures

Fluorescent lights were used for lighting. The hand gestures were performed in front of the camera. We set the measurement distance to 2[m]. The number of pixels of the low-resolution image was 20×15[pixel]. The number of samplings N for FFT was set to 32. The number of normalizations u was set to 32. The number of estimating periodic gestures l was set to 20. These values were defined empirically.

When applying SVM to the proposed method, the radial basis function (RBF) shown in the following formula was utilized.

$$\boldsymbol{K}(\boldsymbol{x}_a,\boldsymbol{x}_b) = \exp\left(\frac{-||\boldsymbol{x}_a-\boldsymbol{x}_b||^2}{\sigma^2}\right). \tag{9}$$

σ is kernel parameters. γ in (7) and σ in (9) were decided by Cross validation and Grid-search.

C. Detection of Periodic Motion Region

Experiments were performed for four subjects. The four subjects were divided into two groups, A and B. First, for the subjects in Group A, an amplitude spectrum of four kinds of periodic gestures and other motions was acquired for 4,000 frames in a bright room. The other motions were various random motions, e.g., walking randomly in the room. These were utilized as training data. Next, for the both group subjects, an amplitude spectrum of the same motions was acquired for 1,000 frames in a bright room. These were utilized as test data.

Detection rates of periodic motion for Group A and B are 90.1% and 91.6% respectively. It is shown that a high recognition rate is achieved for detection of periodic motion. The difference of recognition rates in group A and B seems to be caused by the fact that the difference in intensity between the hand region and the background was larger in Group B.

(a) Bright room 250-315(lx) (b) Dark room 60-190(lx)

Fig. 11. Difference of illumination

TABLE I
RECOGNITION RATE OF PERIODIC GESTURES FOR DIFFERENT LIGHTING CONDITIONS (%)

(a) Bright room 250-315[lx]

Actual class	Predicted class			
	Horizontal	Vertical	Clockwise	Counter clockwise
Horizontal	75	3	2	20
Vertical	0	74	6	20
Clockwise	3	2	71	24
Counterclockwise	5	8	16	71

(b) Dark room 60-190[lx]

Actual class	Predicted class			
	Horizontal	Vertical	Clockwise	Counter clockwise
Horizontal	73	1	3	23
Vertical	0	84	2	14
Clockwise	3	4	66	27
Counterclockwise	1	10	13	76

D. Recognition of Periodic Gesture for Different Lighting Conditions

Experiments were performed for twenty subjects. The twenty subjects were divided into two groups; sixteen people in Group C and four people in Group D. First, for the subjects in Group C, the phase spectra of four kinds of periodic gestures were acquired for 16,000 frames in a bright room in each case. These spectra were utilized as training data. Next, the Group D subjects repeated each gesture 25 times in bright and dark rooms. Illumination around the hand was 250-315[lx] for condition 1 (bright room) and 60-190[lx] for condition 2 (dark room), as shown in Fig.11. When the intended gesture and the gesture which the system recognized could be matched, the recognition of the periodic gesture was considered to be successful. Fig.12 shows the image sequence of each periodic gesture and recognition results.

Table I shows the experimental results. The average of the diagonal component of Table I is the probability of recognizing each periodic gesture correctly. It is 73% in the bright room, and 75% in the dark room. The experiment shows the validity of utilizing the phase spectrum in order to identify each periodic gesture. It is also shown that a similar recognition rate is obtained for different lighting conditions. However, for both lighting conditions, the false recognition rate of counterclockwise motion is a little higher compared to that of other gestures. Counterclockwise motion is difficult to operate with a suitable trajectory, compared with other gestures, which seems to make phase spectrum of counterclockwise motion vary widely. In general, it may be difficult for a person to move their hand in completely periodic manner. As a result of having trained the spectrum of multiple subjects, the gesture is recognized even in the case of an incomplete period if it is the amplitude of natural gesture motion.

The average of the necessary time for recognition after the start of the gesture was 1.5[s].

TABLE II
RECOGNITION RATE OF PERIODIC GESTURES FOR DIFFERENT HAND SPEEDS (%)

(a) 2[Hz]

Actual class	Predicted class			
	Horizontal	Vertical	Clockwise	Counter clockwise
Horizontal	88	0	4	8
Vertical	0	80	5	15
Clockwise	16	7	59	18
Counterclockwise	8	18	10	64

(b) 3[Hz]

Actual class	Predicted class			
	Horizontal	Vertical	Clockwise	Counter clockwise
Horizontal	76	0	6	18
Vertical	0	83	0	17
Clockwise	7	2	69	22
Counterclockwise	6	11	8	75

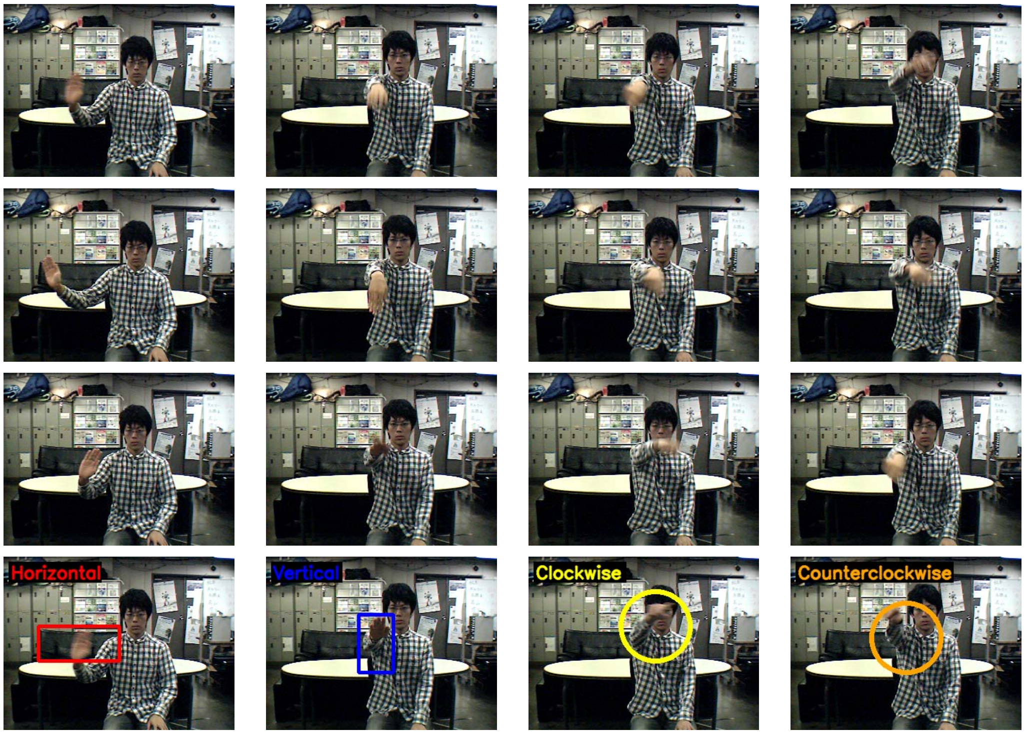
(c) 4[Hz]

Actual class	Predicted class			
	Horizontal	Vertical	Clockwise	Counter clockwise
Horizontal	62	2	6	30
Vertical	0	84	5	11

E. Recognition of Periodic Gesture for Different Hand Speeds

The phase spectra data of Group C acquired for Sec. III-D was also used for training and the experiments in this section. The Group D subjects repeated each gesture 25 times for hand speed, i.e., 2[Hz], 3[Hz], and 4[Hz]. The clockwise and counterclockwise motions were performed only at 2[Hz] and 3[Hz], because it is difficult for a subject to move a hand 4[Hz]. A metronome was used in this section, and the experiments were performed, checking hand speed.

Table II shows the experimental results. The average of the diagonal component of Table II is 73% for 2[Hz] and 76% for 3[Hz]. It is also shown that a similar recognition rate is



(a) Horizontal (b) Vertical (c) Clockwise (d) Counterclockwise

Fig. 12. Image sequence of each periodic gesture and recognition result (from top to bottom)

obtained for different hand speeds. The recognition rate for 3[Hz] is high compared to that for 2[Hz].

IV. Conclusions and Future Work

In this study, we have proposed a method to recognize periodic gestures. We realized the recognition of periodic gestures by utilizing the amplitude and phase spectrum that are obtained by FFT to a time series of intensity images. The method does not require color information and is robust for different degrees of illumination and different subjects. The method is also robust for hand speed. The experiments show the stability and robustness of the proposed method.

Future studies will include the improvement of the recognition results, application to Intelligent Room [18].

References

[1] A. Efros, A. Berg, G. Mori, and J. Malik: "Recognizing action at a distance," Proc. of IEEE Int. Conf. on Computer Vision, pp.726-733, 2003.

[2] A. Briassouli, and N. Ahuja: "Estimation of Multiple Periodic Motions from Video," Proc. European. Conf. Computer Vision, pp. 147-159, 2006.

[3] D. J. Sturman and D. Zeltzer: "A Survey of Glove Based Input," IEEE Computer Graphics and Applications, vol.14, no.1, pp.30-39, 1994.

[4] C. Vogler and D. Metaxas: "ASL Recognition Based on a Coupling Between HMMs and 3D Motion Analysis," Proc. of IEEE Int. Conf. on Computer Vision, pp.363-369, 1998.

[5] R. Polana and R.C. Nelson: "Detection and Recognition of Periodic, Nonrigid Motion," Proc. of International Journal of Computer Vision, vol.23, no.3, pp.261-282, 1997

[6] R. Cutler and L. Davis: "Robust Real-Time Periodic Motion Detection, Analysis, and Applications," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.22, no.8, pp.781-796, 2000.

[7] J. Yamato, J. Ohya, and K. Ishii: "Recognizing Human Action in Time-Sequential Images using Hidden Markov Model," Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition, pp.379-385, 1992.

[8] Y. Sato, M. Saito, and H. Koike: "Real-Time Input of 3D Pose and Gestures of a User's Hand and Its Applications for HCI," Proc. of IEEE Virtual Reality Conference, pp.79-86, 2001.

[9] J. Davis and A. Bobick: "The representation and recognition of human movement using temporal templates," Proc. of Computer Vision and Pattern Recognition, 1997.

[10] Y. Hu, L. Cao, F. Lv, S. Yan, Y. Gong, T. S. Huang: "Action Detection in Complex Scenes with Spatial and Temporal Ambiguities," Proc. of IEEE Int. Conf. on Computer Vision, 2009.

[11] P. Wang, G. D. Abowd and J. M. Rehg: "Quasi-Periodic Event Analysis for Video Retrieval," Proc. of IEEE Int. Conf. on Computer Vision, 2009.

[12] K. Irie and K. Umeda: "Detection of Waving Hands from Images - Application of FFT to time series of intensity values," Proc. 3rd China-Japan Symposium on Mechatronics, pp.79-83, 2002.

[13] W. Press and S. Teukolsky, W. Vetterling and B. Flannnery: Numerical Recipes in C. Cambridge University Press, 1988.

[14] V. Vapnik: Statistical learning theory. John Wiley and Sons, 1998.

[15] M. Oren, C. Papageorgiou, P. Sinha, E. Osum, and T. Poggio: "Pedestrian detection using wavelet templates," Proc. of Computer Vision and Pattern Recognition, pp.193-197, 1997.

[16] E. Osuna, R. Freund, and F. Girosi: "Training Support vector machinces: An application to face detection," Proc. of Computer Vision and Pattern Recognition, pp.130-136, 2000.

[17] Christopher M. Bishop: Pattern Recognition and Machine Learning. Springer, 2006.

[18] K. Irie, N. Wakamura, and K. Umeda: "Construction of an Intelligent Room Based on Gesture Recognition -Operation of Electric Appliances with Hand Gestures," Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp.193-198, 2004.

[19] C.-C. Chang and C.-J. Lin, LIBSVM: "A Library for Support Vector Machines," http://www.csie.ntu.edu.tw/~ cjlin/libsvm/.