Detection of Waving Hands from Images Using Time Series of Intensity Values

Kota IRIE, Kazunori UMEDA

Chuo University, Tokyo, Japan
irie@sensor.mech.chuo-u.ac.jp, umeda@mech.chuo-u.ac.jp

Abstract

This paper proposes a method of detecting waving hands from images as a method for man-machine interface. FFT is applied to time series of intensity images. The images are converted to low-resolution ones, and FFT is applied to each pixel of the low-resolution images. The proposed method is robust to lighting condition and individual difference of skin color, because it doesn't use color information at all. Experiments show the stability and robustness of the proposed method.

Key Words: Image Processing, Gesture Recognition, FFT, Low-Resolution Image, Man-Machine Interface

1. Introduction

So as to realize man-machine interface that is natural for an operator, it is important to recognize the existence of the operator and his/her intention of operation. Hand waving is often used for man-man interface to communicate one's intention to other person, and thus it is thought to be efficient for the purpose.

There are several studies that deal with detecting waving objects[1][2]. For detection of hand regions, color information is usually utilized[3][4][5]. However, this methodology is sensitive to lighting condition and individual difference of skin color, because it is indispensable to extract the skin color from images.

In this paper, we propose a method to recognize waving hands from images that doesn't utilize color information. FFT(Fast Fourier Transform)[6] is applied to time series of intensity images. The images are converted to low-resolution ones, and FFT is applied to each pixel of the low-resolution images. The proposed method is robust to lighting condition and individual difference of skin color, because it doesn't use color information at all. Additionally, the method is very simple, because it doesn't require image processing to recognize hand regions.

2. Cyclic change of intensity values of images caused by hand waving

Hand waving is described as the horizontal cyclic motion of a hand, with the frequency of usually 3 or 4Hz. When a hand is waved, the intensity value of a pixel corresponding to the hand region vibrates between hand region and background. As a pre-processing, we make the resolution of the image lower. By this process, the pattern of the vibration is smoothed as shown in Fig.1, and additionally, the robustness for noises is acquired and calculation cost is reduced.

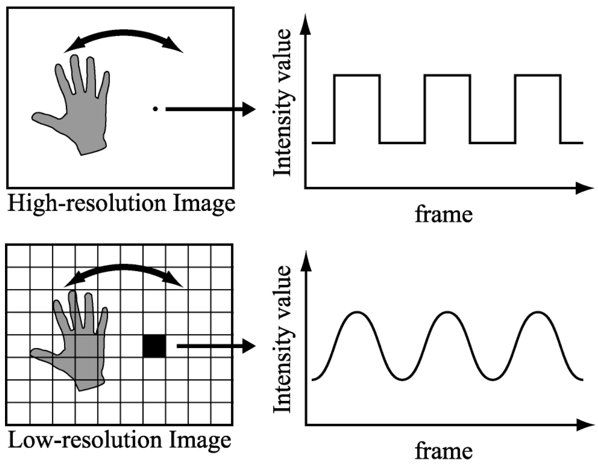


Fig.1 Cyclic change of intensity values by hand waving

The extent of converting an image to low-resolution is evaluated as follows. Suppose the distance to the hand is L[m], the width of hand waving is L_H[m], the horizontal angle of the camera is θ[rad], and number of horizontal pixels of the obtained image is a. Then the width of hand waving in the image H[pixel] (see Fig.2) is obtained by

$$H = \frac{aL_H}{2\tan\frac{\theta}{2}} \cdot \frac{1}{L}. \tag{1}$$

It is necessary that width of hand waving in the low-resolution image is roughly larger than one pixel. Therefore, $H > P_{lim}$ should be satisfied, where P_{lim} is the number of horizontal pixels assigned to the pixel of the low-resolution image.

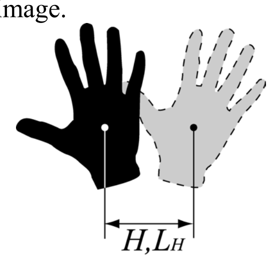


Fig.2 Width of hand waving

3. Application of FFT to time series of intensity values

As mentioned in Chapter 2, each image is converted to low-resolution, and time series of low-resolution images are obtained. Suppose the number of pixels of the images is m×n, and $I(i,j,t)$ is the intensity value of (i,j) pixel (i=1,2,…,m, j=1,2,…,n) of t-th frame, as shown in Fig.3.

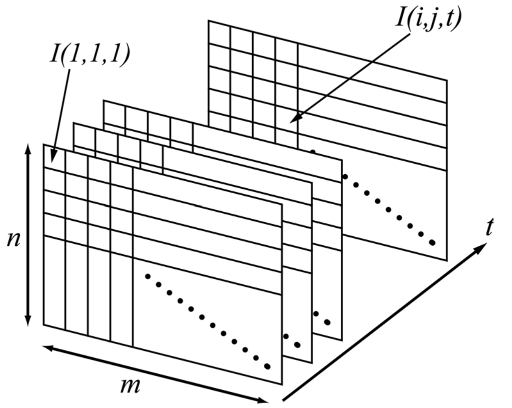
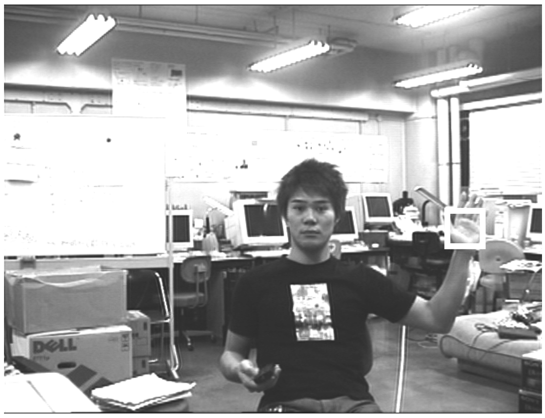


Fig.3 Time series of low-resolution images

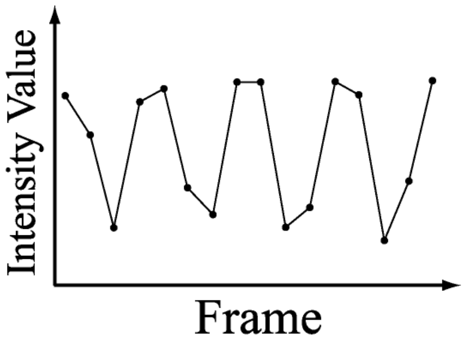
Fig.4 illustrates the application of FFT to low-resolution images. Original image Fig.4(a) is converted to the low-resolution image Fig.4(b). The pixels in the rectangle of Fig.4(b) correspond to the region of the waving hand. The intensity value $I(i,j,t)$ of these pixels changes as illustrated in Fig.4(c), since the rate of the hand and the background changes periodically, according to the waving of the hand. As this change of intensity value is periodic with a constant cycle, we can utilize FFT for quantifying. We apply FFT to intensity values of every pixels $I(i,j,t)$, and detect waving hands from the spectrum, illustrated as Fig.4(d). To remove the effect of noises like flicker of fluorescent light and reduce the calculation cost, FFT is applied to the pixels that satisfie

$$\left|I(i,j,t+1)-I(i,j,t)\right| \geq I_{dif} . \qquad (2)$$

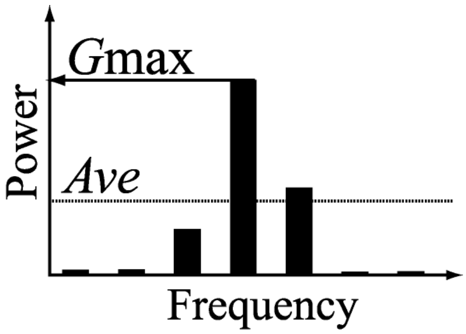


(a)Input image

(b)Low-resolution image



(c)Time series of intensity value



(d) Power spectrum

Fig.4 Application of FFT to time series of intensity values

4. The method of recognizing waving hands

4.1. Features

Features are extracted from the power spectrum obtained from the time series of intensity values. We utilize the two features: the maximum value $G_{\max}$ and the mean value *Ave* of the power of the spectrum. $G_{\max}$ and *Ave* are given by eq.(3) and eq.(4) respectively, where N is the sampling number and W is the twiddle factor of DFT(Discrete Fourier Transform). max(f) represents the maximum value of f.

$$G_{\max} = \max\left(\frac{1}{N}\sum_{k=1}^{N-1} I_k W_{nk}\right) \tag{3}$$

$$Ave = \frac{2}{N}\sum_{i=1}^{N/2} G_i \tag{4}$$

4.2. Recognition by discriminant analysis

The linear discriminant method[7] is applied to the feature space of $G_{\max}$ and *Ave* for recognition of waving hands.

Suppose the feature vector is $\mathbf{x}=[G_{\max},\ Ave]^t$, the class of waving hand is ω_1, the class of other motions is ω_2, and average vector of each class is $\mathbf{m}_1,\mathbf{m}_2$. Then the scatter matrices of each class $\mathbf{S}_1$ and $\mathbf{S}_2$ are defined by eq.(5).

$$\mathbf{S}_i \equiv \sum_{\mathbf{x}\in\chi_i} (\mathbf{x}-\mathbf{m}_i)(\mathbf{x}-\mathbf{m}_i)^t \tag{5}$$

Using all feature vectors of two classes, the within-class scatter matrix and the between-class scatter matrix are defined by eq.(6) and eq.(7) respectively, where n_i is the number of samples of ω_i and $\mathbf{m}$ is the average vector of every sample.

$$\mathbf{S}_W \equiv S_1 + S_2 = \sum_{i=1,2}\sum_{\mathbf{x}\in\chi_i} (\mathbf{x}-\mathbf{m}_i)(\mathbf{x}-\mathbf{m}_i)^t \tag{6}$$

$$\mathbf{S}_B \equiv \sum_{i=1,2} n_i(\mathbf{m}_i-\mathbf{m})(\mathbf{m}_i-\mathbf{m})^t \tag{7}$$

The linear discriminant function $g(\mathbf{x})$ is given by

$$g(\mathbf{x}) = \mathbf{A}^t\mathbf{x} + a_0\ ,\ \mathbf{A} = \mathbf{S}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2)\ . \tag{8}$$

To define the threshold a_0 in eq.(8), we select the method to divide internally with the standard deviation of each class[8]. Therefore,

$$a_0 = \frac{\tilde{m}_1\tilde{\sigma}_2 + \tilde{m}_2\tilde{\sigma}_1}{\tilde{m}_1 + \tilde{m}_2}\ , \tag{9}$$

where $\tilde{m}_i$ and $\tilde{\sigma}_i$ are the average and the standard deviation of points in $G_{\max}$ and *Ave* space projected on normal vector of decision boundary line.

When $g(\mathbf{x}) < 0 \quad (x \in \omega_1)$ for eq.(8), the pixel is regarded as corresponding to the waving hand. So as to make the recognition more robust, the detection of hand waving is done when $g(\mathbf{x})$<0 continues for several frames. It is formulated as eq.(10) and (11).

$$\text{if } g(\mathbf{x})_t<0 \text{ then } D_t=1 \text{ else } D_t=0 \tag{10}$$

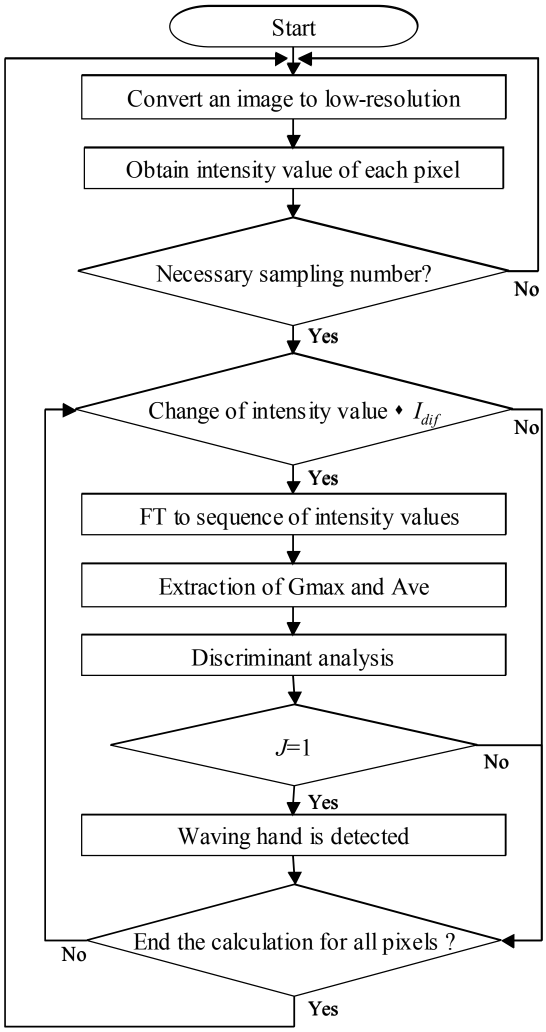


Fig.5 Flow chart of detecting waving hands

$$J = \prod_{k=t}^{t+c} D_k\ ,\ \text{if } J=1 \text{ then hand waving} \tag{11}$$

Fig.5 shows the flow of the recognition of waving hands described in Chapter 2 to 4. Note that FFT is applied to every pixel (except for the pixel with constant value) of low-resolution image, individually.

5. Experiments of detecting waving hands

In this Chapter, we show some experimental results to show the effectiveness of the proposed method to detect waving hands. Every calculation, including FFT for every pixel and recognition, is performed by a PC (Pentium IV 1.4GHz). For inputting images and converting the images to low-resolution, we used an image board PicPort Color(Leutron Vision) and an image processing software HALCON(MVTec). As a CCD camera, we used EVI-G20 (SONY) that has the function of pan-tilt tracking. Therefore, we can pan and tilt the camera so as to move the detected waving hand at the center of the image.

The number of sampling N was set to 16. I_t in eq.(2) was set to 5[pixel]. The sampling period was about 80[ms].

5.1. Decision of image resolution

The resolution of the images was decided by eq.(1). The parameters are, a=640[pixel], $\theta \cong \pi/4$[rad], and H_l =0.3[m]. We set the maximum measurement distance to 8[m]. Then H becomes 29[pixel]. Therefore, we set P_{lim} to 25[pixel] that is less than H, and 25×25 pixels of the original image was assigned to one pixel of the low-resolution image. As a result, the number of the pixels of the low-resolution image was set to 25×19[pixel].

5.2. Decision of the linear discriminant function

The cluster of hand waving ω_1 and the cluster of other motions ω_2 were formed by experiments. The other motions were various random motions, e.g., walking randomly in the room. Fig.6 and Fig.7 show the distribution of $G_{\max}$-Ave for ω_1 and ω_2 respectively. The distance was set from 3 to 8[m]. The number of data for ω_1 and ω_2 was about 2500. Within-class scatter matrix $\mathbf{S}_W$ and average of pattern $\mathbf{m}_i$ of eq.(6) were obtained as follows.

$$\mathbf{S}_W = \begin{bmatrix} S_{G\max,G\max} & S_{G\max,Ave} \\ S_{G\max,Ave} & S_{Ave,Ave} \end{bmatrix} = \begin{bmatrix} 1296.3 & 4616.3 \\ 4616.3 & 2310.2 \end{bmatrix}$$

$$\mathbf{m}_1 = \begin{bmatrix} \overline{G_{\max}} \\ \overline{Ave} \end{bmatrix} = \begin{bmatrix} 214.6 \\ 75.6 \end{bmatrix}, \mathbf{m}_2 = \begin{bmatrix} \overline{G_{\max}} \\ \overline{Ave} \end{bmatrix} = \begin{bmatrix} 61.5 \\ 48.6 \end{bmatrix}$$

Then the matrix A in eq.(8) became

$$\mathbf{A} = \begin{bmatrix} -0.02657 \\ 0.04141 \end{bmatrix}$$

and a_0 in eq.(9) became –0.143. Consequently, the linear discriminant function was given as

$$g(\mathbf{x}) = \begin{bmatrix} -0.02657 \\ 0.04141 \end{bmatrix}^t \begin{bmatrix} G_{\max} \\ Ave \end{bmatrix} - 0.143 . \qquad (12)$$

Here we evaluate the obtained $g(\mathbf{x})$. We define the rate of discriminant error p_i as

$$p_i = \frac{nw_i}{n_i} . \qquad (13)$$

where n_i is the number of samples in cluster i and nw_i is the number of discriminant error. The rates for the data in Fig.6 and Fig.7 are, p_1=4.20% (type 1 error: recognizing the pixel for hand waving as for other motion), and p_2=2.28% (type 2 error: recognizing the pixel for other motion as for hand waving). The error rates are pretty small, and additionally, the rate can be improved by considering the series of $g(\mathbf{x})$ as described in Section 4.2.

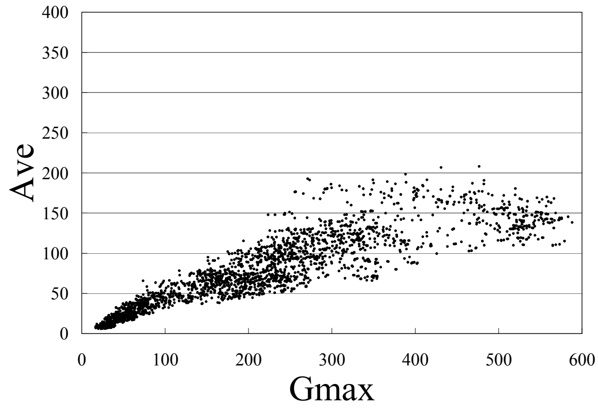


Fig.6 Distribution of waving hand: ω_1

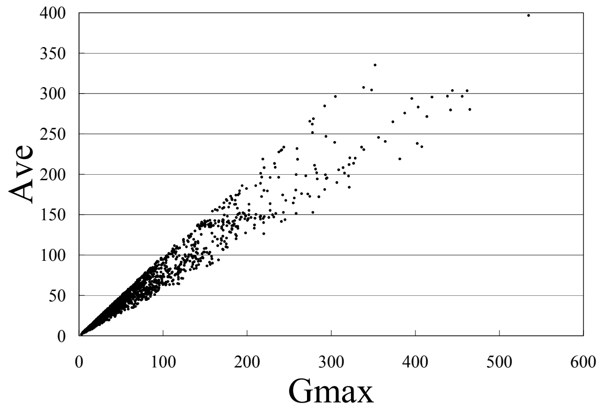


Fig.7 Distribution of other motions: ω_2

5.3 Detection of waving hands

5.3.1. Recognition rate of hand waving

Experiments were performed for 5 subjects, by changing distance and lighting condition. Fluorescent lights were used for lighting. The illumination around the hand was 60-190[lux] (condition 1: dark), and 250-315[lux] (condition 2: bright). The executed motion was as follows.

(1) Waving a hand at arbitrary positions in the camera angle for about two seconds,
(2) Suspending the hand waving for about two seconds, and then next waving.

The motion was repeated for 20 times. When the hand waving was detected in two seconds, we regarded the recognition of hand waging was successful. Table 1 shows the experimental results. It is shown that pretty high recognition rate is realized for the range of 4-8[m]. When the distance becomes larger (7,8[m]), the width of hand waving becomes small and recognition rate becomes lower. Additionally, it is shown that the results are better for condition 1, darker one. This rather strange phenomenon is caused by the difference of illumination between hand region and background. At the darker condition, the background wall was much darker, 30-50[lux], and the difference of intensity between hand region and background was larger. On the contrary, the background wall at the brighter condition was

Fig.8 Distance: 4m, Lighting: Bright

Fig.9 Distance: 4m, Lighting: Dark

Fig.10 Distance: 8m, Lighting: Bright

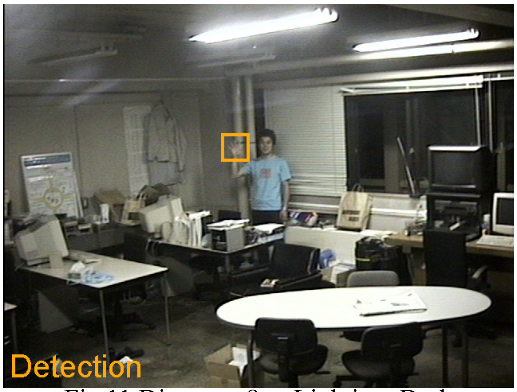


Fig.11 Distance: 8m, Lighting: Dark

230-300[lux], and the difference of intensity was smaller.

Fig.8-11 shows the example of detection of a hand waving.

Table 1 Recognition rate of hand waving

Distance	60-190 lux	250-315 lux
4m	99%	96%
5m	100%	96%
6m	100%	97%
7m	96%	92%
8m	91%	83%

5.3.2. Experiments for recognition error

So as to evaluate the robustness of the proposed method, we made experiments for recognition error. One person made various motions except hand waving, e.g., walking randomly in the room. The distance varied from 1 to 8[m]. The motions were performed continuously for 600 seconds. In this experiments, the recognition error did not occur at all. The flicker of the fluorescent light did not affect the results. This result indicates the robustness of the proposed method.

6. Conclusions

We proposed a method to detect waving hands, which doesn't require color information and is robust for illumination. FFT is applied to time series of intensity values for low-resolution images. The calculation cost is not high, and the method is applicable for practical use. Because FFT can be constructed by hardware, it is possible to make the calculation cost much lower. Experiments show the stability and robustness of the proposed method. Future works include more experimental evaluation, and the improvement as the problem of pattern recognition.

References

[1] R. Cutler, L.Davis, "View-based Detection and Analysis of Periodic Motion", *Proc. of the International Conference on Pattern Recognition*, 1998.

[2] R. Cutler, L. S. Davis, "Robust Real-Time Periodic Motion Detection, Analysis, and Applications", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, pp. 781-796, 2000.

[3] J. Sherrah and S. Gong, "VIGOUR: A System for Tracking and Recognition of Multiple People and their Activities", *Proc. of the International Conference on Pattern Recognition*, 2000.

[4] P. Hong, M.Turk, T. S. Huang, "Gesture Modeling and Recognition Using Finite State Machines", *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, 2000.

[5] H. Wu, T. Shioyama, and H. Kobayashi, "Spotting Recognition of Head Gestures from Color Image Series", *Proc. of the International Conference on Pattern Recognition*, pp.83-85, 1998.

[6] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C*, Cambridge University Press, 1988.

[7] Duda, R.O. and Hart, P.E. *Pattern Classification and Scene Analysis*, John Wiley & Sons, 1973.

[8] Fukunaga K., Introduction to Statistical Pattern Recognition (2nd ed.), Academic Press, 1990.