Face Image Recognition by 2-Dimensional Discrete Walsh Transform and Neural Network

Masahiro Yoshida†, Takeshi Kamio‡ and Hideki Asai†

†Dept. of Systems Eng., Shizuoka Univ., Hamamatsu 432-8561, Japan
Phone: +81-53-478-1237, Fax: +81-53-478-1269
Email: hideasai@sys.eng.shizuoka.ac.jp
‡Dept. of Inf., Machines and Interfaces, Hiroshima City Univ., Hiroshima 731-3194, Japan
Email: kamio@im.hiroshima-cu.ac.jp

1. Introduction

A variety of studies on pattern recognition have been done [1]-[3]. In this field, the researches of the face image recognition also have been done from the viewpoint of security system, database retrieval and so on. However, a computational model of face image recognition is difficult. Therefore, the practical system has been hardly reported.

The conventional pattern recognition has been done based on the extraction techniques of many features and relations among the positions each other. The face image recognition also uses feature-based techniques. These techniques consist of shapes and relations among the positions of eyes, mouth and so on. However, in face image recognition, it is difficult to extract many features exactly because of unclear border of eyes, mouth and so on.

Therefore, the face image recognition by using multi-layer neural networks (MNN)[4], which have been frequently used as a powerful tool for pattern classification, has been done. In the previous researches, face image recognition, which is used on the gray levels on each pixel of the whole face image, has high recognition accuracy without feature extraction [5]-[7]. However, in these methods, a lot of data which are not neccessarily useful for recognition are inputted. Therefore, MNN needs a large number of neurons and training is difficult. If it is possible to recognize by using a part of data in the face image, training would become easier. In this research, we make use of orthogonal transform by which the conventional image processing has been done, where the 2-dimentional discrete Walsh transform (2D-DWT) is used in order to transform the face image into frequency characteristics and reduce the data to be inputted. The 2D-DWT easily transforms the image into frequency characteristics because the calculation of 2D-DWT does not require multiplications. We try to recognize the face image by using MNN, where a part of data, which is the 2D-DWT of the face image, is given as the input to the input layer.

The organization of the rest of the paper is as follows. In Section 2 we describe the 2D-DWT. The face image recognition is described in Section 3. In Section 4, we present the validity of our face image recognition from the simulation results. Section 5 contains the summary of the paper.

2. 2-dimentional discrete Walsh transform

The discrete Walsh transform (DWT) is one of the most important techniques as well as the discrete Fourier transform in the field of signal processing [8]-[9]. The DWT works well for digital signals due to the fundamental function called the Walsh function. The Walsh function has only ±1, and is the system of orthogonal functions. In general, the Walsh function can be generated by the Kronecker's product of the Hadamard matrix **H**'s.

First, the 2-by-2 Hadamard matrix $\mathbf{H_2}$ is defined by

$$\mathbf{H}_2 = \begin{bmatrix} + & + \\ + & - \end{bmatrix}, \tag{1}$$

where the symbols + and - mean +1 and -1, respectively. Furthermore, calculating the Kroneker's product between two $\mathbf{H_2}$'s, the 4-by-4 Hadamard matrix $\mathbf{H_4}$ is easily given as follows:

$$\mathbf{H}_2 \otimes \mathbf{H}_2 = \begin{bmatrix} +\mathbf{H}_2 & +\mathbf{H}_2 \\ +\mathbf{H}_2 & -\mathbf{H}_2 \end{bmatrix}$$

$$= \begin{bmatrix} + & + & + & + \\ + & - & + & - \\ + & + & - & - \\ + & - & - & + \end{bmatrix} = \mathbf{H}_4, \qquad (2)$$

where the symbol $\otimes$ indicates the Kroneker's product. Similarly, 2^k-by-2^k Hadamard matrix is obtained by

$$\mathbf{H}_{2^k} = \mathbf{H}_2 \otimes \mathbf{H}_{2^{k-1}} = \mathbf{H}_2 \otimes \mathbf{H}_2 \otimes \cdots \otimes \mathbf{H}_2 . \quad (3)$$

Frequency characteristics can be given by the Hadamard matrix. Along each row of the Hadamard matrix the frequency is expressed by the number of changes in sign. The number of changes is called “sequency”. The sequency has the characteristics similar to the frequency.

The Walsh function can be expressed as each row of **H** ($\in R^{n\times n}$). Therefore, DWT is known as a kind of the Hadamard transform, where **H** has some useful following characteristics.

i) $\mathbf{H} = \mathbf{H}^T$, (**H** is symmetric.)
ii) $\mathbf{H}\mathbf{H}^T = \mathbf{H}^T\mathbf{H} = n\mathbf{I}$. (**H** is orthogonal.)

Thus, the DWT and the inverse DWT are defined as follows:

DWT: $$\mathbf{V} = \frac{1}{n}\mathbf{H}\mathbf{B}, \qquad (4)$$

IDWT: $$\mathbf{B} = \mathbf{H}\mathbf{V}, \qquad (5)$$

where n is the number of sampled points, **B** is the sampled data vector, **V** is the DWT of **B**, and **H** is Hadamard matrix, i.e. Hadamard-ordered Walsh functions.

The 2D-DWT does the DWT toward the image of m-by-n pixels. The 2D-DWT and the 2D-IDWT are defined as follows:

2D-DWT: $$\mathbf{F} = \frac{1}{n}\mathbf{H}_{\mathbf{m,m}}\mathbf{f}\mathbf{H}_{\mathbf{n,n}}, \qquad (6)$$

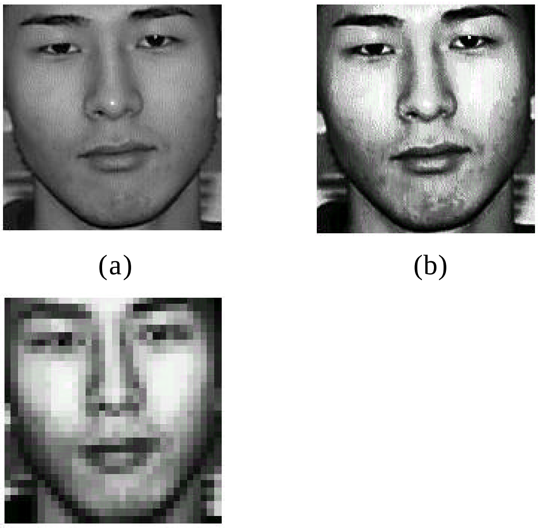
2D-IDWT: $$\mathbf{f} = \frac{1}{n}\mathbf{H}_{\mathbf{m,m}}\mathbf{F}\mathbf{H}_{\mathbf{n,n}}, \qquad (7)$$

where $\mathrm{f} \in R^{m\times n}$ is the sample data matrix and $\mathbf{F} \in \mathbf{R}^{m\times n}$ is the 2D-DWT of **f**. In case of orthogonal transform of the image, the 2D-DWT is more efficient than the DWT. However, to use 2D-DWT, the row and column numbers of sample data matrix must be 2^n (n is a natural number) respectively, because Hadamard matrix can be generated by the Kronecker's product of Hadamard matrix $\mathbf{H}_2$.

3. Face image recognition by multi-layer neural network

A. Extraction of input data from face image

Consider the face image given by the digital picture, which has 256 gray levels as shown in Fig.1(a). At first, histogram modification is done as the gray scale transformation in order to decrease the exposure effect and emphasize contrast as shown in Fig.1(b). Next, the image, for which histogram modification was done, is divided into many subareas, in order to decrease the disagreement of position between the face images and in order to reduce the calculation cost toward 2D-DWT. Furthermore, a certain gray level is determined by taking an average for all the pixels in each subarea as shown in Fig.1(c). This is defined as mosaic processing. In mosaic processing, the numbers of subareas affect recognition accuracy. By this procedure, 32-by-32 subareas are obtained in mosaic image, namely 1024 subareas. The gray level in each subarea is normalized in the range of [0,1]. Furthermore the mosaic image is divided into 16 groups, where one group is 8×8 subareas, and the 2D-DWT is done on each group. By doing 2D-DWT on each group, the calculation cost is reduced. As a result, only the energy with low frequencies can be extracted, and we can apply as the input data to MNN. Mosaic processing is almost equivalent to removal of high frequency elements. In general, the image has the characteristics that energy of low frequencies is big and the one of high frequencies is small. The energy of low frequencies has the rough characteristics. Based on these facts, we extract some energies that correspond to low frequencies in each group and input them to MNN. However,



(a) (b) (c)

Fig.1 (a) the original image. (b) the image after histogram modification. (c) the mosaic image.

DC component is ignored. As a result, the number of neurons at the input layer can be reduced. In our simulation, the number of extraction data is changed for the number of the faces to be recognized. We investigate the suitable numbers of extraction data which correspond to the numbers of neurons and try to recognize several decades of face images.

B. Leaning and recognition

In this research, 3-layer neural network [11], which is constructed by the input, hidden and output layers, is used. And we use the moment method that is one of the back propagation learning methods. The back propagation method is the most frequently used as the training procedure for MNN. In this network, the number of neurons at the input layer is same as the number of the extraction data, the number of neurons at the hidden layer is changed on observation, and the number of neurons at the output layer is n, where n is the number of faces to be recognized. The target signal that is fed to the neurons at the output layer is decided corresponding to the input image. For target signal, a face image to be recognized corresponds to a neuron at the output layer. In the learning process, the value 0.95 is given for the neuron at the output layer that corresponds to the input face image and the value 0.05 is done for other neurons. The initial synaptic weights are random values ranging from -0.1 to 0.1 and are determined through the above training. Using MNN, which completed training, the face image recognition is done. Discrimination of the input image is done by the highest output value in output layer. When an unclear image is inputted, the difference between the highest value and the second value is small and both of them are not near to 1.0. In this case, the discrimination based on only the highest value in output layer causes a mistake. Therefore we judge that the recognition is impossible for the input image, which is satisfied with the following condition.

$$\frac{(\max 1 - \max 2)}{\max 1} < 0.6 , \tag{8}$$

where max1 and max2 are the highest and the second values in output layer respectively.

C. A strategy for recognition of the unlearned face images

The method mentioned above enables to recognize learned face images. On the other hand, from the viewpoint of pattern recognition, unlearned face images have to be distinguished from the learned face images. Then, one neuron is added to the output layer. Henceforth, in this paper, this neuron is called "unlearned neuron". This neuron is used for recognition of the unlearned face images. Indeed many face images except for the learned images are learned as samples of unlearned face images and the synaptic weights of this neuron are determined. As a result, an unlearned image can be detected by fire of the unlearned neuron.

4. Simulation results

Two kinds of face images per person have been prepared for simulations. One kind of face image is used for learning the MNN. Another one is used for the test data. We have examined two types of recognition.

The first attempt is as follows. At first, we extract the energy with the lowest frequency from every group for which 2D-DWT was done. In other words, 16 data extracted from every group are inputted to MNN. By using the network configuration of 16-35-32, we could recognize correctly 32 face images. Next, we extract the energies with the lowest and second frequencies from every group for which 2D-DWT was done. Namely using 32 data, 46 face images could be recognized correctly by the network configuration of 32-35-46.

The second one is as follows. 30 face images were used for learning and other 10 face images were used for learning of unlearned neuron. We tried to recognize 60 test data. We extract the energies with the lowest and second frequencies from ever group for which 2D-DWT was done. When the network configuration was 32-35-31, we could recognize correctly 60 face images.

As long as we know, we have never had any reports, which have shown the complete recognition for many face images by a small number of neurons. We have verified that our method enables to recognize precisely many face images even by a small scale of MNN.

5. Conclusions

In this paper, face image recognition using 3-layer neural network and the 2D-DWT has been described. The face image recognition by MNN is much simpler compared with the method by using the extraction techniques. First, we explained the 2D-DWT. Next, preprocessing for the given image, network structure and the learning method have been shown. Finally, we have tried to test the face image recognition by using a part of frequency characteristic data, which is obtained by the 2D-DWT of face image. It has been shown that, our method can recognize correctly 46 face images. In addition, the scale of MNN is smaller than the scales in the conventional methods. Therefore, we have confirmed that the present face image recognition is valid.

References

[1] Turk M, Pentland A., “Face recognition using eigenfaces” Proc. CVPR'91, pp.586-591, 1991.

[2] R. Brunelli, T. Poggio, “Face recognition: Features versus templates” IEEE Trans. Patt. Anal. Machine Intell., vol.15, no.10, pp.1042-1052, Oct. 1993.

[3] Tatsuo Sato, Kiotaka Miura, Yohko Fukuda, “Face Area Segmentation using Partial Features and Face Recognition” Technical report of IEICE, PRMU97-70, pp103-110, 1997.

[4] R. P. Lippmann, “An Introduction to Computer with Neural Nets” IEEE, ASSP MAGAZINE, April 1987.

[5] M. Kosugi, “Human-Face Recognition Using Mosaic Pattern and Neural Networks” IEICE Trans. Fundamentals, vol. J76-D-2, No.6, pp.1132-1139, June 1993.

[6] M. Kosugi, K. Yamashita, “Automated Identification System of Person Robust for Face Direction by Using Gray-Level Mosaic of a Face” IEICE Trans. Fundamentals, vol. J80-D-2, No.8, pp.2244-2247, August 1997.

[7] J. L. Perry, “Human face recognition using a multilayer perceptron” Proc. IJCNN90- WASH, 2, Jan. 1990.

[8] T. Kamio, H. Ninomiya,H. Asai, “A Neural Net Approach to Discrete Walsh Transform” IEICE Trans. Fundamentals, vol. E77-A, no.11, pp.1882-1886, Nov. 1994.

[9] Mar H. T. L., Sheng C. L., “Fast Hadamard transform using the H diagram” IEEE Trans. Conput., vol. C-22, pp.957-960, 1973.

[10] W. K. Pratt, J. Kane, H. C. Andrews, ”Hadamard Transform Image Coding” Proc.IEEE, 57, 1, pp58-68, 1969.

[11] Masahiro Yoshida, Takeshi Kamio, Hideki Asai, “Neuro-Based Human-Face Recognition with 2-Dimensional Discrete Walsh Transform” IJCNN 2000, July 2000.