2003 International Symposium on Advanced Intelligent Systems
September 25-28, 2003, Jeju, Korea

Face Recognition Using Fuzzy Fusion and Wavelet Decomposition Method

Keun-Chang Kwak[1,2] , Jun-Oh Min[1], Myung-Geun Chun[1]
Witold Pedrycz[2]
[1]Dept. of Electrical and Computer Engineering, Chungbuk National University, Cheongju, Korea
[2]Dept. of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada
San 48, Gaesin Dong, Heungduk Ku, Chungju, Chungbuk, Korea, 361-763
e-mail : kwak@ee.ualberta.ca

Abstract - **In this study, we develop a method for recognizing face images by combining wavelet decomposition, fisherface method, and fuzzy integral. The proposed approach comprises of four main stages. The first stage uses the wavelet decomposition. As a result of this decomposition, we obtain four subimages. The second stage of the approach applies a fisherface method to these four subimage sets. The two last phases are concerned with the generation of the degree of fuzzy membership and the aggregation of the individual classifiers by means of the fuzzy integral. The experimental results obtained for the CNU and Yale face databases reveal that the approach presented in this study yields better classification performance in comparison to the results produced by other classifiers.**

1. Introductory comments

Face recognition is one of the most interesting and challenging areas in computer vision and pattern recognition. The popular approaches for face recognition are eigenface and fisherface method. Eigenface method or Principal Component Analysis (PCA) is most well known methods for face recognition [1]. Each of them comes with some advantages but is not free from limitations and drawbacks when cast in the setting of face recognition. The PCA approach exhibits optimality when it comes to dimensionality reduction however it is not ideal for classification purposes as it retains unwanted variations occurring due to lighting and facial expression [2]. To overcome this problem, proposed was an enhancement known as a fisherface method or Fisher's Linear Discriminant (FLD)[2]. This statistically motivated method maximizes the ratio of between-scatter matrix and within-scatter matrix and in this sense attempts to involve information about classes of the patterns under consideration. In general, this method is used in conjunction with the PCA where the PCA technique first projects the set of images to a lower-dimensional space so that the resulting within-class scatter matrix to be used by the FLD becomes nonsingular. There are various enhancements to the generic form of the FLD technique.

On the other hand, the recent trend of approaches in face recognition involves wavelet-based methods where these come in the form of spectroface [3], and discriminant waveletfaces [4]. These methods use images that are usually decomposed into four subimages (approximation, horizontal, vertical, and diagonal detailed image) via the high-pass and low-pass filtering. In general, an approximation (compressed) image that shows the best performance among the four subimages at the same level is used in further classification procedure [3][4]. But, a single set of approximation images may not be enough to capture the complete information due to large and unwanted variations.

In this study, we perform fisherface method based on four subimage sets obtained by wavelet decomposition. Here, the fusion of the individual classifiers is realized through fuzzy integration with fuzzy integral being employed in this construct [5][6]. The purpose of the fuzzy integral is to combine the results of multiple sources of information. Finally, we cover simulation results for the two face databases available at CNU (Chungbuk National University) and Yale .

2. Fuzzy measure and fuzzy integral

In the section, we shall introduce the properties of fuzzy measure and fuzzy integral. Fuzzy measure and fuzzy integral dwell on the concepts of fuzzy sets and are viewed as an interesting aggregation alternative applied to fuzzy sets.

2.1 Fuzzy measure

Let $Y = \{y_1, y_2, \ldots, y_n\}$ be a finite set and let $P(Y) = 2^Y$ denote the family of all subsets of Y. A set function $g : 2^Y \rightarrow [0,1]$ is called a fuzzy measure if

1) Boundary conditions: $g(\phi) = 0, g(Y) = 1$

2) Monotonicity: $g(A) \le g(B)$, if $A \subset B$ and $A, B \in P(Y)$

From this definition, Sugeno[5] introduced the g_λ fuzzy measure satisfying the following additional property

$$g(A \cup B) = g(A) + g(B) + \lambda g(A)g(B) \quad (1)$$

for all $A, B \subset Y$ and $A \cap B = \phi$, and for some $\lambda > -1$. In general, the value of the constant λ can be determined by the properties of the g_λ-fuzzy measure as follows

Let $g : Y \to [0,1]$ be a fuzzy subset of Y and let $A_i = \{y_i, y_{i+1}, \ldots, y_n\}$. Note that when g is a g_λ-fuzzy measure, the values of $g(A_i)$ can be determined recursively as

$$g(A_1) = g(\{y_1\}) = g^1 \quad (2)$$

$$g(A_i) = g^i + g(A_{i-1}) + \lambda g^i g(A_{i-1}), \quad \text{for } 1 < i \le n \quad (3)$$

Because of the boundary condition $g(\mathbf{Y}) = 1$, λ is determined by solving the following polynomial equation

$$\lambda + 1 = \prod_{i=1}^{n}(1 + \lambda g^i) \quad (4)$$

where $\lambda \in (-1, +\infty)$ and $\lambda \neq 0$. The solution can be easily obtained. Obviously we are interested in the unique root greater than −1. Thus the calculations of the fuzzy integral with respect to a g_λ-fuzzy measure can be realized once we are given with the values of the density function g^i available for the individual points.

2.2 Fuzzy integral

The fuzzy integral taken over $\mathbf{Y}$ of the function h with respect to a fuzzy measure g is defined in the form

$$\int_A h(y) \circ g(\cdot) = \sup_{\alpha \in [0,1]} [\min[\alpha, g(\{y \mid h(y) \ge \alpha\})]] \quad (5)$$

When the values of $h(.)$ are ordered in the decreasing sequence, $h(y_1) \ge h(y_2) \ge \cdots \ge h(y_n)$, the fuzzy integral can be calculated as follows

$$\int_A h(y) \circ g(\cdot) = \max_{i=1}^{n}[\min(h(y_i), g(A_i))] \quad (6)$$

It has been shown that Eq. (6) is not a proper extension of the usual Lebesgue integral. In other words, when the measure is additive the above expression does not return the integral in the Lebesgue sense. In order to overcome this drawback, Murofushi[6] proposed a so-called Choquet integral computed in the same way

$$\int_A h(y) \circ g(\cdot) = \sum_{i=1}^{n}[h(y_i) - h(y_{i+1})]g(A_i),\ h(y_{n+1}) = 0 \quad (7)$$

3. Fisher's Linear Discriminant (FLD)

In this section, we briefly describe the fisherface method. Let a face image be a two-dimensional $n \times n$ array of containing levels of intensity of the individual pixels. An image $\mathbf{z}_i$ may be conveniently considered as a vector of dimension n^2. Denote the training set of N face images by $Z = (\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N)$. We define the covariance matrix as follows

$$R = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{z}_i - \bar{\mathbf{z}})^T = \Phi\Phi^T \quad (8)$$

$$\bar{\mathbf{z}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i \quad (9)$$

Then, the eigenvalues and eigenvectors of the covariance matrix R are calculated, respectively. Let $E = (\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_r)$ contain the r eigenvectors corresponding to the r largest eigenvalues. For a set of original face images Z, their corresponding reduced feature vectors $X = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ can be obtained as follows according to the following relationship

$$\mathbf{x}_i = E^T(\mathbf{z}_i - \bar{\mathbf{z}}) \quad (10)$$

The second processing stage is based on the use of the FLD and can be described as follows. Consider c classes in the problem with N samples; let the between-class scatter matrix be defined as

$$S_B = \sum_{i=1}^{c} N_i(\mathbf{m}_i - \bar{\mathbf{m}})(\mathbf{m}_i - \bar{\mathbf{m}})^T \quad (11)$$

where N_i is the number of samples in i'th class C_i and $\bar{\mathbf{m}}$ is the mean of all samples, $\mathbf{m}_i$ is the mean of class C_i. The within-class scatter matrix is defined as follows

$$S_W = \sum_{i=1}^{c}\sum_{\mathbf{x}_k \in C_i}(\mathbf{x}_k - \mathbf{m}_i)(\mathbf{x}_k - \mathbf{m}_i)^T = \sum_{i=1}^{c} S_{W_i} \quad (12)$$

where S_{W_i} is the covariance matrix of class C_i. The optimal projection matrix W_{FLD} is chosen as the matrix with orthonormal columns that maximizes the ratio of the determinant of the between-class matrix of the projected samples to the determinant of the within-class fuzzy scatter matrix of the projected sampled, i.e.,

$$W_{FLD} = \arg\max_{W} \frac{|W^T S_B W|}{|W^T S_W W|} = [\mathbf{w}_1 \ \ \mathbf{w}_2 \ \ \cdots \ \ \mathbf{w}_m] \quad (13)$$

where $\{\mathbf{w}_i \mid i = 1, 2, \cdots, m\}$ is the set of generalized eigenvectors (discriminant vectors) of S_B and S_W corresponding to the $c - 1$ largest generalized eigenvalues $\{\lambda_i \mid i = 1, 2, \cdots, m\}$, i.e.,

$$S_B \mathbf{w}_i = \lambda_i S_W \mathbf{w}_i \qquad i = 1, 2, \ldots, m \quad (14)$$

Thus, the feature vectors $V = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N)$ for any face images $\mathbf{z}_i$ can be calculated as follows

$$\mathbf{v}_i = W_{FLD}^T \mathbf{x}_i = W_{FLD}^T E^T(\mathbf{z}_i - \bar{\mathbf{z}}) \quad (15)$$

To complete classification of a new pattern (face) **z'**, we compute a Euclidean distance between z' and a pattern in the training set **z** that is

$$d(\mathbf{z}, \mathbf{z}') = \|\mathbf{v} - \mathbf{v}'\| \quad (16)$$

4. Wavelet decomposition

The wavelet transform has been applied to image processing and texture classification with an objective to carry out a comprehensive multiresolution decomposition. The previous works [3][4] used only approximation image among the four subimages available. This choice was motivated by an observation that this image is the best approximation to the original image within the lower-dimensional space and contains the highest energy content within the four subimages available. On the other hand, Sergent [7] found that the low-frequency band and high-frequency components band played different roles in the classification task. The low-frequency components contribute to the global description, while the high-frequency components contribute to the finer details required in the identification task. Taking this into account, we consider an approximation images as well as the three detailed images including auxiliary information.

In this paper, we use the most known Daubechies(db1), along with D4(db2, db4, db6, db8). Daubechies, invented a family of compactly supported orthonormal wavelets making a discrete wavelet analysis practicable.

Fig. 1 shows the architecture of the two-dimensional wavelet decomposition realized at level 1. Here, H and L represent the high-pass and low-pass filter, respectively. As shown in Fig. 1, we can obtain four subimages by wavelet decomposition. In the previous works [3][4], the z_{LL} image is used to carry out face recognition; the choice of the z_{LL} is dictated by its best performance among the four subimages occurring at the same level. In our study, we perform the face recognition using z_{LL} as well as the three remaining subimages z_{HH}, z_{HL}, and z_{LH}.

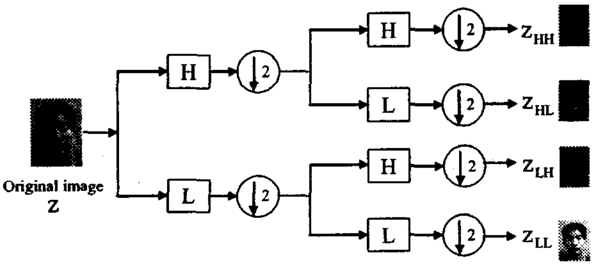

Fig.1. Architecture of two-dimensional wavelet decomposition

5. Face recognition by fuzzy integral

In this section, we combine the wavelet decomposition, fisherface method and fuzzy integral into a single coherent classification platform. The architecture of the overall face recognition system using the proposed method is shown Fig.2. In what follows, we briefly describe the proposed method.

[Step1] Perform the wavelet decomposition for the training image set to extract the intrinsic features of the patterns. Here, we obtain four sets of subimages.

[Step2] Use the fisherface method for four subimage sets, respectively. The feature vectors of the training image set and a given test image are obtained by using Eq. (15). The values of Euclidean distance are computed for the feature vectors produced from the training image set and a given test image using Eq. (16).

[Step3] Generate the membership grades $h(y_i)$ based on the distance information produced in [Step 2].

$$h(y_i) = \sum_{i \in C_t} (\mu_{ij})/N_t \qquad (17)$$

where $h(y_i^t)$ represents the between-class mean of membership grades in each classifier and N_t is the number of samples in t'th class C_t. The membership grades can be determined in many different ways; here we follow the method that their calculations use the distances between the test image and those existing in the training set, namely

$$\mu(y_{ij}) = \frac{1}{1+\left(\frac{d_{ij}}{\bar{d}}\right)} \qquad (18)$$

where i is the number of classifier and j is the index of the training face image. $\bar{d}$ denotes an average distance between all distance values, d_{ij} is the Euclidean distance of feature vector between a given test image and j'th training image in i'th classifier.

[Step 4] Aggregate the output of each classifier $h(y_i)$ and the degree of importance of each classifier g_i using the mechanisms of the fuzzy integral. The class giving rise to the highest value is declared to be the output of the classifier. The fuzzy densities g_i used in the classifier can be either estimated subjectively or obtained from the training data. Here we follow the two approaches and contrast the results. The computations of g_i based on the training data are carried out in the form of the normalized weighted sum

$$g_i = \frac{w_i p_i}{\sum_i p_i} \qquad (19)$$

where p_i is the classification rate in [0,1] of each classifier for the training set of images. These values are obtained by leave-one-out technique for training set. w_i is the subjective weight value.

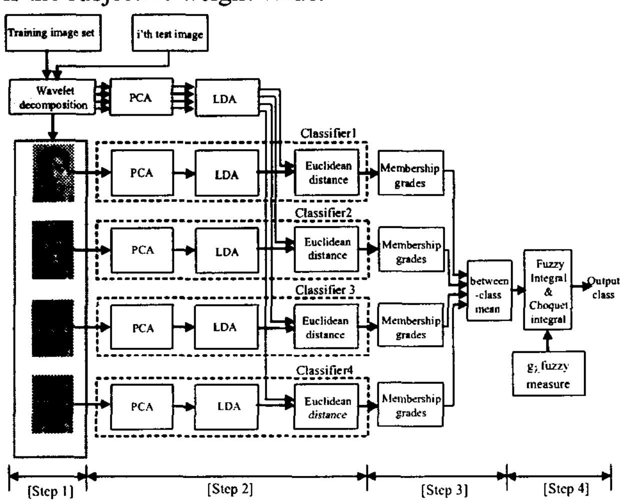

Fig.2 Overall architecture of face recognition

6. Experiments

6.1. CNU Face Databases

The CNU database contains 100 face images coming from 10 individuals. The total number of images for each person is 10. They vary in face pose and exhibit a substantial level of light variation. The size of each original image is 640×480. The wavelet decomposition was completed at level 3. Each image was digitized and presented by a 80×60 pixel (db1) array whose values of the gray levels ranged in between 0 and 255. In the following, we evaluate the performance of face recognition through a 10 fold cross-validation. We first project the image set from N-dimensional space (N-c)-dimensional space and then compute the c-1 discriminant vectors. We selected 80 eigenvectors and 9 discriminant vectors. Fig. 3 includes the pertinent plots that help visualize the differences. In the case of eigenface method, we found the recognition rate to be 84.6%. In the case of fisherface method, we noticed substantial improvement and the recognition rate of 95.4%. On the other hand, the recognition rates by fuzzy integral and Choquet integral are equal to 98% and 98.4%, respectively.

6.2. Yale Face Databases

The Yale face database contains 165 face images of 15 individuals, as shown in Figure 16. There are 11 images per subject and the size of each original image is 243×320. In this experiment we used the face image cropped and resized to remove the background information. The wavelet decomposition was applied at level 2. Each image was digitized and presented by a 61×52 pixel (db1) array. The experiments are completed in the same format as for the previous database. We selected 135 eigenvectors and 14 discriminant vectors. The plots of the classification rates are shown in Fig. 4. The fisherface method yields 96.60% classification rate whereas the integration using fuzzy integral and Choquet integral brings us close to 100% producing 99.11% and 99.24%, respectively

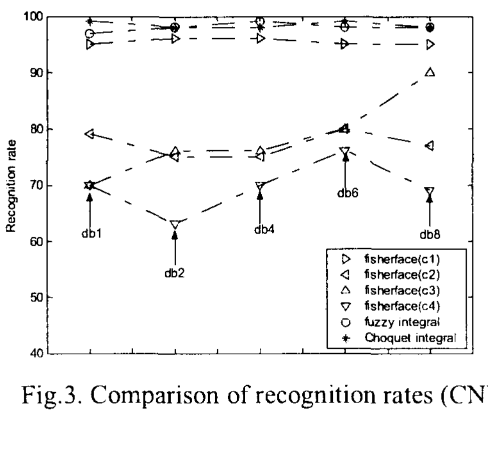

Fig.3. Comparison of recognition rates (CNU)

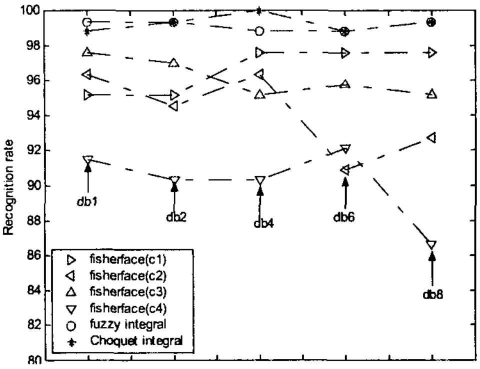

Fig.4. Comparison of recognition rates (Yale)

7. Concluding comments

We have discussed the fusion of classifiers for the face recognition problem realized with the aid of fuzzy integration of outcomes of the individual classifiers. The experiment results revealed that the use of the approximation image as well as three detailed images (including information that is unavailable in the compressed image) we were able to reduce sensitivity caused by varying illumination and viewing conditions associated with the original image.

Reference

[1] M. Turk, A. Pentland, Face recognition using eigenfaces, Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp.586-591, 1991.

[2] P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific Linear Projection, IEEE Trans. on Pattern Analysis and Machine Intell., 19(7), pp.711-720, 1997.

[3] J. H. Lai, P. C. Yuen, G. C. Feng, Face recognition using holistic Fourier invariant features, Pattern Recognition, Vol. 34, 95-109, 2001.

[4] J. T. Chien, C. C. Wu, Discriminant waveletfaces and nearest feature classifiers for face recognition, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 24, No. 2, pp.1644-1649, 2002.

[5] M. Sugeno, Fuzzy measures and fuzzy integrals-A survey, In: M.M Gupta, G.N. Saridis, B.R. Gaines (Eds.), Fuzzy Automata and Decision Processes, North-Holland, Amsterdam, pp.89-102, 1977.

[6] T. Murofushi, M.Sugeno, An interpretation of fuzzy measures and the Choquet integral as an integral with respect to a fuzzy measure, Fuzzy Sets and Systems, Vol. 29, pp.201-227, 1988.

[7] J. Sergent, Microgenesis of face perception, Aspects of Face Processing, Nijhoff, Dordrecht, 1986.