Appl. Math. Inf. Sci. 6 No. 1S pp. 81S-85S (2012)

Applied Mathematics & Information Sciences
An International Journal

Two-Dimensional Supervised Discriminant Projection Method For Feature Extraction

Jianguo Wang[1]
[1]Department of Computer Science & Technology, Tangshan College, Tangshan, China
Email Address: wjgfwjg@163.com

Received: Received May 05, 2011; Revised July 17, 2011; Accepted September 12, 2011
Published online: 1 January 2012

Abstract: For supervised discriminant projection (SDP)method, the image matrix data are vectorized to find the intrinsic manifold structure, and the dimension of matrix data is usually very high, so SDP cannot be performed because of the singularity of scatter matrix. In addition, the matrix-to-vector transform procedure may cause the loss of some useful structural information embedding in the original images. Thus, in this paper, a novel method, called 2D supervised discriminant projection (2DSDP), for face recognition is proposed. The proposed method not only takes into account both the local information of the data and the class information of the data to model the manifold structure, but also preserves the useful information of the image data. To evaluate the performance of the proposed method, several experiments are conducted on the Yale face database, and the FERET face database. The high recognition rates demonstrate the effectiveness of the proposed method.

Keywords: Feature extraction; Face recognition; Supervised discriminant projection (SDP); Two-dimensional SDP (2DSDP); Manifold learning

1 Introduction

Up to now, a number of methods for face feature extraction have been proposed [1,2]; The most well-known appearance-based projection technique is the Eigenfaces method, originally proposed by Kirby and Sirovich and popularized by Turk and Pentland [3]; Another famous projection method termed Fisherfaces was, respectively, put forward by Swets [4] and Belhumeur [5]. This method is based on linear discriminant analysis (LDA) and has shown to yield better results than Eigenfaces under variations in illumination and facial expressions.

As we know, the PCA and the LDA methods all take consideration of the global Euclidean structure of the image data. But they have little to do with the manifold structure of the data. If the data lie on a submanifold which reflects the inherent structure of the data space, it is hard for PCA and LDA to find the hidden manifold.

The UDP (Unsupervised discriminant projection) method is a recently developed face recognition method [6], which characterizes the local scatter as well as the non-local scatter and minimizes the local scatter. This characteristic makes UDP more intuitive and more powerful than the most up-to-date method, such as PCA, LDA and LPP methods. Unfortunately, a common inherent limitation is still existed in UDP method, i.e. class information is not considered in this approach. Unlike the unsupervised learning scheme of UDP, SDP [7] follows the supervised learning scheme, i.e. it uses the class information to model the manifold structure. In SDP method, the local structure of the original data is constructed according to a certain kind of similarity between data points, which takes special consideration of both the local information and the class information.

However, due to the high-dimensional and small sample size problem [8,9] encountered in face recognition, SDP method cannot be performed because of the singularity of scatter matrix. In addition, the matrix-to-vector transform procedure may cause the loss of some useful structural information embedding in the original images. Inspired by the successful application of 2DPCA[10] and 2DLDA [11] to face recognition, we proposed a novel method, called 2D supervised discriminant projection (2DSDP), to handle the above problems by directly projecting the 2D face image matrices rather than using the transformed image vectors. In the experiments conducted on two benchmark face databases, the proposed face recognition method is shown to outperform the SDP, UDP, and 2DUDP methods.

The paper of the rest is organized as follows: Section 2 gives the outline of the proposed method (2DSDP) and describes the details of our algorithm. Section 3 reports experiments carried out and the results. The last section presents our conclusions.

2 The Proposed Method 2DSDP

2.1. Outline of SDP. To make this paper more self-contained, the SDP procedure is given in this section. In contrast with UDP method, the advantage of SDP method is utilizing class information to guide the procedure of feature extraction, i.e. it uses the class information to model the manifold structure [7].

Suppose $X=[x_1,x_2,\cdots,x_M]$ is a set of M training samples in R^n. We can obtain the linear transformation φ by calculating the generalized eigenvectors of the following generalized eigen-equation:

$$S_N\varphi=\lambda S_L\varphi \tag{2.1}$$

where

$$S_L=\frac{1}{2}\frac{1}{MM}\sum_{i=1}^{M}\sum_{j=1}^{M}H_{ij}(x_i-x_j)(x_i-x_j)^T \tag{2.2}$$

$$S_N=\frac{1}{2}\frac{1}{MM}\sum_{i=1}^{M}\sum_{j=1}^{M}(1-H_{ij})(x_i-x_j)(x_i-x_j)^T \tag{2.3}$$

The similarity matrix H in Eq. (2.2) and Eq. (2.3) is defined as follows:

$$H(i,j)=\begin{cases}\exp\left(-\frac{\|x_i-x_j\|^2}{\alpha}\right)\left(1+\exp\left(-\frac{\|x_i-x_j\|^2}{\alpha}\right)\right), \\ \text{if } x_i \text{ is among K nearest neighbors of } x_j, \\ \text{and } x_j \text{ is among K nearest neighbors of } x_i \text{ and } \omega_i=\omega_j. \\ \exp\left(-\frac{\|x_i-x_j\|^2}{\alpha}\right)\left(1-\exp\left(-\frac{\|x_i-x_j\|^2}{\alpha}\right)\right), \\ \text{if } x_i \text{ is among K nearest neighbors of } x_j, \\ \text{and } x_j \text{ is among K nearest neighbors of } x_i \text{ and } \omega_i\neq\omega_j. \\ 0,\ \textit{otherwise}\end{cases} \tag{2.4}$$

where $\exp\left(-\frac{\|x_i-x_j\|^2}{\alpha}\right)$ denotes the local weight as UDP does, $\left(1+\exp\left(-\frac{\|x_i-x_j\|^2}{\alpha}\right)\right)$ and $\left(1-\exp\left(-\frac{\|x_i-x_j\|^2}{\alpha}\right)\right)$ denotes the intra-class discriminating weight and inter-class discriminating weight, respectively. In addition, the parameter α is used as a regulator, which controls the overall scale or the smoothing of the space. But the selection of the parameter α remains an open problem. From Eq. (2.4), we know that if the parameter α is set very low, the value of the $H(i,j)$ will be near zero for all but the closest points.

2.2. Our Method (2DSDP). Considering there is a set of M sample images $A=[A_1,A_2,\cdots,A_M]$, each image is a $m\times n$ matrix, let ω be an n-dimensional unitary column vector. The 2DSDP method is projecting each image $(A_i)_{m\times n}$ onto ω by the following transformation:

$$Y_i=A_i\,\omega,\quad i=1,2\cdots M \tag{2.5}$$

where Y_i is n- dimensional projection feature vector. To obtain the highest recognition rate, it is important to select the optimal projection vector ω. The objective function of the two dimensional SDP method is defined as:

$$S_N\omega=\lambda S_L\omega \tag{2.6}$$

where

$$S_L=\frac{1}{2}\frac{1}{MM}\sum_{i=1}^{M}\sum_{j=1}^{M}W_{ij}(A_i-A_j)(A_i-A_j)^T \tag{2.7}$$

$$S_N=\frac{1}{2}\frac{1}{MM}\sum_{i=1}^{M}\sum_{j=1}^{M}(1-W_{ij})(A_i-A_j)(A_i-A_j)^T \tag{2.8}$$

where W is the weight matrix, whose elements denote the relationship between the image A_i and A_j in the observation space and is defined as:

$$W(i,j)=\begin{cases}\exp\left(-\frac{\|A_i-A_j\|^2}{\alpha}\right)\left(1+\exp\left(-\frac{\|A_i-A_j\|^2}{\alpha}\right)\right), \\ \text{if } A_i \text{ is among } k \text{ nearest neighbors of } A_j, \\ \text{or } A_j \text{ is among } k \text{ nearest neighbors of } A_i \text{ and } \tau_i=\tau_j. \\ \exp\left(-\frac{\|A_i-A_j\|^2}{\alpha}\right)\left(1-\exp\left(-\frac{\|A_i-A_j\|^2}{\alpha}\right)\right), \\ \text{if } A_i \text{ is among } k \text{ nearest neighbors of } A_j, \\ \text{or } A_j \text{ is among } k \text{ nearest neighbors of } A_i \text{ and } \tau_i\neq\tau_j. \\ 0, \ otherwise\end{cases} \quad (2.9)$$

2.3. Feature Extraction and Classification. Step1. Calculating the generalized eigenvectors $\varphi^* = \{\varphi_1, \varphi_2, \cdots \varphi_d\}$ of Eq. (2.6) corresponding to the d largest positive eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$. Step2. Mapping the given sample A_i using the projection matrix $\varphi^* : Y_i = (\varphi^*)^T A_i$, the feature Y_i is used to represent the sample A_i for recognition purposes.

3 Experiments on Face Databases

3.1. Experiments Using the FERET Face Database. The proposed algorithm is tested on a subset of the FERET database. This subset includes 1,400 images of 200 individuals (each individual has seven images). This subset involves variations in facial expression, illumination, and pose. It is composed of the images whose names are marked with two-character strings: “ba”, “bd”, “be”, “bf”, “bg”, “bj”, “bk”. In our experiment, the facial portion of each original image was automatically cropped based on the location of eyes and the cropped image was resized to 40*40 pixels. Some example images of one person are shown in Figure 3.1.

Figure 3.1: Seven images of one person in the subset of FERET face database.

In our test, we use the first, the sixth and the seventh images (i.e., “ba”, “bj” and “bk”) per class for training, and the remaining four images (i.e., “bd” , “be”, “bf” and “bg”) for testing.

To evaluate the performance of the proposed method (2DSDP), we conducted the first experiment on FERET face databases. The number of selected eigenvectors (projection vectors) varies from 1 to 8. Here, let m denotes the projection vector number, then the dimension of corresponding projected feature vector is 40*m . And the kernel parameter α in 2DSDP is chosen as α =2. Finally, a nearest neighbor classifier with Cosine distance is employed to classify in the projected feature space. The recognition rates versus m are shown in Table 1.

Table1 The recognition rates (RR) of 2DSDP versus the dimensions (40*m) when the nearest neighbor classifier with Cosine distance is used on the FERET database

m	1	2	3	4
RR(%)	49.25	45	42.75	37.87
m	5	6	7	8
RR(%)	36.13	34.88	33.12	30.88

To evaluate the effect of the different kernel parameter α , we conducted the second experiment on the FERET database. In this test, the SDP and 2DSDP methods are used for feacture extraction. Finally, a nearest neighbor classifier with Cosine distance is employed to classify in the projected feature space. The recognition rates with different parameter α are shown in Table 2.

Table 2 The recognition rates (%) and corresponding dimension (shown in parentheses) of SDP and 2DSDP versus the parameter α when the nearest neighbor classifier with Cosine distance is used on the FERET database, 3 samples (the first, the sixth, and the seventh) per class are used for training

α	2	5	10
SDP	40.5(40)	46.88(35)	46.38(30)
2DSDP	49.25(40*1)	49.75(40*1)	48.38(40*1)
α	15	20	25
SDP	45.12(20)	45(35)	44.37(20)
2DSDP	48.5(40*1)	48.5(40*1)	48.5(40*1)

From Table 2, we can see that the 2DSDP method outperforms the SDP method significantly. It shows that the proposed method (2DSDP) can extract more discriminative features than the SDP method. this mainly because the proposed one not only takes into account the local information of the data as SDP method does, but also preserves the correlations between variations of rows and those of columns of face images.

3.2. Experiments Using the Yale Face Database. The Yale database contains 165 grayscale images of 15 individuals, each individual has 11 images with

46 * 56 pixels under various facial expressions and lighting conditions (one per different facial expression or configuration: center-light, w/glasses, happy, left-light, w/no glasses, normal, right-light, sad, sleepy, surprised, and wink.). Figure 2 shows some sample images of one people in the Yale face database.

Figure 3.2: Eleven images of one person in the Yale face database

In the experiment, the UDP, SDP, 2DUDP and 2DSDP methods are used for feacture extraction. First, we used the first sixth face images per class for training, and the rest of the five face images for testing. In this experiment, the parameter α was set as 25. Finally, a nearest neighbor classifier with Cosine distance is employed to classify in the projected feature space. The recognition rates and corresponding dimensions are shown in Table 3.

Table 3 The recognition rates (RR) and corresponding dimension of UDP, SDP,2DUDP and 2DSDP methods when the nearest neighbor classifier with Cosine distance is used on the Yale database, the first sixth samples per class are used for training

Methods	UDP	SDP	2DUDP	2DSDP
RR(%)	96	96	93.33	97.33
Dimension	25	20	56*3	56*5

Then, we reevaluate the recognition performance of the proposed method 2DSDP. We repeated the recognition procedure 10 times by randomly choosing six training sets per class. Here, the UDP, SDP, 2DUDP and 2DSDP methods are used for feature extraction. Finally, the nearest neighbor classifier with Euclidean distance is employed for classification. The recognition rates with corresponding standard deviations and corresponding dimensions are shown in Table 4.

Table 4 The average recognition rates (ARR, %) with corresponding standard deviations (Std, %) of four feature extraction methods (UDP, SDP, 2DUDP and 2DSDP) across 10 runs on the Yale database under the nearest neighbor classifier with Cosine distance

Methods	UDP	SDP	2DUDP	2DSDP
ARR(%)	93.87	96.4	96.53	97.6
Std(%)	2.28	2.88	1.43	1.76
Dimension	35	35	56*3	56*5

From Table 3 and Table 4, we can see that the results are consistent with those drawn from the experiments conducted on the FERET face database. The conclusions are, on the whole, consistent with those drawn from the first experiment conducted on the FERET face database. It further shows that the proposed method can extract more discriminative features than the other methods.

4. Conclusions

In this paper, we proposed a novel method, called 2D locally discriminating projection (2DLDP), for face recognition. This method is developed based on the successful application of 2DPCA (two-dimensional principal component analysis) and SDP (supervised discriminant projection) to face recognition. The proposed one not only takes into account both the local information of the data and the class information of the data to model the manifold structure, but also preserves the useful information of the image data. Experimental results show that the 2DSDP method achieves higher recognition than the SDP method and other feature extraction methods.

Acknowledgements

This work was supported by the Research Program of Hebei Education Department under Grant No. Z2010174 and the Research Program of Hebei Municipal Science & Technology Department under Grant No. 10213551.

References

[1] W. Zhao, R. Chellappa, A. Rosenfeld, et.al, Face recognition: A literature survey, ACM Comput. Surv. 35 (2003), 399–458.

[2] M. A. Grudin, On internal representations in face recognition systems, PR. 33 (2000), 1161-1177.

[3] M. Turk and A. Pentland, Eigenfaces for recognition, J. Cogn. Neurosci. 3 (1991), 71-86.

[4] D. L. Swets and J.Weng, Using discriminant eigenfeatures for image retrieval, IEEE Trans. PAMI. 18 (1996) , 831-836.

[5] P. N. Belhumeur, J. P. Hespanha, D. J. Kriengman, Eigenfaces versus Fisherfaces: recognition using class specific linear projection, IEEE Trans. Patt. Anal. Mach. Intell. 19 (1997), 711-720.

[6] J. Yang, D. Zhang, J. y. Yang, et.al, Globally Maximizing, Locally Minimizing: Unsupervised Discriminant Projection with Applications to Face and Palm Biometrics, IEEE Trans. Pattern Anal. Machine

Intell. 29 (2007), 650- 664.

[7] J.G. Wang, J.Z. Hua, C.Y. Zhou, Supervised Discriminant Projection for Feature Extraction, ICCAE 2011. 1 (2011), 200-203.

[8] L. F. Chen, H.Y.M. Liao, M.T.Ko, et al, A new LDA-based face recognition system which can solve the small sample size problem, Pattern Recognition Lett. 33 (2000), 1713–1726.

[9] H. Yu and J. Yang, A direct LDA algorithm for high-dimensional data with application to face recognition, Pattern Recognition Lett. 34 (2001), 2067–2070.

[10] J. Yang, D. Zhang, A. F. Frangi, et al, Two-dimensional PCA: a new approach to appearance- based face representation and recognition, IEEE Trans. PAMI. 26 (2004), 131-137.

[11] M. Li and B. Yuan, 2D-LDA: a statistical linear discriminant analysis for image matrix, PRL. 26 (2005), 527–532.

Jianguo Wang received the MS degree in signal and information processing at Xi'an University of technology in 1997, and the PhD degree in computer application from the Nanjing University of Science and Technology, the Department of Computer Science in 2008. He is currently an professor in the Department of Computer Science & Technology, Tangshan College. His research interests are in the areas of pattern recognition, image processing, face detection and recognition.