Two-Dimensional Nearest Neighbor Discriminant Analysis ⋆

Xipeng Qiu * , Lide Wu

Media Computing & Web Intelligence Lab

Department of Computer Science and Engineering, Fudan University

Abstract

Recently, some feature extraction methods have been developed by representing images with matrix directly, however few of them are proposed to improve accuracy of classification directly. In this paper, a novel feature extraction method, two-dimensional nearest neighbor discriminant analysis(2DNNDA), is proposed from the view of the nearest neighbor classification, which makes use of the matrix representation of images. We apply 2DNNDA to face recognition and the results demonstrate that 2DNNDA outperforms the conventional methods.

Key words: Two-Dimensional Nearest Neighbor Discriminant Analysis, Linear Discriminant Analysis, Face Recognition

1 Introduction

When we apply the traditional pattern recognition technologies to deal with image classification problems, such as face recognition, the raw image data often have very high dimensionality and the limited number of samples. So we have to face with the "curse of dimensionality" problem. Feature extraction [1] is an effective method to map data into a low-dimensional space. Principal

⋆ The research described in this paper has been supported by NSF of China (60435020).

* Corresponding author. TEL:+86-21-55664405.
Email address: xpqiu@fudan.edu.cn (Xipeng Qiu).

component analysis (PCA) [2] and linear discriminant analysis (LDA) [3] are two popular linear feature extraction.
However, these methods are based on vector data. When dealing with image data, we must represent them in vectors, which usually leads to a high dimensional vector space. Thus, some methods, such as LDA, have the small sample size problem because of the relatively small number of training samples [3].
Recently, some tensor based methods [4–6] have been developed for dealing with image data by representing image with matrix directly. They consider image as two-dimensional signal $X \in R^{m \times n}$ and aim to find two transformation matrices $L \in R^{m \times m'}$ and $R \in R^{n \times n'}$ that map each X to a low rank matrix $Y \in R^{m' \times n'}$ by $Y = L^T X R$. An advantage of these methods is computational efficiency since the dimensionality is reduced in two ways. Another advantage is that these methods make use of the image space information, which could be lost when concatenating image matrices to vectors.

Some methods, such as two-dimensional principal component analysis(2DPCA) [4] and generalized low rank approximations of matrices (GLRAM) [6], are proposed to reduce the reconstruction error without considering the classification. Other methods, such as two-dimensional linear discriminative analysis(2DLDA)[5], aim to minimize the within-class distances and maximize between-class distances, which, like LDA, fail when each class does not belong to single Gaussian distribution or their centers overlap.

In this paper, a new feature extraction method, two-dimensional nearest neighbor discriminant analysis (2DNNDA), is proposed from the point of view of nearest neighbor classification (NN). 2DNNDA can be regarded as an two-dimensional extension of nearest neighbor discriminant analysis [7] with matrix based image representation.

The rest of the paper is organized as follows: We describe two-dimensional nearest neighbor discriminant analysis in Section 2. Experimental evaluations of our method are presented in Section 3. Finally, we give the conclusions in Section 4.

2 Two-Dimensional Nearest Neighbor Discriminant Analysis

Considering $X_i \in R^{m \times n}, (i = 1, \cdots, N)$ be the N images in the dataset, clustered into classes $\omega_i (i = 1, \ldots, c)$.

A natural similarity metric between matrices X_i and X_j is $||X_i - X_j||_F$, the Frobenius norm of their difference.

We define the extra-class nearest neighbor of a sample $X_j \in \omega_i, X_j \in R^{m\times n}$ as

$$X_j^E = \arg\min_Z ||Z - X_j||_F, \forall Z \notin \omega_i. \tag{1}$$

The intra-class nearest neighbor of the sample $X_j \in \omega_i$ is defined as

$$X_j^I = \arg\min_Z ||Z - X_j||_F, \forall Z \in \omega_i, Z \neq X_j. \tag{2}$$

Then the nonparametric extra-class and intra-class differences are defined as

$$\Delta_j^E = X_j - X_j^E, \tag{3}$$
$$\Delta_j^I = X_j - X_j^I. \tag{4}$$

From Eq.(3) and (4), we can see that $||\Delta_j^E||_F$ represents the distance between the sample X_j and its nearest neighbor in the different classes, and $||\Delta_j^I||_F$ represents the distance between the sample X_j and its nearest neighbor in the same class. Given a training sample X_j, the accuracy of the nearest neighbor classification can be directly computed as the difference

$$\Theta_j = ||\Delta_j^E||_F^2 - ||\Delta_j^I||_F^2, \tag{5}$$

where Δ_j^E and Δ_j^I are nonparametric extra-class and intra-class differences and defined in Eq.(3) and (4).

If the difference Θ_j is positive, X_j will be correctly classified. Otherwise, X_j will be classified wrongly. The larger the difference Θ_j is, the more accurately the sample X_j is classified.

Assuming that we extract features by the two projection matrices $L \in R^{m\times m'}$ and $R \in R^{n\times n'}$, the projected sample $X_j^{new} = L^T X_j R$. The projected nonparametric extra-class and intra-class differences can be written as $\delta_j^E = L^T \Delta_j^E R$ and $\delta_j^I = L^T \Delta_j^I R$. So we expect to find the optimal L and R to make the difference $||\delta_j^E||_F^2 - ||\delta_j^I||_F^2$ as large as possible in the projected subspace.

$$(\widehat{L}, \widehat{R}) = \arg\max_{L,R} \sum_{j=1}^{N} (||\delta_j^E||_F^2 - ||\delta_j^I||_F^2). \tag{6}$$

Considering that,

$$
\begin{aligned}
&\sum_{j=1}^{N}(||\delta_j^E||_F^2 - ||\delta_j^I||_F^2) \\
=&\sum_{j=1}^{N}(||L^T\Delta_j^E R||_F^2 - ||L^T\Delta_j^I R||_F^2) \\
=&\, tr(\sum_{j=1}^{N}(L^T\Delta_j^E R)(L^T\Delta_j^E R)^T) - tr(\sum_{j=1}^{N}(L^T\Delta_j^I R)(L^T\Delta_j^I R)^T) \\
=&\, tr\Big(L^T(\sum_{j=1}^{N}\Delta_j^E RR^T(\Delta_j^E)^T)L\Big) - tr\Big(L^T(\sum_{j=1}^{N}\Delta_j^I RR^T(\Delta_j^I)^T)L\Big) \qquad (7) \\
=&\, tr\Big(R^T(\sum_{j=1}^{N}(\Delta_j^E)^T LL^T\Delta_j^E)R\Big) - tr\Big(R^T(\sum_{j=1}^{N}(\Delta_j^I)^T LL^T\Delta_j^I)R\Big) \qquad (8)
\end{aligned}
$$

where $tr(\cdot)$ is the trace of matrix.

Due to the difficulty of computing the optimal L and R simultaneously, we derive an iterative algorithm. More specifically, for a fixed R, we can compute the optimal L by solving an optimization problem similar to the one in NNDA [7]. With the computed L, we can then update R by solving another optimization problem.

2.1 Computation of L

For a fix R, the Eq. (6) is equivalent to

$$\widehat{L} = \arg\max_{L} tr(L^T(\hat{S}_b^R - \hat{S}_w^R)L), \ \texttt{subject to} \ L^T L = I, \qquad (9)$$

where $\hat{S}_b^R$ and $\hat{S}_w^R$ the nonparametric between-class and within-class scatter matrix are defined as: $\hat{S}_b^R = \sum_{j=1}^{N}(\Delta_j^E)RR^T(\Delta_j^E)^T$ and $\hat{S}_w^R = \sum_{j=1}^{N}(\Delta_j^I)RR^T(\Delta_j^I)^T$. The projection matrix $\widehat{L}$ must be constituted by the m' eigenvectors of $(\hat{S}_b^R - \hat{S}_w^R)$ corresponding to its m' largest eigenvalues.

2.2 Computation of R

Similarly, for a fix L, the Eq. (6) is equivalent to

$$\widehat{R} = \arg\max_{R} tr(R^T(\hat{S}_b^L - \hat{S}_w^L)R), \ \texttt{subject to} \ R^T R = I, \qquad (10)$$

where $\hat{S}_b^L$ and $\hat{S}_w^L$ the nonparametric between-class and within-class scatter matrix are defined as: $\hat{S}_b^L = \sum_{j=1}^{N}(\Delta_j^E)LL^T(\Delta_j^E)^T$ and $\hat{S}_w^L = \sum_{j=1}^{N}(\Delta_j^I)LL^T(\Delta_j^I)^T$. The projection matrix $\widehat{R}$ must be constituted by the n' eigenvectors of $(\hat{S}_b^L - \hat{S}_w^L)$ corresponding to its n' largest eigenvalues.

Fig. 1 gives the algorithmic flowchart of 2DNNDA.

- Given samples (X_j, c_n), where $X_j \in R^{m\times n}$, $n \in \{1, \cdots, N\}$ and $c_n \in \{\omega_1, \cdots, \omega_C\}$ is label;
- Give the number of iterations T;
- $R \leftarrow I_{n\times n}$;
- For $t = 1, \cdots, T$,

(1) Calculate $\hat{S}_b^R$ and $\hat{S}_w^R$;
(2) Compute the first m' eigenvectors $\{\phi_k^L\}_{k=1}^{m'}$ of $(\hat{S}_b^R - \hat{S}_w^R)$;
(3) Update $L \leftarrow [\phi_1^L, \cdots, \phi_{m'}^L]$;
(4) Calculate $\hat{S}_b^L$ and $\hat{S}_w^L$;
(5) Compute the first n' eigenvectors $\{\phi_k^R\}_{k=1}^{n'}$ of $(\hat{S}_b^L - \hat{S}_w^L)$;
(6) Update $R \leftarrow [\phi_1^R, \cdots, \phi_{n'}^R]$.

- Return L and R.

Fig. 1. Algorithm of 2DNNDA

3 Experiments

In this section, we apply 2DNNDA to face recognition and compare it with the existing classical feature extraction methods, such as PCA[2], 2DPCA[4], LDA[3], 2DLDA[5] and NNDA[7]. We perform the experiments on the famous ATT [8] and UMIST [9] face databases. All experiments are repeated five times independently, and the average results are given.

The ATT database contains 400 images (112 × 92) of 40 persons, 10 images per person. The set of the 10 images for each person is randomly partitioned into a training subset of 5 images and a test set of the other 5.

The UMIST database consists of 564 images (112 × 92) of 20 people. Each covering a range of poses from profile to frontal views. Subjects cover a range of race/sex/appearance. We select 5 images randomly as training set for each person, and the rest images are test set.

Since that PCA and 2DPCA aim to minimize reconstruction error and not to classification error, their results are just regarded as the baselines.

LDA and 2DLDA aim to minimize the average intra-class distance and to maximize the average extra-class distance, so they are effective just when each class has a single Gaussian distribution and their class centers do not overlap. When the real face data cannot meet these conditions, LDA and 2DLDA could result in poor performances.

2DNNDA is proposed from the viewpoint of nearest neighbor classification(NN), so it can improve the performance of NN directly.

Table 1 gives the recognition accuracies of the above methods on the ATT and UMIST face databases, which shows that 2DNNDA, like our analysis, gives the best results on all the two databases.

Table 1

Recognition accuracies on ATT and UMIST face databases

		PCA	LDA	NNDA	2DPCA	2DLDA	2DNNDA
ATT	Dim	50	39	50	112×2	10×10	10×10
	Accuracy	0.94	0.93	0.96	0.94	0.95	0.98
UMIST	Dim	50	19	50	112×2	10×10	10×10
	Accuracy	0.82	0.87	0.88	0.85	0.74	0.94

4 Conclusions

In this paper, we proposed a new feature extraction method, two-dimensional nearest neighbor discriminant analysis(2DNNDA), which extracts feature to improve the performance of nearest neighbor classification. Our experimental results demonstrate that 2DNNDA is a very efficient method. In the further works, we will apply 2DNNDA to other image recognition problems.

References

[1] K. Fukunaga, Introduction to statistical pattern recognition, 2nd Edition, Academic Press, Boston, 1990.

[2] M. Turk, A. Pentland, Eigenfaces for recognition, J. Cognitive Neuroscience 3 (1) (1991) 71–86.

[3] P. Belhumeur, J. Hespanda, D. Kiregeman, Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection, IEEE Trans. Pattern Anal. Machine Intell. 19 (7) (1997) 711–720.

[4] J. Yang, D. Zhang, A. F. Frangi, J. yu Yang, Two-dimensional PCA: A new approach to appearance-based face representation and recognition, IEEE Trans. Pattern Anal. Machine Intell. 26 (1) (2004) 131–137.

[5] J. Ye, R. Janardan, Q. Li, Two-dimensional linear discriminant analysis, Vol. 17, 2004.

[6] J. Ye, Generalized low rank approximations of matrices, Mach. Learn. 61 (1) (2005) 167–191.

[7] X. Qiu, L. Wu, Stepwise nearest neighbor discriminant analysis, in: Proc. of Int. Joint Conf. on Artif. Intell. (IJCAI), Edinburgh, 2005, pp. 829–834.

[8] F. Samaria, A. Harter, Parameterisation of a stochastic model for human face identification, in: Proc. of IEEE Workshop on Appl. of Comput. Vision, 1994.

[9] D. B. Graham, N. M. Allinson, Characterizing virtual eigensignatures for general purpose face recognition, (in) Face Recognition: From Theory to Applications, NATO ASI Series F, Computer and Systems Sciences 163 (1998) 446–456.