Journal of Computational Information Systems 7:4 (2011) 1100-1107
Available at http://www.Jofcis.com

Dual-space Neighborhood Discriminant Embedding for Face Recognition

Zhonghua LIU [1,2,†], Yuan ZHOU, Zhong JIN[1]

[1]*School of Computer Science & Technology, Nanjing University of Science and Technology, Nanjing, China*

[2]*Electronic Information Engineering College, Henan University of Science and Technology, Luoyang, China*

Abstract

In this paper, a novel subspace learning method called neighborhood discriminant embedding (NDE) is proposed for pattern classification. In our algorithm, the neighbor and class relations of training samples data are used to construct the low-dimensional embedding submanifold. After being embedding into a low-dimensional subspace, in a local structure, samples from the same class will be as close as possible, and those from different classes are separated far way. However, NDE will suffer from the small sample size problem when dealing with the high dimensional face data. In order to solve the small sample size problem and take fully advantage of the discriminative information in the face space, we further propose a novel algorithm called dual-space neighborhood discriminant embedding (DSNDE). Since Gabor wavelet representation of face images is robust to variations due to illumination and facial expression changes, we apply the proposed DSNDE (NDE) algorithm on Gabor features for face recognition. Experiments on the face databases demonstrate the effectiveness of our method.

Keywords: Locally Linear Embedding; Neighborhood Discriminant Embedding; Dual-space Neighborhood Discriminant Embedding; Face Recognition.

1. Introduction

In many real world applications such as pattern recognition and information retrieval, the data is often of very high dimensionality. Due to the curse of dimensionality, Procedures that are analytically or computationally manageable in low-dimensional spaces can become completely impractical in a space of several hundreds or thousands dimensions [1]. Dimension reduction is such a technique that attempts to overcome the curse of the dimensionality and to extract relevant features. Classical linear dimensionality reduction methods, such as Principal Component Analysis (PCA) [2] and Linear Discriminant Analysis (LDA) [3], make use of some well-defined statistical measures to evaluate the usefulness of features. Since these measures are usually defined based on the overall properties of the data set, such statistical approaches are often powerful to retain the global structures of the data space. However, the methods usually perform poorly when their underlying assumptions are violated [4].

Compared with linear dimensionality reduction methods, nonlinear dimensionality reduction techniques yield better results if the data samples reside on a manifold. The manifold-embedding-based nonlinear approaches, such as locally linear embedding (LLE) [5], isometric feature mapping (ISOMAP) [6], utilize

[†] Corresponding author.
Email addresses: lzhua@yahoo.cn (Zhonghua LIU).

local neighborhood relation to learn the global structure of nonlinear manifolds. Though these methods can learn the intrinsic manifold embedded in the ambient space of high dimensionality, and proved to be effective, they cannot deal with the out-of-sample problem [7]. In order to overcome the disadvantages, many methods, such as locality preserving projections (LPP) [8], neighborhood preserving projections (NPP) [9], etc., which are linear forms of the original nonlinear manifold methods, are proposed to address the problem. However, the linear forms of the manifold methods are unsupervised dimensionality reduction method and neglect the overall properties of the data that are fundamental to the success of extracting predictive features.

Therefore, we propose a new supervised linear dimensionality reduction method called neighborhood discriminant embedding (NDE). NDE effectively combines local geometric information underlying the data and class label information. That is, NDE maximizes between-class separability and preserves within-class local structure at the same time. NDE can extract features to preserve complex classification structures effectively. In the implementation of NDE, to overcome the singularity of the within-class metric matrix in face recognition, inspired by the dual-space LDA in [10], [11], we further develop a new method called dual-space discriminant neighborhood embedding (DSNDE). DSNDE takes full advantage of the discriminative information in the principal and null subspaces of the within-class metric matrix to fulfill classification recognition. Since the Gabor wavelet representation of face images is robust to variations due to illumination and facial expression changes [12], we apply the proposed DSNDE method on Gabor features for face recognition. We have conducted extensive experiments on the face databases to evaluate the effectiveness of our proposed method and compared it with other subspace methods.

2. Neighborhood Discriminant Embedding (NDE)

Though the distribution of one face under variations such as viewpoint, pose, and illumination is highly nonlinear and nonconvex [13], it is generally believed that the face space is a submanifold embedded in the ambient image space. The manifold learning approaches can learn the intrinsic manifold embedded in the ambient space of high dimensionality. Therefore, we adopt linearization extension of locally linear embedding (LLE) to model the within-class metric matrix. For the modeling of complex intrinsic difference variations, we consider the extra-class samples of k nearest neighbors of each sample. The between-class metric matrix is obtained by computing the difference between each sample and extra-class samples of its k nearest neighbors.

2.1. Local Supervised Linear Dimension Reduction

Locally linear embedding (LLE) [5] is a nonlinear learning technique that looks for an embedding that preserves the local geometry in the neighborhood of each data point. In the LLE algorithm, geometric relationship among the high dimensional data is described by the linear nearest-neighbor reconstruction. It means that each data point is represented by the linear combination of its k nearest neighbors. Given a set of points $x_1, x_2, \cdots, x_m$ in R^N, we characterize the local geometry relation by linear coefficients $W_{ij}^{(c)}$ that reconstruct each data point x_i from its k neighbors x_j. Choose $W_{ij}^{(c)}$ to minimize a cost function of squared reconstruction errors:

$$\min \sum_i \left\| x_i - \sum_{j=1}^{k} W_{ij}^{(c)} x_j \right\|^2 \tag{1}$$

Where $W_{ij}^{(c)}=0$ if x_i and x_j belong to different classes and the rows of the weight matrix sum to one: $\sum_{j=1}^{k} W_{ij}^{(c)} = 1$. The minimization in (1) is different from that in the original unsupervised LLE. Here each sample is reconstructed only from other samples in its class.

In the embedding stage, a low-dimensional data set $\{y_1, y_2, \cdots, y_n\}$, where $y_i \in R^d$ ($d < N$), y_i corresponds to x_i and y_i is constructed with the same weights $W^{(c)}$. Suppose the transformation is linear, that is, $y^T=P^T x$, where the i-th row vector of y is y_i, the i-th column vector of x is x_i. We have the embedding cost function as

$$\min E(P) = \min \sum_i \left\| y_i - \sum_{j=1}^{k} W_{ij}^{(c)} y_j \right\|^2 = \min tr(P^T M_s P) \tag{2}$$

Where $M_s = x(I-W^{(c)})^T(I-W^{(c)})x^T$.

2.2. Maximum Local Inter-class Distance

Suppose that c_i denotes the class label of x_i and knn(i) denotes the set of k nearest-neighbors of x_i , then the adjacent matrix W of graph G which models the underlying supervised manifold structure is as follows:

$$W_{ij}^{(r)} = \begin{cases} 1 & (x_i \in knn(j) \vee x_j \in knn(i)) \wedge (c_i \neq c_j) \\ 0 & otherwise \end{cases} \tag{3}$$

In order to make inter-class separability, we define a criterion function as

$$\delta(P) = \sum_{i,j} \left\| y_i - y_j \right\|^2 = \sum_{i,j} \left\| P^T x_i - P^T x_j \right\|^2 \tag{4}$$

where $(x_i \in knn(j) \vee x_j \in knn(i)) \wedge (c_i \neq c_j)$, P is a linear transformation matrix, and $y_i^T = P^T x_i$.

Maximizing (4) is an attempt to maximize the distance between different classes. Using (3), the criterion (4) can also be expressed as

$$\delta(P) = \sum_{i,j} \left\| y_i - y_j \right\|^2 W_{ij}^{(r)} = \sum_{i,j} \left\| P^T x_i - P^T x_j \right\|^2 W_{ij}^{(r)} = 2\, tr(P^T M_b P) \tag{5}$$

where $x=[x_1,x_2,\cdots x_m]$, and D is a diagonal matrix; its entries are column (or row, since $W^{(r)}$ is symmetric) sum of $W^{(r)}$, $D_{ii} = \sum_j W_{ij}^{(r)}$, $M_b = x(D - W^{(r)})x^T$.

2.3. NDE

From the above analysis, a novel subspace learning method called neighborhood discriminant embedding (NDE) is proposed. The criterion function of NDE is as follows:

$$P = \arg\max \frac{tr(P^T M_b P)}{tr(P^T M_s P)} \tag{6}$$

NDE method make data points in the same class form compact submanifold whereas the gaps between the submanifolds corresponding to different classes become wider than before. At the same time, NDE method overcomes the deficiency of traditional fisher discriminate analysis method (such as multimodal data or non-Gaussian distribution).

The linear transformation matrix P can be acquired by using generalized eigenvalue decomposition. That is, P can be computed from the eigenvectors of $M_s^{-1}M_b$. However, for face recognition, the number of images in the training set is usually much smaller than the dimension of data, NDE can also suffer from the small sample size problem. In these cases, the within-class scatter matrix is singular and it is difficult to compute M_s^{-1}.

3. Dual-space Neighborhood Discriminant Embedding (DSNDE)

3.1. DSNDE

When the number of samples of each class is small, directly solving eigenvectors of $M_s^{-1}M_b$ is infeasible. In order to solve the small sample size problem and avoid the singularity of M_s, several approaches, such as two-stage PCA+LDA method [2], null space method [14] and Direct LDA [15], etc., have been proposed to address this problem. These methods only make use of the discriminative information in the null space or the principal space. However, it has been proved that both the principal space and the null space contain the discriminative information. In order to fully utilize the discriminative information contained in the principal space and the null space of the intra class scatter matrix, inspired by the dual-space LDA in [10], we further propose a new method called dual-space neighborhood discriminant embedding (DSNDE) to simultaneously make use of the principal and null subspace of within-class scatter matrix. This helps stabilize the transformation and thus improves the generalization performance.

The detailed algorithm of the DSNDE is given as follows:

At training stage:

1. Compute M_s and M_b from the training set.
2. Apply PCA to M_s, and calculate the principal space F spanned by the eigenvectors V of M_s with nonzero eigenvalues $\{\lambda_i\}_{i=1}^K$.
3. M_b is transformed to

$$K_b^P = \Lambda^{-1/2} V^T M_b V \Lambda^{-1/2} \tag{7}$$

Where Λ is the eigenvalue matrix of M_s.

4. Apply PCA to K_b^P, compute l_p eigenvectors Ψ_P with the largest eigenvalues. The l_P discriminative vectors in F are defined as

$$W_P = V\Lambda^{-1/2}\Psi_P \tag{8}$$

5. In the null space $\overline{F}$ spanned by the eigenvectors of M_s with zero eigenvalues $\{\lambda_i\}_{i=K+1}^{N}$, M_b is transformed to

$$K_b^C = (I - VV^T)M_b(I - VV^T) \tag{9}$$

Compute l_C eigenvectors Ψ_C of K_b^C with the largest eigenvalues. The l_C discriminative vectors in the $\overline{F}$ are defined as

$$W_C = (I - VV^T)\Psi_C \tag{10}$$

At the recognition stage,

1. All the face images $\{x_i\}$ in the gallery and the probe image y are respectively projected to the principal space F and the null space $\overline{F}$.

$$\alpha_i^P = W_P^T x_i\,,\quad \alpha_i^C = W_c^T x_i\,, \alpha^P = W_P^T y\,,\quad \alpha^C = W_C^T y \tag{11}$$

2. Class is found to minimize the distance measure

$$d_i = \left\|\alpha_i^P - \alpha^P\right\|^2 + \left\|\alpha_i^C - \alpha^C\right\|^2 / \rho \tag{12}$$

Where $\rho = (1/(N-K))\sum_{i=K+1}^{N}\lambda_i^*$, the unknown λ_i^* in $\overline{F}$ are estimated by fitting a nonlinear function to the available portion of the eigenvalues spectrum in F.

3.2. Estimation of the Parameter ρ in DSNDE Algorithm

The ρ is the mean value of all the eigenvalues in $\overline{F}$. However, we do not compute all the eigenvalues in practice due to the high dimensionality, and we can only compute the eigenvalues λ_i $(i=1, \cdots, K)$ in F. The eigenvalues $\lambda_i^*(i = K+1,\cdots,N)$ in $\overline{F}$ can be computed by extrapolating the eigenspace spectrum. Figure 1 shows the fitting function of the eigenvalues in F on ORL database.

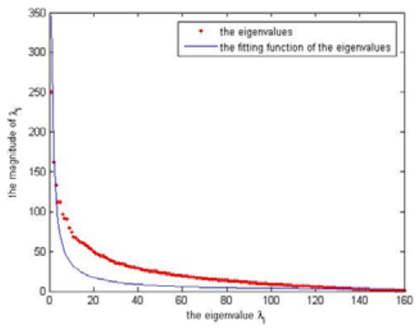

Fig. 1 The Fitting Image of the Eigen Values in F .

4. Experimental Results

In this section, we implement experiments on ORL, AR, CMU PIE and FERET face databases to evaluate the proposed algorithm. For comparison purpose, LDA [3], LFDA [16], and LPP [8] are used to feature extraction, and we compare the classification ability of our DSNDE algorithm with these algorithms after feature extraction. In addition, the concrete steps of extracting Gabor transform features are as follows: 40 Gabor wavelet images are firstly obtained for each face image. Then we downsample each Gabor wavelet image by a factor to reduce the space dimension. Finally, we concatenate all these Gabor wavelet images and derive an augmented feature vector for face recognition.

4.1. Experiments on ORL Face Database

The ORL face database is used in this test. It contains 400 images of 40 individuals. The images were captured at different times and have different variations including expressions and facial details. The size of each face image is 56×46 pixels.

In this experiment, we do not make Gabor transform for face images, and the proposed DSNDE is tested directly on the original face images. We choose the first p different images per individual to form the training set, and the rest of the database is used for testing. Figure 4 illustrates the recognition rates of each method versus the variation of training sample size.

It can be seen from Figure 4 that the recognition rate of DSNDE is higher than other methods irrespective of the variation of training sample size.

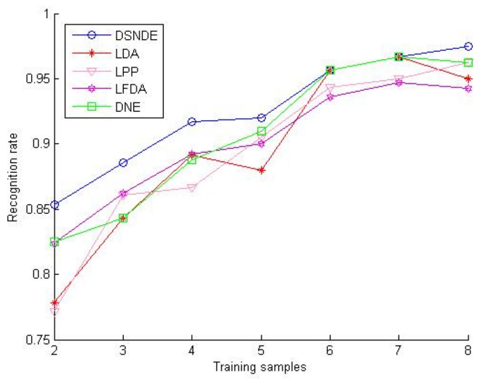

Fig. 4 Recognition Rate vs. Training Sample Size of DSNDE, LDA, PCA, LFDA, DNE on ORL Database.

4.2. Experiments on AR Face Database

The AR face database contains over 4000 face images of 126 people, including the frontal view of faces with different facial expressions, lighting conditions, and occlusions. The images of 120 individuals were taken in two sessions and each section contains 13 color images. Here, we select the training data and the testing data from the face images of these 120 persons. For the training data, we select 120×7 non included face images of 120 persons from the first session. The testing data are composed of 120×7 non included face images from the second session. Each image was cropped to 50×40 pixels.

In this experiment, we firstly make Gabor transform on face images, then we investigate the performance of the DSNDE on Gabor wavelet images. The results of the experiment are reported in Table 1.

Table 1 Recognition Rate Comparison of DSNDE with other Methods on the Gabor Wavelet Images on the AR Database

Methods	Recognition Rate (%)
LDA	74.67
LPP	73.93
LFDA	74.17
DNE	74.88
DSNDE	79.64

It can be known from the Table 1 that the recognition accuracy of the suggested method is much higher than those of the other methods.

4.3. Experiments on CMU PIE Face Database

The CMU PIE face database contains 68 subjects with 41,368 face images as a whole. In this paper, we use a subset that only contains five near frontal poses and all of the images are under different illuminations and expressions. Hence, there are 170 images with 50×50 pixels for each individual. We choose the 43 different images per individual to form the training set. The rest are used for testing.

In this experiment, we test the performance of the methods on Gabor wavelet images. Table 2 gives the results. It can be also seen from Table 2 that the recognition accuracy of the suggested DSNDE is higher than those of the other methods.

Table 2 Recognition Rate Comparison of DSNDE with other Methods on the Gabor Wavelet Images on the PIE Database

Methods	Recognition Rate (%)
LDA	93.56
LPP	94.42
LFDA	94.07
DNE	93.42
DSNDE	95.92

4.4. Experiments on FERET Face Database

The FERET face database includes 1400 images of 200 individuals (each one has 7 images). We choose the former 3 different images per individual to form the training set. The rest are used for testing.

Table 3 Recognition Rate Comparison of DSNDE with other Methods on the Gabor Wavelet Images on the FERET Database

Methods	Recognition Rate (%)
LDA	62.88
LPP	57.88
LFDA	69
DNE	59.25
DSNDE	70.13

In this experiment, we also evaluate the performance of the methods on Gabor wavelet images. The results are reported in Table 3. Once again we observe that DSNDE performs better than the other methods.

5. Conclusion

In this paper, we propose a novel linear dimensionality reduction algorithm called NDE. It takes into account the local neighboring information and the between-class neighboring information for constructing an optimal projection. After being embedding into a low-dimensional subspace, the samples of same class maintain their intrinsic neighbor relations, whereas the neighboring samples of different class are well separated. In order to solve the small sample size problem and fully utilize the discriminant information in both the principal space and the null space of the intraclass scatter matrix, we further propose a novel algorithm called DSNDE. Since Gabor wavelets have been shown to outperform original appearance features, we apply the suggested DSNDE methods on Gabor features for face recognition. Experimental results on several face databases show the superiority of DSNDE (NDE).

Acknowledgement

This project is supported by NSFC of China (Grant Nos. 60973098, 60632050, 60873151).

Reference

[1] Changshui Zhang, Jun Wang et al. Reconstruction and analysis of multi-pose face images based on nonlinear dimensionality reduction. Pattern Recognition, 2004, pp. 325-336.

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

[3] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711−720.

[4] Fukunnaga, K. Introduction to Statistical Pattern Recognition. 2nd ed. Academic Press, London (1991).

[5] Roweis, S., Saul, L.K. Nonlinear Dimensionality Reduction by Locally Linear Embedding. Science 290(22), 2323−2326 (2000).

[6] Tenenbaum, J.B., Silva, V., Langford, J.C. A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science 290 (22), 2319−2323 (2000).

[7] Bengio,Y., Paiement,J., Vincent,P., Dellallaeu,O., Roux,N.L, Quimet,M. Out of sample Extensions for LLE, Isomap, MDS, Eigenmaps, and Spectral Clustering. Neural Information Processing Systems, 2003.

[8] X. He, Yan,S., Hu,Y., Zhang,H. Learning a Locality Preserving Subspace for Visual Recognition. In Proc. IEEE International Conference on Computer Vison, 2003.

[9] Y.W. Pang, L. Zhang, Z.K. Liu, N.H. Yu, H.Q. Li. Neighborhood Preserving Projections (NPP): A Novel Linear Dimension Reduction Method. Proc. international conference on intelligent computing, pp. 117−125, 2005.

[10] X. Wang and X. Tang. Dual-Space Linear Discriminant Analysis for Face Recognition. Proc. IEEE Int’l Conf. Computer Vision and Pattern Recognition, vol. 2, pp. 564–569, Washington, DC, 2004.

[11] Zhifeng Li, Xiaoou Tang. Nonparametric Discriminant Analysis for Face Recognition. IEEE Transactions on Pattern analysis and machine intelligence, 2009, 31(4): 755-761.

[12] Chengjun Liu, Harry Wechsler. Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition. IEEE Transactions on Image Processing, 2002, 11(4): 467-475.

[13] Yan Yan, Yu-Jin Zhang. Discriminant projection embedding for face and palmprint recognition. Neurocomputing, 2008, pp. 3534−3543.

[14] L.Chen, H. Liao, M. Ko, J. Lin, and G. Yu. A New LDA Based Face Recognition System Which can Solve the Small Sample Size Problem. Pattern Recognition, Vol. 33, No. 10, pp. 1713-1726, Oct. 2000.

[15] H.Yu and J.Yang. A direct LDA Algorithm for High-Dimensional data - with Application to Face Recognition. Pattern Recognition, Vol. 34, pp. 2067-2070, 2001.

[16] M. Sugiyama. Local fisher discriminant analysis for supervised dimensionality reduction. Proceedings of 23rd International Conference on Machine Learning, pages 905-912, Pittsburgh, USA,Jun. 25-29 2006b.