Neighborhood Discriminant Projection for Face Recognition

Qubo You, Nanning Zheng, Shaoyi Du, Yang Wu
Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University
Xi'an,Shaanxi Province 710049 P. R. China
qbyou@aiar.xjtu.edu.cn, nnzheng@mail.xjtu.edu.cn, sydu@aiar.xjtu.edu.cn, ywu@aiar.xjtu.edu.cn

Abstract

We propose a novel manifold learning approach, called Neighborhood Discriminant Projection (NDP), for robust face recognition. The purpose of NDP is to preserve the within-class neighboring geometry of the image space, while keeping away the projected vectors of the samples of different classes. For representing the intrinsic within-class neighboring geometry and the similarity of the samples of different classes, the within-class affinity weight and the between-class affinity weight are used to model the within-class submanifold and the between-class submanifold of the samples, respectively. Several experiments on face recognition are conducted to demonstrate the effectiveness and robustness of our proposed method.

1. Introduction

Face recognition has become one of the most challenging problems in computer vision and pattern recognition [11]. Numerous methods have been proposed for face recognition over the past few decades. Among these methods, Principal Component Analysis (PCA) [10] and Linear Discriminant Analysis (LDA) [1] are the most popular techniques, which assume that the samples lie on a linearly embedded manifold.

However, a lot of research has shown that facial images possibly lie on a nonlinear submanifold [6, 9]. When using PCA and LDA for dimensionality reduction, they will fail to discover the intrinsic dimension of the image space. Recently, a number of manifold learning methods are proposed to discover the nonlinear structure of the manifold by investigating the local geometry of the samples, such as LLE [6], Isomap [9] and Laplacian Eigenmap [2]. Because of the difficult issue that how to map a new test sample to the low dimensional space, these algorithms can't be applied to classification problems. Some manifold-based algorithms resolve the difficulty [4, 5, 3]; however, these methods are designed to preserve the locality of the samples in the lower dimensional space rather than good discrimination ability. As a result, the projected vectors of different classes may overlap.

In order to overcome the above shortcoming, we propose a novel manifold learning algorithm, called Neighborhood Discriminant Projection (NDP), which explicitly considers the within-class submanifold and the between-class submanifold by integrating the within-class neighboring information and the between-class neighboring information. The aim of NDP is to preserve the within-class neighboring geometry of the image space, while keeping away the projected vectors of the samples of different classes.

The rest of the paper is organized as follows: the Neighborhood Discriminant Projection (NDP) algorithm is described in Section 2. In Section 3, experimental results are presented to demonstrate the effectiveness and robustness of NDP. Finally, conclusions are summarized in Section 4.

2. Neighborhood Discriminant Projection

2.1. Modeling Within-class Submanifold

Let samples $\{\mathbf{x}_i|\mathbf{x}_i \in \mathbb{R}^d\}_{i=1}^n$ belong to c classes, the number of the samples in the ith class be n_i, and $\mathbf{x}_j^i$ denotes the jth samples in the ith class. Recall the LLE algorithm [6], LLE supposes that the data point, sampling from the manifold of the data, can be reconstructed by a linear combination of its k-nearest neighbors; furthermore, the locally geometric characteristics is valid for the local neighbors on the manifold of the data. As a result, the low dimensional embedding of LLE preserves the neighboring geometry of the high dimensional space [8]. One of the aims of NDP is to preserve the within-class neighboring geometry. Therefore, similar to LLE, we assume that every facial image $\mathbf{x}_j^i$, sampling from the nonlinear submanifold of the image space, can be reconstructed by the linear combination of the other samples in the ith class. Moreover, the weight w_{ij}, reflecting the contribution of the jth facial image to the reconstruction of the ith facial image, should be preserved in the lower dimensional face space. We call w_{ij} the *within-class*

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

IEEE COMPUTER SOCIETY

affinity weight. The weight matrix $\mathbf{W}$ can be computed by minimizing the following reconstruction error [8]:

$$\min \sum_i \left\| \mathbf{x}_i - \sum_j w_{ij}\mathbf{x}_j \right\|^2 \quad (1)$$

with two constraints [8]:

1. $w_{ij} = 0$, if the pair of the samples $\mathbf{x}_i$ and $\mathbf{x}_j$ from different classes
2. $\sum_j w_{ij} = 1, i = 1, 2, \cdots, n$

Due to space limits, the details resolving $\mathbf{W}$ can see [8].

Let $\mathbf{Q} \in \mathbb{R}^{d\times\ell}$ be the transformation matrix, and $\{\mathbf{y}_i = \mathbf{Q}^T\mathbf{x}_i | \mathbf{y}_i \in \mathbb{R}^\ell\}_{i=1}^n$ are the projected vectors. In order to make the projected vectors preserve the local geometry of the image space, according to LLE [8], we should minimize the following cost function:

$$\begin{aligned} J_{\min}(\mathbf{Q}) &= \sum_i \left\| \mathbf{y}_i - \sum_j w_{ij}\mathbf{y}_j \right\|^2 \\ &= \sum_i \left\| \mathbf{Q}^T\mathbf{x}_i - \sum_j w_{ij}(\mathbf{Q}^T\mathbf{x}_j) \right\|^2 \quad (2) \\ &= tr(\mathbf{Q}^T\mathbf{X}(\mathbf{I}-\mathbf{W})^T(\mathbf{I}-\mathbf{W})\mathbf{X}^T\mathbf{Q}) \end{aligned}$$

where the symbol "*tr*" denotes the operation of trace, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n]$ and $\mathbf{I} = diag(1, 1, \cdots, 1)$.

2.2. Modeling Between-class Submanifold

Since the purpose of NDP is to solve the classification problems, we should make the projected vectors of the samples of different classes far from each other. Let weight w'_{ij} reflect the similarity between $\mathbf{x}_i$ and $\mathbf{x}_j$ from different classes. If $\mathbf{x}_i$ is one of $\mathbf{x}_j$'s k-nearest neighbors or $\mathbf{x}_j$ is one of $\mathbf{x}_i$'s k-nearest neighbors, the weight w'_{ij} can be defined as heat kernel [2]:

$$w'_{ij} = \exp(-\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2 / t) \quad (3)$$

Otherwise, $w'_{ij} = 0$. We call w'_{ij} the *between-class affinity weight*. It is obvious that the matrix $\mathbf{W}'$ is sparse and symmetric. Because the weight w'_{ij} will be very small or further $w'_{ij} = 0$ if the pair of facial images $\mathbf{x}_i$ and $\mathbf{x}_j$ from different classes is distant, the weight w'_{ij} reinforces the pair of facial images $\mathbf{x}_i$ and $\mathbf{x}_j$ from different classes if $\mathbf{x}_i$ is one of $\mathbf{x}_j$'s k-nearest neighbors or $\mathbf{x}_j$ is one of $\mathbf{x}_i$'s k-nearest neighbors. In order to make the projected vectors of the samples of different classes far from each other, we can maximize the following cost function:

$$\begin{aligned} J_{\max}(\mathbf{Q}) &= \sum_{ij} \left\|\mathbf{y}_i - \mathbf{y}_j\right\|^2 w'_{ij} \\ &= tr\{\sum_{ij} \mathbf{Q}^T(\mathbf{x}_i - \mathbf{x}_j)w'_{ij}(\mathbf{x}_i - \mathbf{x}_j)^T\mathbf{Q}\} \\ &= tr\{\mathbf{Q}^T(2\mathbf{X}\mathbf{D}'\mathbf{X}^T - 2\mathbf{X}\mathbf{W}'\mathbf{X}^T)\mathbf{Q}\} \quad (4) \\ &\propto tr(\mathbf{Q}^T\mathbf{X}(\mathbf{D}'-\mathbf{W}')\mathbf{X}^T\mathbf{Q}) \end{aligned}$$

where $\mathbf{D}'$ is an diagonal matrix with diagonal element $d'_{ii} = \sum_j w'_{ij}$

2.3. Low Dimensional Embedding

From the above theoretic analysis, we should not only maximize the $J_{\max}(\mathbf{Q})$ but also minimize the $J_{\min}(\mathbf{Q})$, i.e. maximizing the following criterion:

$$J_{NDP}(\mathbf{Q}_{opt}) = \arg\max_{\mathbf{Q}} \frac{tr(\mathbf{Q}^T\mathbf{S}_{BN}\mathbf{Q})}{tr(\mathbf{Q}^T\mathbf{S}_{WN}\mathbf{Q})} \quad (5)$$

where $\mathbf{S}_{BN} = \mathbf{X}(\mathbf{D}' - \mathbf{W}')\mathbf{X}^T$ and $\mathbf{S}_{WN} = \mathbf{X}(\mathbf{I} - \mathbf{W})^T(\mathbf{I}-\mathbf{W})\mathbf{X}^T$. The matrix $\mathbf{S}_{BN}$ is called the *between-class neighborhood scatter matrix*. The matrix $\mathbf{S}_{WN}$ is called the *within-class neighborhood scatter matrix*. The rank of the matrix $\mathbf{S}_{WN}$ is at most $n-c$, while the size of the matrix $\mathbf{S}_{WN}$ is $d\times d$. Due to the small sample size problem $d \gg n$, however, the matrix $\mathbf{S}_{WN}$ is singular and can't be applied directly to compute the transformation matrix $\mathbf{Q}$ based on eq.(5).

In a special case, where $tr(\mathbf{Q}^T\mathbf{S}_{WN}\mathbf{Q}) = 0$, the criterion in eq.(5) can be rewritten as:

$$J_{NDP}(\mathbf{Q}_{opt}) = \underset{tr(\mathbf{Q}^T\mathbf{S}_{WN}\mathbf{Q})=0}{\arg\max_{\mathbf{Q}}} \; tr(\mathbf{Q}^T\mathbf{S}_{BN}\mathbf{Q}) \quad (6)$$

Since $\mathbf{S}_{WN}$ are symmetric positive semi-definite, $tr(\mathbf{Q}^T\mathbf{S}_{WN}\mathbf{Q}) = 0$ means that the column vectors of transformation matrix $\mathbf{Q}$ are the eigenvectors corresponding to the zero eigenvalues of $\mathbf{S}_{WN}$. Therefore, we can project the training samples onto the null space of the $\mathbf{S}_{WN}$ and then compute the projection directions by maximizing $tr(\mathbf{Q}^T\mathbf{S}_{BN}\mathbf{Q})$. As a result, we must first obtain the eigenvectors which span the null space of $\mathbf{S}_{WN}$. Due to the small sample size problem, directly computing the bases spanning the null space of $\mathbf{S}_{WN}$ is intractable. However, we can quickly compute the bases spanning the range space of $\mathbf{S}_{WN}$ and then project the training samples onto the null space of $\mathbf{S}_{WN}$.

Suppose $\mathbb{R}^d$ be the image space, B be the null space of $\mathbf{S}_{WN}$, $B^\perp$ be the range space of $\mathbf{S}_{WN}$ and $rank(\mathbf{S}_{WN}) = r$, where

$$B = span\{\gamma_k | \gamma_k\mathbf{S}_{WN} = 0, k = 1, \cdots, d-r\}$$

$B^{\perp} = span\{\gamma_k | \gamma_k \mathbf{S}_{WN} \neq 0, k = d-r+1, \cdots, d\}$

Let $\mathbf{U} = [\gamma_1, \cdots, \gamma_{d-r}]$ and $\bar{\mathbf{U}} = [\gamma_{d-r+1}, \cdots, \gamma_d]$, then we can obtain

$$\mathbf{x}_j^i = \mathbf{U}\mathbf{U}^T\mathbf{x}_j^i + \bar{\mathbf{U}}\bar{\mathbf{U}}^T\mathbf{x}_j^i \quad (7)$$

Let $\mathbf{z}_j^i = \mathbf{U}\mathbf{U}^T\mathbf{x}_j^i = \sum_{k=1}^{d-r} < \gamma_k, \mathbf{x}_j^i > \gamma_k$. Therefore, $\mathbf{z}_j^i$ means that the sample $\mathbf{x}_j^i$ is projected onto the null space of $\mathbf{S}_{WN}$. According to eq.(7), we can rewrite $\mathbf{z}_j^i$ as

$$\mathbf{z}_j^i = \mathbf{x}_j^i - \bar{\mathbf{U}}\bar{\mathbf{U}}^T\mathbf{x}_j^i \quad (8)$$

After projected all the training samples onto the null space of $\mathbf{S}_{WN}$, the criterion $J_{NDP}(\mathbf{Q}_{opt})$ of NDP can be represented as

$$J_{NDP}(\mathbf{Q}_{opt}) = \arg\max_{\mathbf{Q}} tr(\mathbf{Q}^T\mathbf{S}'_{BN}\mathbf{Q}) \quad (9)$$

where $\mathbf{S}'_{BN} = \mathbf{Z}(\mathbf{D}' - \mathbf{W}')\mathbf{Z}^T$, $\mathbf{Z} = [\mathbf{Z}^1, \cdots, \mathbf{Z}^c]$ and $\mathbf{Z}^i = [\mathbf{z}_1^i, \cdots, \mathbf{z}_{n_i}^i]$. From the eq.(9), we can find that the projection directions are the leading eigenvectors corresponding to the nonzero eigenvalues of $\mathbf{S}'_{BN}$. Due to $rank(\mathbf{S}'_{BN}) = c-1$, the number of the projection directions is $c-1$, i.e. $\mathbf{Q} = [\mathbf{q}_1, \cdots, \mathbf{q}_{c-1}]$

Then, the embedding is as follows:

$$\mathbf{y} = \mathbf{Q}^T\mathbf{x} \quad (10)$$

Theorem 1 *The projected vector $\mathbf{y}_j^i$ corresponding to $\mathbf{x}_j^i$ is the discriminative common vector of the ith class, i.e.*

$$\mathbf{y}_{com}^i = \mathbf{Q}^T\mathbf{x}_j^i, 1 \leq i \leq c; 1 \leq j \leq n_i \quad (11)$$

Because of space limits, detailed proof of **Theorem 1** is omitted.

Since all the samples in a given training class correspond to a unique discriminative common vector, we only need to register a discriminative common vector for each training class, which greatly saves storage space; and the projected vector $\mathbf{y}_{test}$ of a new test facial image $\mathbf{x}_{test}$ is only compared to $\{\mathbf{y}_{com}^i\}_{i=1}^{c-1}$, which greatly improves the computational efficiency in the face recognition systems.

3. Experiments

To verify the proposed NDP approach, the well-known ORL databases [7] was used; and the system performance of NDP was compared to the ones of PCA [10], LDA [1], Supervised NPE (SNPE) [3] and Supervised Laplacianfaces (SLF) [5]. For its simplicity, the nearest-neighbor method using Euclidean metric was employed.

In ORL database, there are ten different grey images for each of 40 distinct subjects. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). All the images were taken against a dark homogeneous background. The size of each image is 92×112 pixels with 256 grey levels per pixel. We illustrate five facial images of one individual in Fig. 1.

Figure 1. Five facial images of one individual

In Section 2, we have discussed how to extract a linear face subspace spanned by a set of projection directions $\{\mathbf{q}_i\}_{i=1}^{c-1}$. Then, the projected vector of a facial image can be obtained from eq.(10). Similar to PCA and LDA, we can display the projection direction $\mathbf{q}_i$ as an image, called NDP-faces. Using all the facial images in the ORL database as the training set, we illustrate the first 5 projection directions of PCA, LDA and NDP in Fig. 2. It is very interesting to see that the NDP-faces are similar to Fisherfaces.

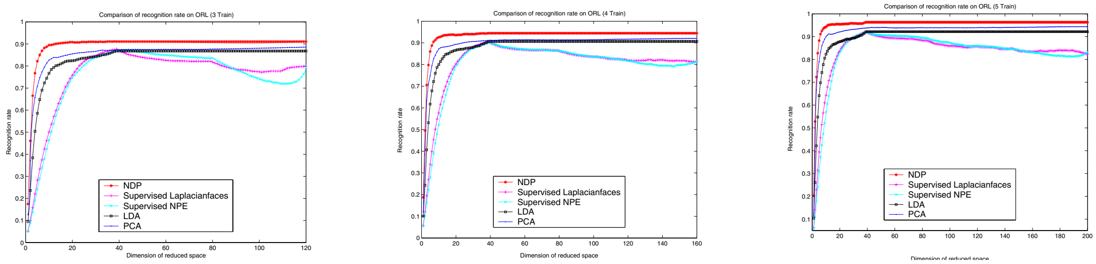
We chose randomly $\xi (= 3, 4, 5)$ different images per individual to form the training set. The rest of the ORL database was used for testing set. For each given ξ, we performed 50 times to choose randomly the training set. The final result is the average recognition rate over 50 random training sets. Fig. 3 illustrates the plot of recognition rate versus the dimension ℓ of reduced space. The top recognition rate achieved by each method and the corresponding dimension ℓ are also shown in Table 1. Note that the upper bound of the dimension ℓ of reduced space is $n-1$, $c-1$, n, $n-1$ and $c-1$ for PCA, LDA, SNPE, SLF and NDP, respectively.

Table 1. Comparison of top recognition rate and corresponding dimension ℓ

Method	3 Train	4 Train	5 Train
PCA	88.55%(119)	92.12%(155)	94.40%(196)
LDA	86.78%(39)	90.67%(39)	92.17%(39)
SNPE	87.04%(40)	90.15%(40)	91.86%(40)
SLF	87.69%(39)	90.52%(39)	91.99%(39)
NDP	**91.11%(36)**	**94.45%(39)**	**96.31%(39)**

From Fig. 3, it is very obvious that the NDP method outperforms the other methods across all the values of ℓ. Although LDE, SNPE and SLF perform well when the value of ℓ is small, their performance impairs as the dimension ℓ increases. From Table 1, we can see that the dimension ℓ of NDP corresponding to the top recognition rate is very low and the performance of NDP improves significantly as

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE



(a) 3 Train (b) 4 Train (c) 5 Train

Figure 3. Recognition rate vs. dimension ℓ of reduced space on the ORL database

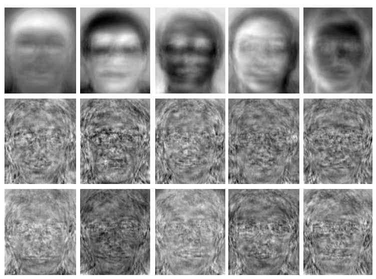


Figure 2. The first 5 projection directions. From top to bottom: Eigenfaces, Fisherfaces and NDP-faces

the number ξ of training samples per individual increases. Although the top recognition rate of PCA is comparative to the one of NDP as the size of the training set increases, the dimension ℓ of PCA corresponding to the top recognition rate is very high.

4. Conclusions

In this paper, we propose a novel manifold learning method named Neighborhood Discriminant Projection for face recognition. In order to preserve the within-class neighboring geometry of the image space and make the projected vectors of the samples of different classes far from each other, NDP explicitly considers the within-class submanifold and the between-class submanifold by the *within-class affinity weight* and the *between-class affinity weight*. Experimental results on ORL database show the effectiveness and robustness of our proposed method.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under grant no. 60021302, and the National High-Tech Research and Development Plan of China under Grant Nos. 2005AA147060.

References

[1] P. N. Belhumeur, J. P. Hepanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[2] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in Neural Information Processing System 14*, 2001.

[3] X. He, D. Cai, S. Yan, and H.-J. Zhang. Neighborhood preserving embedding. In *Proc. of the Tenth IEEE International Conference on Computer Vision*, volume 2, pages 1208–1213, 2005.

[4] X. He and P. Niyogi. Locality preserving projections. In *Advances in Neural Information Processing System 16*, 2003.

[5] X. He, S. Yan, Y. Hu, and H.-J. Zhang. Face recognition using laplacianfaces. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(3):328–340, 2005.

[6] S. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, 2000.

[7] F. S. Samaria and A. C. Harter. Parameterisation of a stochastic model for human face identification. In *Proc. of the 2nd IEEE Workshop on Applications of Computer Vision*, pages 138–142, 1994.

[8] L. K. Saul and S. Roweis. Think globally, fit locally: Unsupervised learning of low dimensional manifolds. *Journal of Machine Learning Research*, 4:119–155, 2003.

[9] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, 2000.

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

[11] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: a literature survey. *ACM Computing Surveys*, 35(4):399–459, 2003.