www.ietdl.org

Published in IET Computer Vision
Received on 11th December 2009
Revised on 7th October 2010
doi: 10.1049/iet-cvi.2009.0138

ISSN 1751-9632

Face recognition using regularised generalised discriminant locality preserving projections

G.-F. Lu[1,2] *Z. Lin*[1] *Z. Jin*[1]

[1]*School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, People's Republic of China*
[2]*Department of Computer Science and Engineering, AnHui University of Technology and Science, WuHu, AnHui 241000, People's Republic of China*
E-mail: luguifu_jsj@163.com

Abstract: Discriminant locality preserving projection (DLPP) is a recently proposed algorithm, which is an extension of locality preserving projections (LPP) and can encode both the geometrical and discriminant structure of the data manifold. However, DLPP suffers from small sample size (SSS) problem which is often encountered in face recognition tasks. To deal with this problem, the authors propose a novel regularised generalised discriminant locality preserving projections (RGDLPP) method for facial feature extraction and recognition. First, locality preserving within-class scatter in DLPP method is replaced by locality preserving total scatter and all the training samples are projected into the range of locality preserving total scatter. Then the authors regularise the small and zero eigenvalues of locality preserving within-class scatter since the small eigenvalues are sensitive to noise. RGDLPP address SSS problem by removing the null space of locality preserving total scatter without loss of discriminant information. Meanwhile, RGDLPP can alleviate the problem of noise disturbance of the small eigenvalues. Experiments on the ORL, Yale, FERET and PIE face databases show the effectiveness of the proposed RGDLPP.

1 Introduction

Dimensionality reduction has been a key problem in many fields of information processing, such as machine learning, data mining, information retrieval and pattern recognition. In the past several decades, many useful techniques for dimensionality reduction have been developed. Linear combination of features is of particular interest since it is simple to calculate and analytically analyse. That is, dimensionality reduction is realised via linear projection. The most well-known techniques are the principal component analysis (PCA) [1] and the Fisher linear discriminant analysis (LDA) [2]. PCA looks for a subspace where the samples have the minimum reconstruction error. LDA aims to better discriminate patterns of different classes by searching the projection axes on which the data points of different classes are far from each other, while constraining the data points of the same to be as close to each other as possible. Unfortunately, it cannot be applied directly to small size sample (SSS) problem [3]. To address this problem, extensive methods have been proposed in the literature [4–13].

Belhumeur *et al.* [4] proposed a two-stage PCA + LDA method, also known as the Fisherface method, in which PCA is first used for dimension reduction so as to make within-class scatter matrix non-singular before the application of LDA. However, to make within-class scatter matrix non-singular, some useful discriminatory information may be lost. Direct LDA [5] method removes null space of the between-class scatter matrix and extracts the eigenvectors corresponding to the smallest eigenvalues of the within-class scatter matrix. In [6] a LDA-based method that makes use of the null space of within-class scatter matrix was proposed. All the samples are first projected onto the null space of within-class scatter matrix, where the within-class scatter is zero, and then the optimal discriminant vectors of LDA are those vectors that maximise the between-class scatter. PCA is used to yield them. However, the computational complexity of determining the null space of within-class scatter matrix is also very high because of the high dimension of within-class scatter matrix. Huang *et al.* [7] proposed a PCA + null space method to deal with SSS problem. In this method, at first, PCA is applied to remove the null space of total scatter matrix of the training sets samples. Then, the optimal projection vectors are found in the remaining lower dimensional space by using the null space method. In [8] Yang and Yang proposed a variation of this method which can extracts features separately from the principal and null space of within-class scatter matrix. In [9] Lu *et al.* proposed a direct fractional-step LDA (DF-LDA) which combines the strengths of the direct LDA and fractional-step LDA [10] approaches while at the same time overcomes their shortcoming and limitations. The above approaches focus on the problem of singularity of within-class scatter matrix. In fact, the instability and noise disturbance of the small eigenvalues cause great problems when the inverse of matrix is applied in the whitening process of various LDA approaches. Dai *et al.* [11] proposed a three-parameter regularised discriminant analysis (RDA) method which regularises the eigenvalues of within-class scatter matrix to solve SSS problem and the problem of noise disturbance of the small eigenvalues. However, it is difficult to determine three optimal parameters because of the computational complexity. To overcome the complexity problem, Chen *et al.* [12] further proposed a single parameter RDA algorithm. Jiang *et al.* [13] also proposed an approach for facial eigenfeature regularisation and extraction. In this method, image space spanned by the eigenvectors of within-class scatter matrix is first decomposed into three subspaces. Then eigenfeatures are regularised differently in these subspaces based on an eigenspectrum model and the optimal projection vectors can derive features both from the principal and null space of within-class scatter matrix. Lately local descriptors such as local binary patterns (LBP) [14, 15] have also gained attention because of their robustness to challenge such as pose and illumination changes.

Recent studies have shown that the face images possibly reside on a non-linear submanifold [16–24]. Many manifold-based learning algorithms have been proposed, for example, Isomap [16], locally linear embedding (LLE) [17] and Laplacian eigenmap [18]. Recently, Yan *et al.* [19] introduced a general framework for dimensionality reduction, called graph embedding, where a large number of popular dimensionality reduction, for example, PCA, LDA, Isomap, LLE and Laplacian eigenmap, could be considered as special cases within the framework. Manifold-based learning algorithms have been shown to be effective in discovering the geometrical structure of the underlying manifold. However, how to evaluate the maps they generated on novel test data points remains unclear. He *et al.* [20, 21] proposed the locality preserving projections (LPP) that build a graph incorporating neighbourhood information of the data set and provide a way to the projection of the novel test data points. In contrast to most manifold learning algorithms, LPP possess a remarkable advantage that it can generate an explicit map. The objective is linear and can be easily computed, like PCA and LDA. Based on LPP, the Laplacianfaces was further developed for face recognition [21], giving encouraging performance. To consider the discriminant information of recognition task, several locality preserving discriminant analysis methods have been mentioned in recent years. Hu [22] proposed an orthogonal neighbourhood preserving discriminant analysis method, which effectively combines the characteristics of LDA and LPP. Yu *et al.* [23] presented a discriminant locality preserving projections (DLPP) method to improve the classification performance of LPP. All the mentioned locality preserving methods also suffer from SSS problem too. So PCA approach, which discards some useful discriminatory information, is often used before LPP or DLPP. Yang *et al.* [24] proposed a null space discriminant locality preserving projections (NDLPP) algorithms. However, NDLPP merely utilises the discriminant information in the null space of the locality preserving within-class scatter.

In this paper, to overcome SSS problem encountered by DLPP, we propose a regularised generalised discriminant locality preserving projections (RGDLPP) method. At first, we replace locality preserving within-class scatter in DLPP approach by locality preserving total scatter. Then, to alleviate the problem of unreliable small and zero eigenvalues caused by noise and the limited number of training samples, a method for regularising the small and zero eigenvalues of locality preserving within-class scatter is proposed. It enables RGDLPP to be executed in the full sample space and alleviates the over-fitting problem. An efficient algorithm for implementing RGDLPP is also developed without any loss of effective discriminatory information. Extensive experimental studies on the ORL, Yale, FERET and PIE face databases show the effectiveness of the proposed RGDLPP method.

The organisation of rest of this paper is as follows. In Section 2, we review briefly the discriminant LPP. In Section 3, we propose the idea and describe the new method in detail. In Section 4, experiments with face images data are presents to demonstrate the effectiveness of the RGDLPP algorithm. Conclusions are made in Section 5.

2 Outline of discriminant LPP

A set of face image samples $\{\boldsymbol{x}_i\}$ can be represented as an $M \times N$ matrix $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N]$, where M is the number of pixels in the images and N is the number of samples. Each face image $\boldsymbol{x}_i$ belongs to one of the C face classes $X_1, \ldots, X_C$. DLPP tries to maximise an objective function as follows:

$$\frac{\Sigma_{i,j=1}^{C}(\boldsymbol{m}_i - \boldsymbol{m}_j)B_{ij}(\boldsymbol{m}_i - \boldsymbol{m}_j)^{\mathrm{T}}}{\Sigma_{c=1}^{C}\Sigma_{i,j=1}^{n_c}(\boldsymbol{y}_i^c - \boldsymbol{y}_j^c)W_{ij}^c(\boldsymbol{y}_i^c - \boldsymbol{y}_j^c)^{\mathrm{T}}} \tag{1}$$

where n_c is the number of samples in the cth class, $\boldsymbol{y}_i^c$ represents the ith projected vector in the cth class, $\boldsymbol{m}_i$ and $\boldsymbol{m}_j$ are separately the mean projected vector for the ith class and jth class, that is, $\boldsymbol{m}_i = (1/n_i)\Sigma_{k=1}^{n_i}\boldsymbol{y}_k^i$ and $\boldsymbol{m}_j = (1/n_j)\Sigma_{k=1}^{n_j}\boldsymbol{y}_k^j$, where n_i and n_j are the number of samples in the ith and jth class, separately. W_{ij}^c represents the elements of within-class weight matrix and $W_{ij}^c = \exp(-\|\boldsymbol{x}_i^c - \boldsymbol{x}_j^c\|^2/\sigma^2)$, and B_{ij} represents the elements of between-class weight matrix and $B_{ij} = \exp(-\|\boldsymbol{f}_i - \boldsymbol{f}_j\|^2/\sigma^2)$, where σ is an empirically determined parameter, $\boldsymbol{x}_i^c$ represents the ith vector in the cth class, $\boldsymbol{f}_i$ is the mean of the ith class, that is, $\boldsymbol{f}_i = (1/n_i)\Sigma_{k=1}^{ni}\boldsymbol{x}_k^i$. Thus, the between-class weight matrix is $\boldsymbol{B} = [B_{ij}]$ $(i, j = 1, 2, \ldots, C)$, the within-class weight matrix is $\boldsymbol{W} = \mathrm{diag}(W^{(1)}, \ldots, W^{(C)})$, where $W^{(i)} = [W_{jk}^{(i)}]$ $(j, k = 1, 2, \ldots, n_i)$. It is clear that both $\boldsymbol{B}$ and $\boldsymbol{W}$ are symmetric positive semi-definite matrices.

Suppose that the mapping from $\boldsymbol{x}_i$ to $\boldsymbol{y}_i$ is A, that is, $\boldsymbol{y}_i = A^{\mathrm{T}}\boldsymbol{x}_i$, then the objective function (1) can be rewritten as

$$J_1(A) = \frac{A^{\mathrm{T}}F\boldsymbol{H}F^{\mathrm{T}}A}{A^{\mathrm{T}}X\boldsymbol{L}X^{\mathrm{T}}A} \tag{2}$$

where $\boldsymbol{L}$ and $\boldsymbol{H}$ are Laplacian matrices. $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$, $\boldsymbol{D} = \mathrm{diag}(D_1, \ldots, D_C)$, $\boldsymbol{D}_i$ is a diagonal matrix and its elements are column (or row) sum of $W^{(i)}$; $\boldsymbol{H} = \boldsymbol{E} - \boldsymbol{B}$, $\boldsymbol{E}$ is a diagonal matrix and its elements are column (or row) sum of $\boldsymbol{B}$, that is, $E_{ii} = \Sigma_j B_{ij}$, $F = [\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_C]$.

Now we would give the following definitions:

- *locality preserving within-class scatter*: $\boldsymbol{S}_w^L = XLX^{\mathrm{T}}$;
- *locality preserving between-class scatter*: $\boldsymbol{S}_b^L = FHF^{\mathrm{T}}$;
- *locality preserving total scatter*: $\boldsymbol{S}_t^L = \boldsymbol{S}_b^L + \boldsymbol{S}_w^L$.

It is clear that $\boldsymbol{S}_w^L$, $\boldsymbol{S}_b^L$ and $\boldsymbol{S}_t^L$ are all symmetric positive semi-definite matrices. The transformation matrix $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k]$ that maximises the objective function (2) can be obtained by solving the generalised eigenvalues problem

$$(\boldsymbol{FHF}^{\mathrm{T}})\boldsymbol{a}_i = \lambda_i(X\boldsymbol{L}X^{\mathrm{T}})\boldsymbol{a}_i, \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \tag{3}$$

or

$$\boldsymbol{S}_b^L\boldsymbol{a}_i = \lambda_i\boldsymbol{S}_w^L\boldsymbol{a}_i, \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \tag{4}$$

DLPP requires that matrix $X\boldsymbol{L}X^{\mathrm{T}}$ be non-singular. For many applications involving SSS problem, this matrix is singular. In fact, as long as the dimension of sample M is greater than the number of samples N, $\boldsymbol{XL}X^{\mathrm{T}}$ must be singular. This can be induced from simple matrix computation knowledge, $\mathrm{rank}(X\boldsymbol{L}X^{\mathrm{T}}) \leq \mathrm{rank}(\boldsymbol{L}) \leq N < M$. Thus, DLPP cannot be applied directly. So PCA approach, which discards some useful discriminatory information, is used before DLPP.

3 Regularised generalised discriminant LPP

3.1 Fundamentals

To overcome SSS problem encountered by DLPP and alleviate the problem of noise disturbance of the small eigenvalues, in this section, we propose a novel approach named RGDLPP. First, we replace $\boldsymbol{S}_w^L$ in DLPP criterion by $\boldsymbol{S}_t^L$. Therefore, (2) is changed to

$$J_2(A) = \frac{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{S}_b^L\boldsymbol{A}}{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{S}_t^L\boldsymbol{A}} \tag{5}$$

The transformation matrix $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k]$ that maximises the objective function (5) can be obtained by solving the generalised eigenvalues problem

$$S_b^L\boldsymbol{a}_i = \lambda_i S_t^L\boldsymbol{a}_i, \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \tag{6}$$

The rank of the matrix $\boldsymbol{S}_t^L$ is, in general, greater than that of the matrix $\boldsymbol{S}_w^L$. But $\boldsymbol{S}_t^L$ can still be singular. If we use PCA approach to make $\boldsymbol{S}_t^L$ non-singular, some effective discriminatory information will be discarded.

The matrices $\boldsymbol{S}_w^L$ and $\boldsymbol{S}_b^L$ are both positive semi-definite, so the intersection of their null spaces is equal to the null space of $\boldsymbol{S}_t^L$, namely, $\boldsymbol{x}|\boldsymbol{S}_t^L\boldsymbol{x} = 0$. As the null space of $\boldsymbol{S}_t^L$ does not contain discriminating information for the training data ($\boldsymbol{x}^{\mathrm{T}}\boldsymbol{S}_w^L\boldsymbol{x} = 0$ and $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{S}_b^L\boldsymbol{x} = 0$), it may be removed from the solution space without accuracy. Assume that the eigenvalue decomposition of matrix $\boldsymbol{S}_t^L$ is

$$\boldsymbol{S}_t^L = U\Lambda U^{\mathrm{T}} \tag{7}$$

where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{m'})$, $\lambda_i > 0$, $i = 1, 2, \ldots, m'$, and m' is the number of positive singular values of $\boldsymbol{S}_t^L$, $U = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{m'}]$ are the eigenvectors of $\boldsymbol{S}_t^L$ corresponding to eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_{m'}$. Therefore, (5) can be changed to

$$J_2(P) = \frac{\boldsymbol{P}^{\mathrm{T}}\tilde{S}_b^L P}{P^{\mathrm{T}}\tilde{S}_t^L P} = \frac{P^{\mathrm{T}}\tilde{S}_b^L P}{P^{\mathrm{T}}\Lambda P} \tag{8}$$

where $\tilde{S}_b^L = U^{\mathrm{T}}S_b^L U$, $\tilde{S}_t^L = U^{\mathrm{T}}S_t^L U = \Lambda$, and $A = UP$, where $\boldsymbol{P} = [\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_k] \in \boldsymbol{R}^{m'\times k}$ is the transform matrix that maximises the objective function (8). Then, the denominator of the objective function (8) is always positive for non-zero P, that is, $\tilde{S}_t^L$ is positive definite.

Let λ to be eigenvalue of (8). It is obvious that $\lambda \leq 1$. When $\boldsymbol{p}$ is in the null space of $\tilde{S}_w^L = U^{\mathrm{T}}S_w^L U$, we have $\tilde{S}_w^L\boldsymbol{p} = 0$. Note that $\tilde{S}_t^L = \tilde{S}_w^L + \tilde{S}_b^L$; thus, we have

$$\boldsymbol{p}^{\mathrm{T}}\tilde{S}_t^L\boldsymbol{p} = \boldsymbol{p}^{\mathrm{T}}(\tilde{S}_w^L + \tilde{S}_b^L)\boldsymbol{p} = \boldsymbol{p}^{\mathrm{T}}\tilde{S}_b^L\boldsymbol{p} \tag{9}$$

From (9), we have

$$\frac{\boldsymbol{p}^{\mathrm{T}}\tilde{S}_b^L\boldsymbol{p}}{\boldsymbol{p}^{\mathrm{T}}\tilde{S}_t^L\boldsymbol{p}} = 1 \tag{10}$$

From (10), we obtain that all the eigenvectors in null space of $\tilde{S}_w^L$ share the same maximal eigenvalue (=1). Obviously, the null space of $\tilde{S}_w^L$ is unduly overemphasised, which leads to over-fitting problem and poor generalisation. Besides, the small eigenvalues are very sensitive to noise, which may cause great problems when the inverse of $\tilde{S}_t^L$ is applied. So, we should regularise the small and zero eigenvalues.

Inspired by the literature [11–13], we propose a method for regularising the small and zero eigenvalue of $\tilde{S}_w^L$. Assume that the singular value decomposition of matrix $\tilde{\boldsymbol{S}}_w^L$ is

$$\tilde{\boldsymbol{S}}_w^L = \tilde{U}_w\tilde{\Lambda}_w\tilde{U}_w^{\mathrm{T}} \tag{11}$$

where $\tilde{\Lambda}_w = \mathrm{diag}\{\tilde{\lambda}_1^w, \tilde{\lambda}_2^w, \ldots \tilde{\lambda}_r^w, \tilde{\lambda}_{r+1}^w, \ldots, \tilde{\lambda}_{m'}^w\}$, $r = \mathrm{rank}(\tilde{S}_w^L)$, $\tilde{U}_w = [\tilde{\boldsymbol{u}}_1^w, \tilde{\boldsymbol{u}}_2^w, \ldots, \tilde{\boldsymbol{u}}_{m'}^w]$ is the eigenvectors of $\tilde{\boldsymbol{S}}_w^L$ corresponding to eigenvalues $\tilde{\lambda}_1^w, \tilde{\lambda}_2^w, \ldots, \tilde{\lambda}_{m'}^w$, and the eigenvalues are sorted in descending order $\tilde{\lambda}_1^w \geq \cdots \geq \tilde{\lambda}_{m'}^w$. The regularised eigenspectrum $\hat{\lambda}_k^w$ is given by

$$\hat{\lambda}_i^w = \begin{cases} \tilde{\lambda}_i^w, & i < m \\ \dfrac{\alpha}{i+\beta}, & m \leq i \leq r \\ \dfrac{\alpha}{r+1+\beta}, & r < i \leq m' \end{cases} \tag{12}$$

where α, β are constant and are given by $\alpha = (\tilde{\lambda}_1^w\tilde{\lambda}_m^w(m-1)/\tilde{\lambda}_1^w - \tilde{\lambda}_m^w)$ and $\beta = (m\tilde{\lambda}_m^w - \tilde{\lambda}_1^w/\tilde{\lambda}_1^w - \tilde{\lambda}_m^w)$, respectively, m is estimated by $\tilde{\lambda}_{m+1}^w = \max\{\forall\tilde{\lambda}_i^w|\tilde{\lambda}_i^w < (\tilde{\lambda}_{\mathrm{med}}^w + 0.9(\tilde{\lambda}_{\mathrm{med}}^w - \tilde{\lambda}_r^w))\}$, where $\tilde{\lambda}_{\mathrm{med}}^w = \mathrm{median}\{\forall\tilde{\lambda}_i^w|i < r\}$. Therefore the criterion of RGDLPP can be defined as follows

$$J_2(T) = \frac{T^{\mathrm{T}}\hat{S}_b^L T}{T^{\mathrm{T}}\hat{S}_t^L T} = \frac{T^{\mathrm{T}}\hat{S}_b^L T}{T^{\mathrm{T}}(\hat{S}_b^L + \hat{\Lambda}_w)T} \tag{13}$$

where $\hat{\Lambda}_w = \mathrm{diag}(\hat{\lambda}_i^w)$, $i = 1, 2, \ldots, m'$, $\hat{S}_b^L = \tilde{U}_w^{\mathrm{T}}\tilde{S}_b^L\tilde{U}_w$, and $\hat{S}_t^L = \hat{S}_b^L + \hat{\Lambda}_w$. The optimal projection matrix is $\boldsymbol{A} = U\tilde{U}_w T$ where the column vectors of T are the leading eigenvectors of $(\hat{S}_b^L + \hat{\Lambda}_w)^{-1}\hat{S}_b^L$.

3.2 Computational consideration

In the above subsection, we should compute eigenvalue decomposition of matrix $\boldsymbol{S}_t^L$ (5). However, in real-world application of such face recognition, gene expression and web document recognition, the dimension M of the vector samples is usually large, so it is difficult to solve the eigenvector of the $M \times M$ matrix $\boldsymbol{S}_t^L$ directly. Besides, there is still attendant problem of numerical accuracy when diagonalising large matrix directly [25].

The Laplacian matrices $\boldsymbol{L}$ and $\boldsymbol{H}$ are always real symmetric positive semi-definite, so $\boldsymbol{L}$ and $\boldsymbol{H}$ can be decomposed as follows

$$L = V_L \Lambda_L V_L^{\mathrm{T}}, \quad H = V_H \Lambda_H V_H^{\mathrm{T}} \tag{14}$$

where Λ_L is the eigenvalue matrix of $\boldsymbol{L}$, that is, $\Lambda_L = \mathrm{diag}(\lambda_1^L, \lambda_2^L, \ldots, \lambda_N^L)$, the column of V_L are the orthogonal eigenvectors corresponding to eigenvalues of L; Λ_H is the eigenvalue matrix of $\boldsymbol{H}$, that is, $\Lambda_H = \mathrm{diag}(\lambda_1^H, \lambda_2^H, \ldots, \lambda_C^H)$, the column of V_H are the orthogonal eigenvectors corresponding to eigenvalues of $\boldsymbol{H}$.

It is easy to know that all the eigenvalues of both $\boldsymbol{L}$ and $\boldsymbol{H}$ are non-negative since both L and H are real symmetric semi-positive definite matrices. Consequently, $\boldsymbol{S}_w^L$, $\boldsymbol{S}_b^L$ and $\boldsymbol{S}_t^L$ can be rewritten as

$$S_w^L = \mathrm{XLX}^{\mathrm{T}} = H_w H_w^{\mathrm{T}} \tag{15}$$

$$S_b^L = \mathrm{FHF}^{\mathrm{T}} = H_b H_b^{\mathrm{T}} \tag{16}$$

$$S_t^L = S_w^L + S_b^L = H_t H_t^{\mathrm{T}} \tag{17}$$

where $H_w = XV_L\Lambda_L^{1/2} \in \boldsymbol{R}^{M\times N}$, $H_b = FV_H\Lambda_H^{1/2} \in \boldsymbol{R}^{M\times C}$ and $H_t = [H_w \backslash H_b] \in \boldsymbol{R}^{M\times(N+C)}$. Assume that the thin singular value decomposition of matrix $\boldsymbol{H}_t$ is

$$H_t = U\Lambda_t Q^{\mathrm{T}} \tag{18}$$

where Λ_t is the singular value matrix of $\boldsymbol{H}_t$ and $\Lambda = \Lambda_t^2$, $U = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{m'}]$ is the left singular vector matrix of $\boldsymbol{H}_t$, and Q is the right singular vector matrix of $\boldsymbol{H}_t$. So, we can obtain the eigenvectors $U = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{m'}]$ of $\boldsymbol{S}_t^L$ by the thin singular value decomposition of matrix $\boldsymbol{H}_t$. Note that the size of H_w, H_b and H_t are much smaller than that of $\boldsymbol{S}_t^L$, since usually $N \ll M$ and $C \ll M$.

Now, the algorithmic procedure of RGDLPP is formally summarised as follows:

1. Construct the within-class weight matrix $\boldsymbol{W}$ and between-class weight matrix $\boldsymbol{B}$. Then calculate the within-class Laplacian matrix $\boldsymbol{L}$ and between-class Laplacian matrix $\boldsymbol{H}$.
2. Compute $\boldsymbol{H}_w$, $\boldsymbol{H}_b$ and $\boldsymbol{H}_t$ by (14)–(17).
3. Perform the thin singular value decomposition of matrix $\boldsymbol{H}_t$ as (18).
4. Compute $\tilde{S}_b^L = U^{\mathrm{T}} S_b^L U$ and $\tilde{S}_w^L = U^{\mathrm{T}} S_w^L U$.
5. Solve the singular value decomposition of matrix $\tilde{\boldsymbol{S}}_w^L$ as (11).
6. Regularisation of $\tilde{\Lambda}_w$ to $\hat{\Lambda}_w$ according to (12).
7. Solve the eigenvalue problem $(\hat{S}_b^L + \hat{\Lambda}_w)^{-1}\hat{S}_b^L \boldsymbol{t} = \lambda \boldsymbol{t}$. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$ be the k largest eigenvalues of $(\hat{S}_b^L + \hat{\Lambda}_w)^{-1}\hat{S}_b^L$ and $\boldsymbol{t}_1, \boldsymbol{t}_2, \ldots, \boldsymbol{t}_k$ be the associated eigenvectors.
8. The optimal projection matrix is given by $A = U\tilde{U}_w T$, where $U = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{m'}]$, $\tilde{U}_w = [\tilde{\boldsymbol{u}}_1^w, \tilde{\boldsymbol{u}}_2^w, \ldots, \tilde{\boldsymbol{u}}_{m'}^w]$ and $\boldsymbol{T} = [\boldsymbol{t}_1, \boldsymbol{t}_2, \ldots, \boldsymbol{t}_k]$.

This algorithm has several novelty features. First, we construct within-class Laplacian matrix L and between-class Laplacian matrix H directly in the image space, where the local structure of face data points is exactly described. Second, the dimension of feature space is first greatly reduced without loss on discriminative information by removing the null space of $\boldsymbol{S}_t^L$. Third, discriminant evaluation is performed in the full space dimension of the image data. Four, the over-fitting problem of GDLPP is alleviated after regularising the unreliable small and zero eigenvalues caused by noise and the limited number of training samples.

3.3 Theoretical analyses of RGDLPP

As discussed previously, the transformation matrix $\boldsymbol{A}$ that maximises the objective function (5) can be obtained by solving the eigenvalues problem $(\boldsymbol{S}_t^L)^{-1}\boldsymbol{S}_b^L$. When $\boldsymbol{S}_t^L$ is singular, we can use pseudo-inverse to deal with the singularity problem, that is, we can replace $(\boldsymbol{S}_t^L)^{-1}$ by pseudo-inverse $(\boldsymbol{S}_t^L)^+$. The following theorem shows $A = UP$ is the discriminant transformation of $(\boldsymbol{S}_t^L)^+\boldsymbol{S}_b^L$.

Theorem 1: Let $A = UP$, then the columns of $\boldsymbol{A}$ are eigenvectors of $(\boldsymbol{S}_t^L)^+\boldsymbol{S}_b^L$.

Proof: Let $\boldsymbol{p} \in \boldsymbol{R}^{m'\times 1}$ be the eigenvector of (8) corresponding to the eigenvalue λ, then we have

$$\tilde{S}_b^L \boldsymbol{p} = \lambda\Lambda\boldsymbol{p} \tag{19}$$

Since Λ non-singular, (19) can be changed to

$$\Lambda^{-1}\tilde{S}_b^L \boldsymbol{p} = \lambda\boldsymbol{p} \tag{20}$$

Because of $\tilde{S}_b^L = U^{\mathrm{T}} S_b^L U$, (20) can be changed to

$$U\Lambda^{-1}U^{\mathrm{T}} S_b^L (U\boldsymbol{p}) = \lambda(U\boldsymbol{p}) \tag{21}$$

Because of $(\boldsymbol{S}_t^L)^+\boldsymbol{S}_b^L = \boldsymbol{U}\Lambda^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{S}_b^L$, $\boldsymbol{Up}$ is an eigenvector of $(\boldsymbol{S}_t^L)^+\boldsymbol{S}_b^L$. □

4 Experiments and results

In this section, experiments are conducted on four well-known face image databases, that is, ORL, Yale, FERET and PIE to evaluate the performance of the proposed RGDLPP algorithm. PCA [1], LDA [2], LBP [14, 15], LPP [20, 21], DLPP [23] and the proposed method are used for feature extraction. LDA, LPP and DLPP involve a preceding PCA stage to avoid the singularity problem and 98% image energy is kept in PCA phase. For LPP, DLPP and RGDLPP algorithms, the Gaussian Kernel $\exp(-\|x-y\|^2/\sigma^2)$ is used. Note that the value of σ has a great impact on the ultimate performances of LPP, DLPP and RGDLPP algorithms. However, until now, it is still unclear how to choose the optimal σ [19]. The method in [19] for choosing the value of σ is used in our experiments, that is, σ is set as $2^{(e-10)/25}\sigma_0$, $e = 0, 1, \ldots, 20$, where σ_0 is the standard derivation of the training data set. A nearest neighbour classifier with cosine distance is employed to classify in the projected feature space. Cosine distance measure between two vectors, $\boldsymbol{a}$ and $\boldsymbol{b}$, is defined as

$$\cos\_\mathrm{dis}(\boldsymbol{a}, \boldsymbol{b}) = \frac{\langle \boldsymbol{a}, \boldsymbol{b}\rangle}{\|\boldsymbol{a}\|_2 \|\boldsymbol{b}\|_2} \tag{22}$$

where $\| \|_2$ is the norm 2 operator. For LBP method, the $\mathrm{LBP}_{8,2}^{u2}$ operator is used for feature extraction and the χ^2

doi: 10.1049/iet-cvi.2009.0138

measure is used to measure the difference between the histograms [14].

The experiments are implemented on a Mobile DualCore Intel Pentium (1666 MHz) processor Hasee Computer with 898M RAM and the programming environment is MATLAB 7.0.

4.1 Database

The ORL, Yale, FERET and PIE face databases are used in our experiments. The ORL face database consists of a total of 400 face images, of a total of 40 people (10 samples per person). For some subjects, the images were taken at different times, varying the lighting, facial expressions (open/closed eyes, smiling/not smiling) and facial details (glassed/no glassed). All the images were taken against a dark homogeneous background with the subjects in an upright, front position (with tolerance for some side movement). In our experiments, each image in ORL database was manually cropped and resized to 32 × 32.

The Yale face database contains 165 grey scale images of 15 individuals, each individual has 11 images. The images demonstrate variations in lighting condition, facial expression (normal, happy, sad, sleepy, surprised and wink). In our experiments, each image in Yale database was manually cropped and resized to 32 × 32.

The FERET face database contains 14 126 images from 1199 individuals. In our experiments, we select a subset which contains 1400 images of 200 individuals (each individual has seven images). It is composed of the images whose names are marked with two-character strings: ‘ba’, ‘bj’, ‘bk’, ‘be’, ‘bf’, ‘bd’ and ‘bg’. The subset involves variations in facial expression, illumination and pose ($\pm 15^\circ$ and $\pm 25^\circ$). In our experiments, each image in FERET database was manually cropped and resized to 32 × 32.

The CMU PIE face database contains 68 individuals with 41 368 face images as a whole. The face images were captured under varying pose, illumination and expression. In our experiments, we select a subset (C29) that contains

Fig. 1 *Images of one person in ORL*

Fig. 2 *Images of one person in Yale*

Fig. 3 *Images of one person in FERET*

Fig. 4 *Images of one person in PIE*

Fig. 5 *Some reconstructed images (the numbers in parentheses are the reconstruction error)*

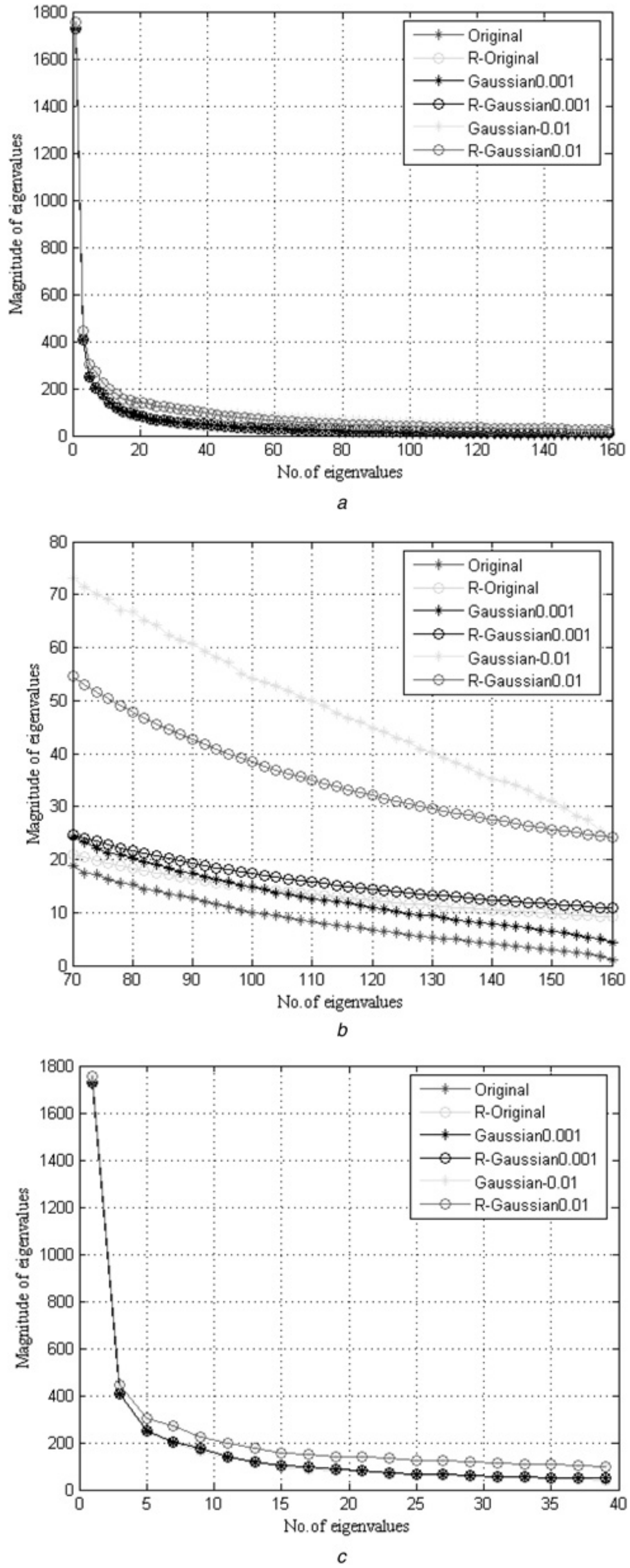

Fig. 6 *Comparison of eigenspectrum*

a Eigenspectrum comparison of all eigenvalues
b Eigenspectrum comparison of the last 90 small eigenvalues
c Eigenspectrum comparison of the first 40 big eigenvalues

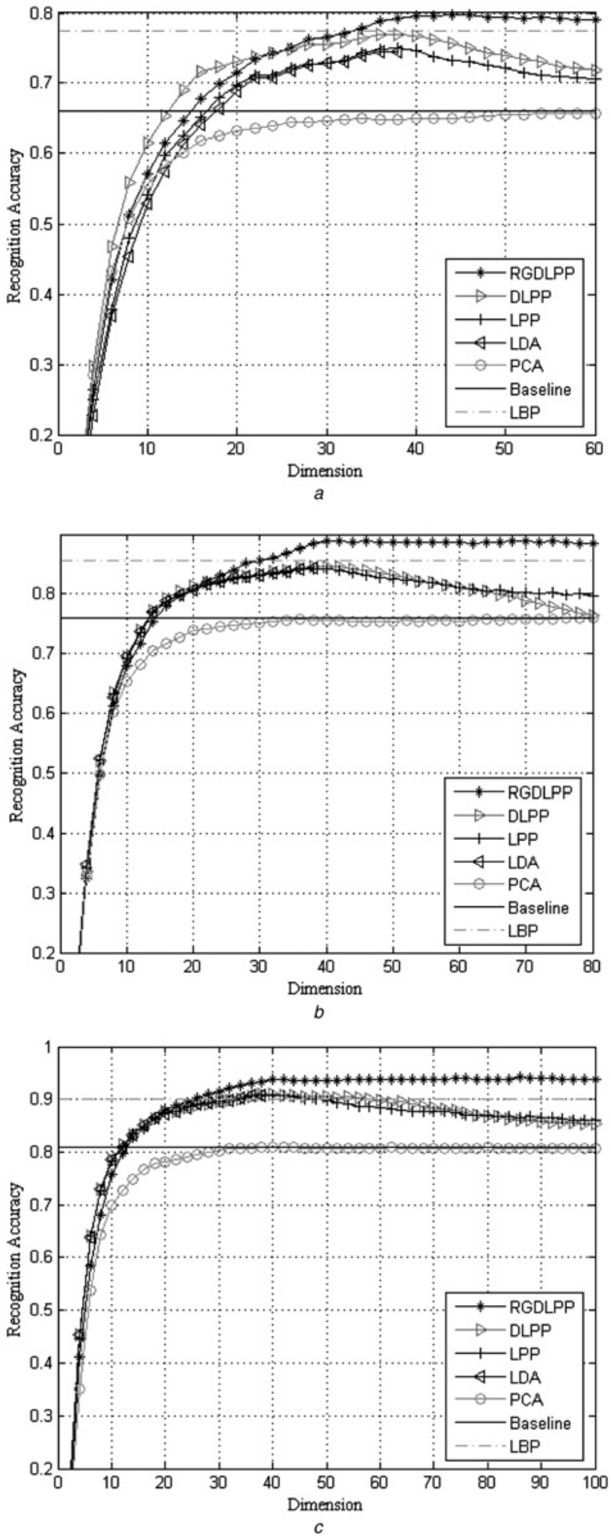

Fig. 7 *Recognition rate against dimension of reduced space on the ORL database*

a Two trains
b Three trains
c Four trains

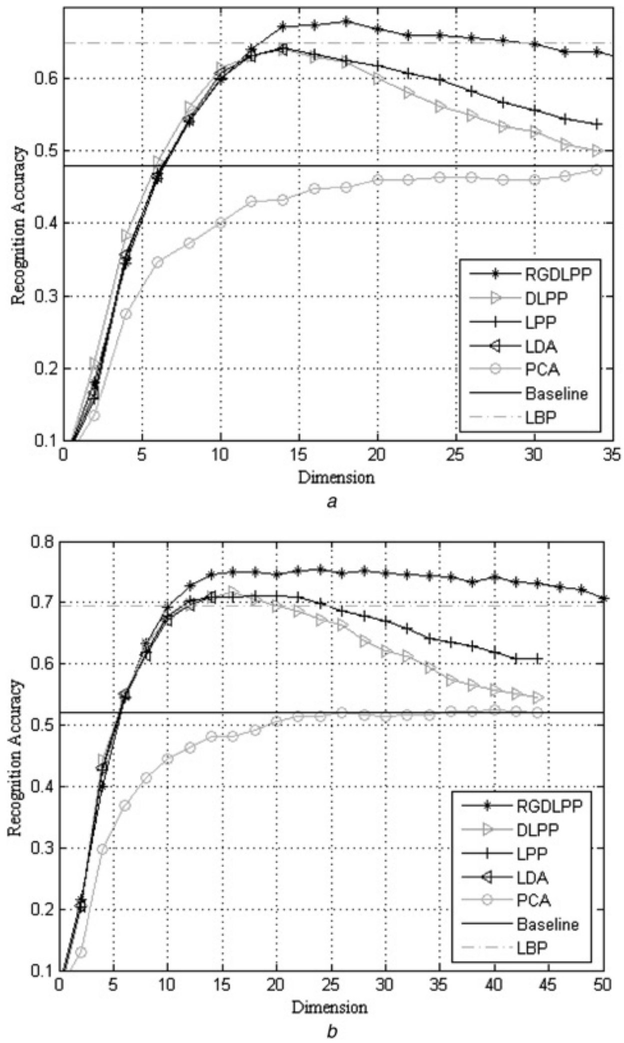

Fig. 8 *Recognition rate against dimension of reduced space on the Yale database*

a Three trains
b Four trains

1632 images of 68 individuals (each individual has 24 images). The C29 subset involves variations in illumination, facial expression and pose. All of these face images are aligned based on eye coordinates and cropped to 64 × 64. Figs. 1–4 show the sample images from the four databases.

4.2 Experiment on images reconstruction

In this subsection, an experiment on images reconstruction is performed to verify that the dimension of the feature space is greatly reduced without loss on discriminative information. We randomly select 160 images from ORL face database. According to the analysis in Section 3.1, the dimension can be reduced from M to m', where M is the number of pixels in the images and m' is the rank of $\boldsymbol{S}_t^L$, respectively. Then, in this experiment, M is 1024 and m' is 159. Obviously, the dimension of the feature space is greatly reduced. Fig. 5 shows an original image from ORL and its five reconstructed images using the first k ($k = 10$, 50, 100, 150, 199) features. The reconstructed images become clearer and the reconstruction errors decrease as the number of features is increased. When the number of feature is 199, the reconstruction error is zero and the original image is perfectly reconstructed.

4.3 Experiment on the robustness of RGDLPP

In this experiment, we evaluate the robustness of RGDLPP to the noise disturbance of the small eigenvalues. We randomly select 200 images from ORL face database. To demonstrate that the small eigenvalues are sensitive to noise, two kinds

doi: 10.1049/iet-cvi.2009.0138

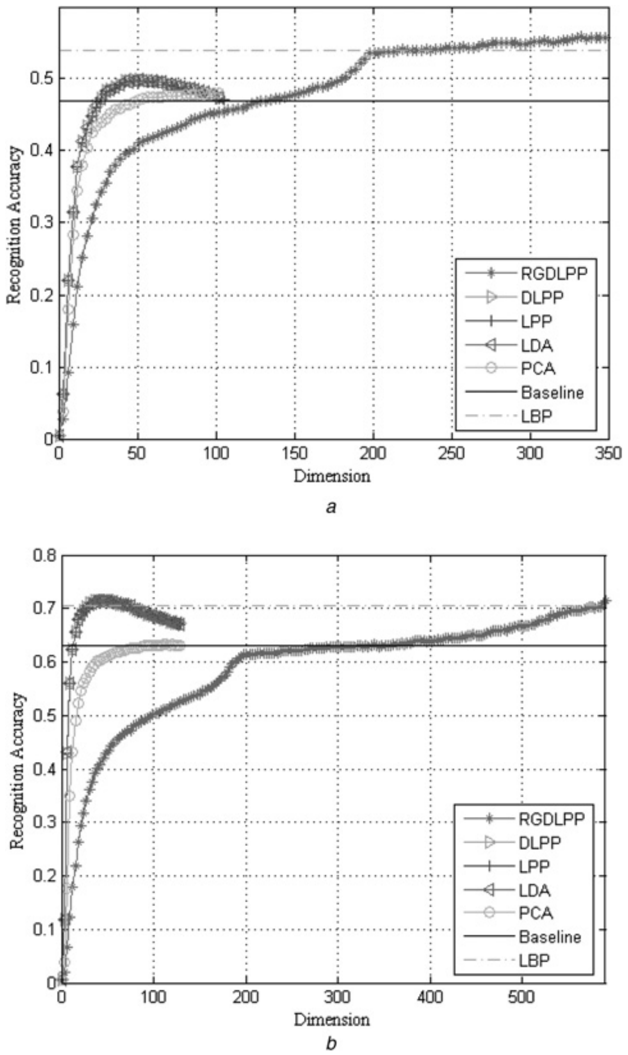

Fig. 9 *Recognition rate against dimension of reduced space on the FERET database*

a Two trains
b Three trains

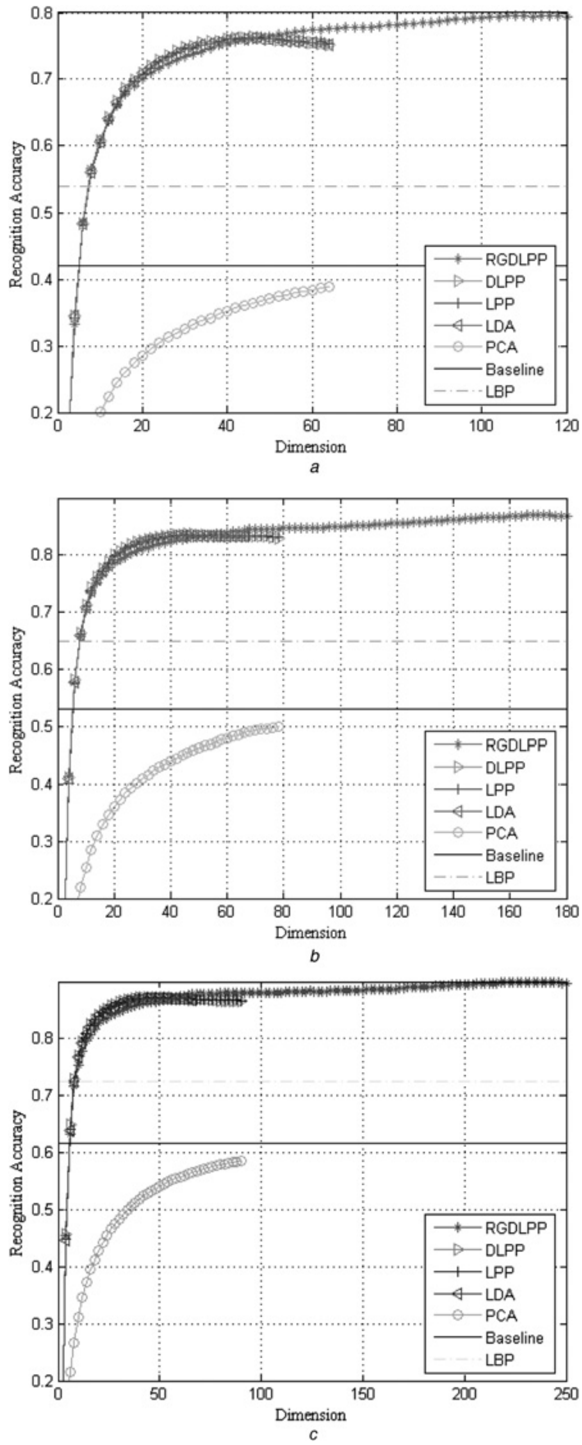

Fig. 10 *Recognition rate against dimension of reduced space on the PIE database*

a Two trains
b Three trains
c Four trains

of Gaussian noise, that is, zero mean with 0.001 variance and zero mean with 0.01 variance, are added to the images. Fig. 6 shows the eigenspectrum of $\tilde{\boldsymbol{S}}_w^L$ of original images (the curve of Original), corrupted images with Gaussian noise of zero mean with 0.001 variance (the curve of Gaussian 0.001), and corrupted images with Gaussian noise of zero mean with 0.01 variance (the curve of Gaussian 0.01). Their corresponding regularised eigenspectrum are also shown in Fig. 6, that is, the curve of R-Original, the curve of R-Gaussian 0.001 and R-Gaussian 0.01, respectively.

As shown in Fig. 6, we observe that the small eigenvalues are more sensitive to noise than the big eigenvalues. The differences among R-Original, R-Gaussian 0.001 and R-Gaussian 0.01 are much smaller than the differences among Original, Gaussian 0.001 and Gaussian 0.01. Then we can find the proposed regularised method is much more robust to noise.

4.4 Experimental results and analysis

In this experiment, we compare the performances of different algorithms. We randomly select i ($i = 2$, 3, 4 for ORL and PIE, $i = 2$, 3 for Yale and $i = 2$, 3 for FERET) different samples of each individual for training, and the remaining ones are used for testing. For each given i, we perform 20 times to randomly choose the training set and calculate the average recognition rates as well as the standard deviation. Figs. 7–10 illustrate the plot of recognition rate against the dimension of reduced space for different algorithms. Note that the recognition rates of LPP, DLPP and LDA on FERET and PIE databases are very similar and their performance curves are very similar in Figs. 9 and 10. For the baseline method, we simply perform face matching

Table 1 Recognition accuracy (%) on ORL (mean ± std)

Size	Baseline	LBP	PCA	LDA	LPP	DLPP	RGDLPP
2	66.8 ± 3.4	78.3 ± 2.4	66.4 ± 3.4(60)	75.4 ± 3.1(39)	75.9 ± 3.1(38)	77.7 ± 3.4(38)	80.8 ± 3.5(45)
3	77.0 ± 2.5	85.8 ± 2.6	76.7 ± 2.4(81)	85.1 ± 1.9(39)	85.1 ± 2.1(38)	85.4 ± 2.0(38)	89.8 ± 1.7(42)
4	81.7 ± 2.3	90.3 ± 1.8	82.0 ± 2.1(45)	91.3 ± 1.9(39)	91.3 ± 1.7(40)	91.7 ± 1.8(40)	95.1 ± 1.5(87)

Table 2 Recognition accuracy (%) on Yale (mean ± std)

Size	Baseline	LBP	PCA	LDA	LPP	DLPP	RGDLPP
3	48.1 ± 4.3	66.5 ± 3.7	47.8 ± 4.0(34)	65.4 ± 4.6(14)	65.6 ± 4.7(14)	66.4 ± 4.8(14)	69.6 ± 4.1(18)
4	52.9 ± 4.2	69.7 ± 4.0	53.8 ± 4.8(41)	72.1 ± 5.4(14)	73.3 ± 5.4(18)	73.2 ± 4.8(16)	78.5 ± 3.6(25)

Table 3 Recognition accuracy (%) on FERET (mean ± std)

Size	Baseline	LBP	PCA	LDA	LPP	DLPP	RGDLPP
2	47.8 ± 9.3	54.3 ± 10.2	48.4 ± 8.9(102)	51.0 ± 7.4(51)	51.0 ± 7.3(49)	51.1 ± 7.4(51)	58.0 ± 6.5(357)
3	63.1 ± 8.5	70.5 ± 6.6	63.8 ± 7.9(111)	72.5 ± 8.0(49)	72.6 ± 7.9(48)	72.5 ± 7.9(45)	73.0 ± 5.8(591)

Table 4 Recognition accuracy (%) on PIE (mean ± std)

Size	Baseline	LBP	PCA	LDA	LPP	DLPP	RGDLPP
2	41.9 ± 1.6	54.1 ± 1.9	39.3 ± 1.5(65)	76.6 ± 1.8(46)	76.8 ± 1.7(47)	76.8 ± 1.8(46)	80.0 ± 2.0(119)
3	53.0 ± 1.7	65.1 ± 1.6	50.3 ± 1.7(78)	83.9 ± 1.3(49)	84.0 ± 1.3(46)	84.1 ± 1.3(46)	87.4 ± 1.1(175)
4	61.7 ± 1.9	72.5 ± 1.7	58.8 ± 2.0(91)	87.6 ± 1.1(51)	87.7 ± 0.9(49)	87.8 ± 1.0(50)	90.3 ± 0.9(235)

without any face extractor. Thus, the feature dimensions of the baseline method are 4096 (64×64) for PIE and 1024 (32×32) for ORL, Yale and FERET. For the LBP method, the images are divided into 4×4 regions for PIE and 2×2 regions for ORL, Yale and FERET. Then the size of each sub-region is 16×16. The best performances obtained by different algorithms as well as the corresponding dimensionality of reduced subspace (the numbers in parentheses) on ORL, Yale, FERET and PIE databases are given in Tables 1–4, respectively. To evaluate the computational efficiency of different algorithms, we also give the average total CPU time of each method involved in Tables 5–8. Note that the CPU time of baseline is not reported in Table 2, since we use the raw data without dimensionality reduction.

From Figs. 7–10 and Tables 1–8, we can obtain the following conclusions:

1. For each method, the recognition accuracy increases with the increase of training samples sizes. The reason may be that a large set of training data can sample the underlying distribution more accurately than a smaller set.
2. PCA is simple to perform, but it generally performs much worse than LBP, LDA, LPP, DLPP and RGDLPP. Its recognition rates are just close to the baseline on all used databases.
3. The LDA, LPP, DLPP and RGDLPP methods all outperform the baseline method. The low dimensionality of the face subspace obtained in our experiment show that dimensionality reduction is indeed necessary as a pre-processing for face recognition.
4. On the tested databases, our proposed RGDLPP consistently outperforms PCA, LBP, LPP and DLPP methods.

Table 5 Comparison of CPU time (s) for each method on ORL

Size	LBP	PCA	LDA	LPP	DLPP	RGDLPP
2	2.328	0.078	0.080	0.082	0.156	0.313
3	2.328	0.140	0.156	0.201	0.235	0.641
4	2.328	0.218	0.230	0.264	0.360	1.078

Table 6 Comparison of CPU time (s) for each method on Yale

Size	LBP	PCA	LDA	LPP	DLPP	RGDLPP
3	0.844	0.031	0.032	0.034	0.047	0.094
4	0.844	0.047	0.049	0.050	0.062	0.157

Table 7 Comparison of CPU time (s) for each method on FERET

Size	LBP	PCA	LDA	LPP	DLPP	RGDLPP
2	12.328	1.531	1.752	1.945	1.957	14.297
3	12.328	4.062	4.290	4.634	4.954	39.719

Table 8 Comparison of CPU time (s) for each method on PIE

Size	LBP	PCA	LDA	LPP	DLPP	RGDLPP
2	60.063	1.016	1.150	1.342	1.406	4.297
3	60.063	1.547	1.782	1.961	1.969	7.453
4	60.063	2.250	2.550	2.756	2.828	11.359

IET Comput. Vis., 2011, Vol. 5, Iss. 2, pp. 107–116
doi: 10.1049/iet-cvi.2009.0138

Table 9 *p*-Values between LBP and GDLPP on different face databases

Data	ORL(2)	ORL(3)	ORL(4)	Yale(3)	Yale(4)	FERER(2)	FERET(3)	PIE(2)	PIE(3)	PIE(4)
p-value	0.0636	$4.0054 \times 10^{-5*}$	$1.907 \times 10^{-6*}$	0.3593	$7.6294 \times 10^{-5*}$	0.5034	0.1153	$1.907 \times 10^{-6*}$	$1.907 \times 10^{-6*}$	$1.907 \times 10^{-6*}$

The asterisks indicate a statistically significant difference between LBP and GDLPP at a significance level of 0.05

Table 10 *p*-Values between PCA and GDLPP on different face databases

Data	ORL(2)	ORL(3)	ORL(4)	Yale(3)	Yale(4)	FERER(2)	FERET(3)	PIE(2)	PIE(3)	PIE(4)
p-value	$1.907 \times 10^{-6*}$	$1.907 \times 10^{-6*}$	$1.907 \times 10^{-6*}$	$1.907 \times 10^{-6*}$	$1.907 \times 10^{-6*}$	$1.907 \times 10^{-6*}$	$1.907 \times 10^{-6*}$	$1.907 \times 10^{-6*}$	$1.907 \times 10^{-6*}$	$1.907 \times 10^{-6*}$

The asterisks indicate a statistically significant difference between PCA and GDLPP at a significance level of 0.05

Table 11 *p*-Values between LDA and GDLPP on different face databases

Data	ORL(2)	ORL(3)	ORL(4)	Yale(3)	Yale(4)	FERER(2)	FERET(3)	PIE(2)	PIE(3)	PIE(4)
p-value	$1.907 \times 10^{-6*}$	$1.907 \times 10^{-6*}$	$1.907 \times 10^{-6*}$	$4.0245 \times 10^{-4*}$	$7.6294 \times 10^{-5*}$	$1.907 \times 10^{-6*}$	0.5034	$1.907 \times 10^{-6*}$	$1.907 \times 10^{-6*}$	$1.907 \times 10^{-6*}$

The asterisks indicate a statistically significant difference between LDA and GDLPP at a significance level of 0.05

Table 12 *p*-Values between LPP and GDLPP on different face databases

Data	ORL(2)	ORL(3)	ORL(4)	Yale(3)	Yale(4)	FERER(2)	FERET(3)	PIE(2)	PIE(3)	PIE(4)
p-value	$4.0054 \times 10^{-5*}$	$1.907 \times 10^{-6*}$	$1.907 \times 10^{-6*}$	0.0044^{*}	0.000402^{*}	$1.907 \times 10^{-6*}$	0.5034	$1.907 \times 10^{-6*}$	$1.907 \times 10^{-6*}$	$1.907 \times 10^{-6*}$

The asterisks indicate a statistically significant difference between LPP and GDLPP at a significance level of 0.05

Table 13 *p*-Values between DLPP and GDLPP on different face databases

Data	ORL(2)	ORL(3)	ORL(4)	Yale(3)	Yale(4)	FERER(2)	FERET(3)	PIE(2)	PIE(3)	PIE(4)
p-value	$1.907 \times 10^{-6*}$	$1.907 \times 10^{-6*}$	$3.814 \times 10^{-6*}$	0.0044^{*}	$7.6294 \times 10^{-5*}$	$1.907 \times 10^{-6*}$	0.5034	$1.907 \times 10^{-6*}$	$1.907 \times 10^{-6*}$	$1.907 \times 10^{-6*}$

The asterisks indicate a statistically significant difference between DLPP and GDLPP at a significance level of 0.05

5. LBP is competitive with LDA, LPP and DLPP methods on ORL, Yale and FERET face databases. The recognition rates of LBP can be still improved by using appropriate weights [14, 15].
6. RGDLPP is slightly slower than PCA, LBP, LPP and DLPP. In fact, RGDLPP also requires more storage than LDA, LPP and DLPP methods, since RGDLPP can extract more features than other methods.

4.5 Evaluation of the experimental results

Is the proposed method statistically better than other methods in terms of its recognition rate? To answer this question, let us evaluate the experimental results in Tables 1–4 using McNemar's significance test [26, 27]. McNemar's test is essentially a null hypothesis statistical test based on a Bernoulli mode. If the resulting *p*-value is below the desired significance level (e.g. 0.05), the null hypothesis is rejected and the performance difference between two algorithms is considered to be statistically significant. The *p*-values between LBP and GDLPP, PCA and GDLPP, LDA and GDLPP, LPP and GDLPP, DLPP and GDLPP on different face databases are reported in Tables 9–13, respectively. Note that the numbers in parentheses denote the sample size.

From Tables 9–13, we can obtain the following conclusions:

1. The proposed GDLPP statistically significantly outperforms LBP in the trials except with two training samples for ORL, three training samples for Yale and two, three training samples for FERET.
2. The proposed GDLPP statistically significantly outperforms PCA in all the experimental cases.
3. The proposed GDLPP statistically significantly outperforms LDA, LPP and DLPP in all the experimental cases except with three training samples for FERET.
4. The proposed GDLPP statistically significantly outperforms all the other compared algorithms on PIE database.

5 Conclusions and future work

In this paper, we proposed the RGDLPP method for face recognition. At first, we replace locality preserving within-class scatter $\boldsymbol{S}_w^L$ in DLPP approach by locality preserving total scatter $\boldsymbol{S}_t^L$. All training samples are projected into the range of $\boldsymbol{S}_t^L$ to reduce dimensionality without loss on discriminative information. Second, to alleviate the problem of unreliable small and zero eigenvalues caused by noise and the limited number of training samples, a method for regularising the small and zero eigenvalues of locality preserving within-class scatter $\boldsymbol{S}_w^L$ is proposed. Experimental results on ORL, Yale, FERET and PIE face databases indicate that RGDLPP performs significantly better than DLPP, LPP, LDA, PCA and LBP methods in terms of recognition accuracy.

However, when there is only one training sample per person available, which is called one sample problem [28–34], the Laplacian matrix $\boldsymbol{L}$ and locality preserving within-class scatter $\boldsymbol{S}_w^L$ are both zero matrices. Then DLPP and our proposed RGDLPP fail to work. We will investigate how to apply DLPP and RGDLPP to one sample problem in the future work.

6 Acknowledgments

This research is supported by NSFC of China (nos 60632050, 60705006, 60873151, 60973098) and the 2010 Graduates' Research Innovation Program of Higher Education of Jiangsu Province (no. 178). The authors would like to thank the anonymous reviewers and the editor for their helpful comments and suggestions. We would to thank one of the anonymous reviewers for pointing out the weakness of the RGDLPP algorithm.

7 References

1 Fukunaga, K.: 'Introduction to statistical pattern recognition' (Academic Press, 1990, 2nd edn.)
2 Duda, R.O., Hart, P.E., Stork, D.G.: 'Pattern classification' (Wiley, 2000, 2nd edn.)
3 Raudys, S.J., Jain, A.K.: 'Small sample size effects in statistical pattern recognition: recommendations for practitioners', *IEEE Trans. Patt. Anal. Mach. Intell.*, 1991, **13**, (3), pp. 252–264
4 Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: 'Eigenfaces vs. Fisherfaces: recognition using class specific linear projection', *IEEE Trans. Patt. Anal. Mach. Intell.*, 1997, **19**, (7), pp. 711–720
5 Yu, H., Yang, J.: 'A direct LDA algorithm for high-dimensional data with application to face recognition', *Patt. Recogn.*, 2000, **33**, (1), pp. 1731–1726
6 Chen, L.F., Liao, H.Y.M., Ko, M.T., Yu, G.J.: 'A new LDA-based face recognition system which can solve the small sample size problem', *Patt. Recogn.*, 2000, **33**, (1), pp. 1713–1726
7 Huang, R., Liu, Q., Lu, H., Ma, S.: 'Solving the small size problem of LDA'. Proc. 16th Int Conf. Pattern Recognition, 2002, pp. 29–32
8 Yang, J., Yang, J.Y.: 'Why can LDA be performed in PCA transformed space?', *Patt. Recogn.*, 2003, **36**, (3), pp. 563–566
9 Lu, J., Plataniotis, K.N., Venetsanopoulos, A.N.: 'Face recognition using LDA-based algorithms', *IEEE Trans. Neural Netw.*, 2003, **14**, (1), pp. 195–200
10 Lotlikar, R., Kothari, R.: 'Fractional-step dimensionality reduction', *IEEE Trans. Patt. Anal. Mach. Intell.*, 2000, **22**, (6), pp. 623–627
11 Dai, D.Q., Yuen, P.C.: 'Regularized discriminant analysis and its application to face recognition', *Patt. Recogn.*, 2003, **36**, pp. 845–847
12 Chen, W.S., Yuen, P.C., Huang, J.: 'Kernel machine-based one-parameter regularized Fisher discriminant method for face recognition', *IEEE Trans. Syst. Man Cybern. – Part B: Cybern.*, 2005, **35**, (4), pp. 659–669
13 Jiang, X., Mandal, B., Kot, A.: 'Eigenfeature regularization and extraction in face recognition', *IEEE Trans. Patt. Anal. Mach. Intell.*, 2008, **30**, (3), pp. 383–394
14 Ahonen, T., Hadid, A., Pietikäinen, M.: 'Face description with local binary patterns: application to face recognition', *IEEE Trans. Patt. Anal. Mach. Intell.*, 2006, **28**, (12), pp. 2037–2041
15 Zhao, G., Pietikäinen, M.: 'Dynamic texture recognition using local binary patterns with an application to facial expressions', *IEEE Trans. Patt. Anal. Mach. Intell.*, 2007, **29**, (6), pp. 915–928
16 Tenenbaum, J.B., Silva, V.D., Langford, J.C.: 'A global geometric framework for nonlinear dimensionality reduction', *Science*, 2000, **290**, pp. 2319–2323
17 Roweis, S.T., Saul, L.K.: 'Nonlinear dimension reduction by locally linear embedding', *Science*, 2000, **290**, pp. 2323–2326
18 Belkin, M., Niyogi, P.: 'Laplacian eigenmaps for dimensionality reduction and data representation', *Neural Comput.*, 2003, **15**, (6), pp. 1373–1396
19 Yan, S., Xu, D., Zhang, B., Zhang, H.-J., Yang, Q., Lin, S.: 'Graph embedding and extensions: a general framework for dimensionality reduction', *IEEE Trans. Patt. Anal. Mach. Intell.*, 2007, **29**, (1), pp. 40–51
20 He, X., Yan, S., Hu, Y., Zhang, H.: 'Learning a locality preserving subspace for visual recognition'. Proc. Ninth Int. Conf. Computer Vision, France, 2003, pp. 385–392
21 He, X., Yan, S., Hu, Y., Niyogi, P., Zhang, H.: 'Face recognition using Laplacian faces', *IEEE Trans. Patt. Anal. Mach. Intell.*, 2005, **27**, (3), pp. 328–340
22 Hu, H.: 'Orthogonal neighborhood preserving discriminant analysis for face recognition', *Patt. Recogn.*, 2008, **412**, pp. 2045–2054
23 Yu, W., Teng, X., Liu, C.: 'Face recognition using discriminant locality preserving projections', *Image Vis. Comput.*, 2006, **24**, pp. 239–248
24 Yang, L., Gong, W., Gu, X., Li, W., Liang, Y.: 'Null space discriminant locality preserving projections for face recognition', *Neurocomputing*, 2008, **71**, pp. 3644–3649
25 Rosipal, R., Girolami, M., Trejo, L.J., Cichocki, A.: 'Kernel PCA for feature extraction and de-noising in nonlinear regression', *Neural Comput. Appl.*, 2001, **10**, (3), pp. 231–243
26 Yambor, W., Draper, B., Beveridge, R.: 'Analyzing PCA-based face recognition algorithms: eigenvectors selection and distance measures'. Empirical Evaluation Methods in Computer Vision, Singapore, 2002, pp. 1–14
27 Draper, B.A., Baek, K., Bartlett, M.S., Beveridge, J.R.: 'Recognizing faces with PCA and ICA', *Comput. Vis. Image Understand.*, 2003, **91**, (1–2), pp. 115–137
28 Tan, X., Chen, S., Zhou, Z.-H., Zhang, F.: 'Face recognition from a single image per person: a survey', *Patt. Recogn.*, 2006, **39**, (9), pp. 1746–1762
29 James, A.P., Dimitrijev, S.: 'Face recognition using local binary decisions', *IEEE Signal Process. Lett.*, 2008, **15**, (11), pp. 821–824
30 Gao, Y., Wang, Y., Feng, X., Zhou, X.: 'Face recognition using most discriminative local and global features'. 18th Int. Conf. Pattern Recognition, ICPR, Hong Kong, 2006, pp. 351–354
31 Tan, X., Triggs, B.: 'Fusing Gabor and LBP feature sets for kernel-based face recognition', *Lect. Notes Comput. Sci.*, 2007, **4778**, pp. 235–249
32 Nguyen, H.V., Bai, L., Shen, L.: 'Local Gabor binary pattern whitened PCA: a novel approach for face recognition from single image per person', *Lect. Notes Comput. Sci.*, 2009, **5558**, pp. 269–278
33 Guo, Y., Xu, Z.: 'Local Gabor phase difference pattern for face recognition'. 19th Int. Conf. Pattern Recognition, Tampa, FL, 2008, pp. 1–4
34 Deng, W., Hu, J., Guo, J., Cai, W., Feng, D.: 'Robust, accurate and efficient face recognition from a single training image: a uniform pursuit approach', *Patt. Recogn.*, 2010, **43**, (5), pp. 1748–1762

doi: 10.1049/iet-cvi.2009.0138