

On the Dimensionality Reduction for Sparse Representation based Face Recognition

Lei Zhang, Meng Yang, Zhizhao Feng and David Zhang

Biometric Research Center, Dept. of Computing, The Hong Kong Polytechnic University

E-mail: {cslzhang, csmyang, cszfeng, csdzhang}@comp.polyu.edu.hk

Abstract

Face recognition (FR) is an active yet challenging topic in computer vision applications. As a powerful tool to represent high dimensional data, recently sparse representation based classification (SRC) has been successfully used for FR. This paper discusses the dimensionality reduction (DR) of face images under the framework of SRC. Although one important merit of SRC is that it is insensitive to DR or feature extraction, a well trained projection matrix can lead to higher FR rate at a lower dimensionality. An SRC oriented unsupervised DR algorithm is proposed in this paper and the experimental results on benchmark face databases demonstrated the improvements brought by the proposed DR algorithm over PCA or random projection based DR under the SRC framework.

1. Introduction

Automatic face recognition (FR) has been, and remains being, one of the most visible and challenging research topics in computer vision, machine learning and biometrics. Although the facial images have a high dimensionality, they usually lie on a lower dimensional subspaces or sub-manifolds. Therefore, subspace learning and manifold learning methods have been dominantly and successfully used in appearance based FR [1-6]. Among various subspace analysis based FR methods, the classical Eigenfaces and Fisherfaces [1-2] algorithms are the most representative ones based on principal component analysis (PCA) and linear discriminant analysis (LDA), respectively. However, PCA and LDA consider only the global scatter of training samples and they fail to reveal the essential data structures nonlinearly embedded in high dimensional space. To overcome these limitations, the manifold learning methods were

proposed by assuming that the data lie on a low dimensional manifold of the high dimensional space [3-4]. The representative manifold learning methods include locality preserving projection (LPP) [5], unsupervised discriminant projection (UDP) [6], etc.

The success of manifold learning implies that the high dimensional face images can be sparsely represented or coded by the representative samples on the manifold. Very recently, an interesting work was reported by Wright *et al.* [7] by using the sparse representation (SR) technique for robust FR. In Wright *et al.*'s pioneer work, the training face images are used as the dictionary of representative samples, and an input testing image is coded as a sparse linear combination of these sample images via l_1 -norm minimization. The results in [7] clearly validated the effectiveness of SR techniques in FR, which can not only lead to high classification accuracy, but also well handle the problem of face occlusion.

The dimensionality of the face images needs to be reduced before applying SRC. In [7], Randomfaces was proposed and compared with Eigenfaces and Fisherfaces for the dimensionality reduction (DR) in SRC based FR. The author claimed that it achieves the best performance with SRC in Yale-B and AR databases. However, Randomfaces, as a type of "universal" feature extractor, does not exploit the information in the specific training database, and hence may not be able to better capture the features of objects than the training based feature extractor such as Eigenfaces.

In this paper, we investigate how to compute the optimal unsupervised DR matrix of the given dataset under the framework of SRC. Compared with PCA, our method minimizes the reconstructed error by imposing sparse constraints on the coefficients. The experimental results demonstrated that our method can achieve better performance than both Eigenfaces and Randomfaces by using SRC as the classifier.

The rest of the paper is organized as follows. Section 2 briefly reviews SRC. Then proposed DR algorithm is presented in Section 3. Section 4 conducts experiments to validate the proposed method and Section 4 concludes the paper.

2. Sparse Representation based Classification for Face Recognition

Denote by $A_i = [s_{i,1}, s_{i,2}, \dots, s_{i,n_i}] \in \mathbb{R}^{m \times n_i}$ the set of training samples of the i^{th} object class, where $s_{i,j}$, $j=1,2,\dots,n_i$, is an m -dimensional vector stretched by the j^{th} sample of the i^{th} class. For a test sample $\mathbf{y} \in \mathbb{R}^m$ from this class, intuitively, \mathbf{y} could be well approximated by the linear combination of the samples in A_i , i.e. $\mathbf{y} \approx \sum_{j=1}^{n_i} \alpha_{i,j} s_{i,j} = A_i \boldsymbol{\alpha}_i$, where $\boldsymbol{\alpha}_i = [\alpha_{i,1}, \alpha_{i,2}, \dots, \alpha_{i,n_i}]^T \in \mathbb{R}^{n_i}$ are the coefficients. Suppose we have K object classes, and let $A = [A_1, A_2, \dots, A_K]$ be the concatenation of the n training samples from all the K classes, where $n = n_1 + n_2 + \dots + n_K$. If we use A to represent the input test image \mathbf{y} , there is $\mathbf{y} = A\boldsymbol{\alpha}$, where $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1; \dots; \boldsymbol{\alpha}_i; \dots; \boldsymbol{\alpha}_K]$. Since \mathbf{y} is from the i^{th} class and $\mathbf{y} = A_i \boldsymbol{\alpha}_i$ holds well, a naturally good solution to $\boldsymbol{\alpha}$ will be that all the coefficients in $\boldsymbol{\alpha}_k$, $k=1,2,\dots,K$ and $k \neq i$, are nearly zero and only the coefficients in $\boldsymbol{\alpha}_i$ have significant values. In other words, the sparse non-zero entries in $\boldsymbol{\alpha}$ can well encode the identity of the test sample \mathbf{y} . This is the essential idea of SR based classification (SRC) in [7].

The SRC algorithm presented in [7] is as follows.

1. Normalize the columns of A to have unit l_2 -norm.

2. Solve the l_1 -minimization problem:

$$\hat{\boldsymbol{\alpha}}_i = \arg \min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \quad \text{s.t.} \quad \|A\boldsymbol{\alpha} - \mathbf{y}\|_2 \leq \varepsilon$$

3. Compute the residuals

$$r_i(\mathbf{y}) = \|\mathbf{y} - A\delta_i(\hat{\boldsymbol{\alpha}}_i)\|_2, \quad \text{for } i = 1, \dots, k,$$

where $\delta_i(\boldsymbol{\alpha}) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the characteristic function which selects the coefficients associated with the i -th class.

4. Output that identity(\mathbf{y}) = argmin $r_i(\mathbf{y})$.

3. Sparse Dimensionality Reduction

The SRC scheme proposed in [7] is a very powerful classifier. As pointed out in [7], one important merit of SRC is that it is insensitive to DR or feature extraction. Even using random projection or direct down-sampling for DR, robust recognition results can also be obtained by SRC. Nonetheless, in low dimension space a training based DR method such as PCA will lead to better FR performance under the framework of SRC. For example, with PCA for DR, a much lower

dimensionality can be used in SRC than random projection while keeping similar FR accuracy. Therefore, it deserves to investigate the optimal DR for SRC based FR. This section presents an unsupervised algorithm to compute the desired projection matrix.

Denote by $\mathbf{z}_k \in \mathbb{R}^{m \times 1}$ the k -th training sample of A and by $D_k = [\mathbf{z}_1, \dots, \mathbf{z}_{k-1}, \mathbf{z}_{k+1}, \dots, \mathbf{z}_n] \in \mathbb{R}^{m \times (n-1)}$ the collection of training samples without the k -th sample, a projection matrix P , whose dimension is $l \times m$, $l < m$, is to be learned. By requiring P is orthogonal, i.e. $PP^T = I$, we define the following objective function to determine P :

$$J_{P, \{\boldsymbol{\beta}_k\}} = \arg \min_{P, \{\boldsymbol{\beta}_k\}} \left\{ \begin{aligned} & \sum_{k=1}^n \left(\|P\mathbf{z}_k - PD_k\boldsymbol{\beta}_k\|_F^2 + \lambda_1 \|\boldsymbol{\beta}_k\|_1 \right) \\ & + \lambda_2 \|A - P^T P A\|_F^2 \end{aligned} \right\} \quad \text{s.t.} \quad PP^T = I \quad (1)$$

where $\boldsymbol{\beta}_k$ is the SR coefficient vector of \mathbf{z}_k over D_k ; λ_1 and λ_2 are scalar parameters. The first and second items on the right side of Eq. (1) are approximation and sparse constraints respectively; and the third item requires that the dataset A can be well reconstructed from the projected subspace by P . Eq. (1) is a joint optimization of projection P and SR coefficients $\{\boldsymbol{\beta}_k\}$. The minimization of J can be implemented by optimizing P and $\{\boldsymbol{\beta}_k\}$ alternatively. The following algorithm describes the procedures.

Step 1. Initialize P . A good initialization of P can be obtained by applying PCA to A . Such an initialization can speed up the convergence of the optimization and lead to a robust solution.

Step 2. Fix P , compute each $\boldsymbol{\beta}_k$. In computing $\boldsymbol{\beta}_k$, the objective function J is reduced to

$$J_{\boldsymbol{\beta}_k} = \arg \min_{\boldsymbol{\beta}_k} \left\{ \|P\mathbf{z}_k - PD_k\boldsymbol{\beta}_k\|_F^2 + \lambda_1 \|\boldsymbol{\beta}_k\|_1 \right\} \quad (2)$$

Some standard convex optimization techniques or the method in [8] can be used to solve $\boldsymbol{\beta}_k$.

Step 3. Fix Z , update P . Now the objective function is reduced to

$$J_P = \arg \min_P \left\{ \begin{aligned} & \sum_{k=1}^n \left(\|P\mathbf{z}_k - PD_k\boldsymbol{\beta}_k\|_F^2 \right) \\ & + \lambda_2 \|A - P^T P A\|_F^2 \end{aligned} \right\} \quad \text{s.t.} \quad PP^T = I \quad (3)$$

Let $\Gamma = [\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_n]$, $\boldsymbol{\gamma}_k = \mathbf{z}_k - D_k \boldsymbol{\beta}_k$, Eq. (3) becomes

$$J_P = \arg \min_P \left\{ \|P\Gamma\|_F^2 + \lambda_2 \|A - P^T P A\|_F^2 \right\} \quad \text{s.t.} \quad PP^T = I \quad (4)$$

There is

$$\begin{aligned} J_P &= \arg \min_P \left\{ \lambda_2 \|A - P^T P A\|_F^2 + \|P\Gamma\|_F^2 \right\} \\ &= \arg \min_P \text{tr} \left\{ \lambda_2 (A^T - A^T P^T P) (A - P^T P A) + P\Gamma\Gamma^T P^T \right\} \\ &= \arg \min_P \text{tr} \left\{ P\Gamma\Gamma^T P^T - \lambda_2 P A A^T P^T + \lambda_2 A^T A \right\} \\ &= \arg \min_P \text{tr} \left\{ P(\Gamma\Gamma^T - \lambda_2 A A^T) P^T + \lambda_2 A^T A \right\} \end{aligned}$$

Since the last term $\lambda_2 A^T A$ has no effect on the solution of P , obviously P can be determined by applying SVD (singular value decomposition) to $(\Gamma\Gamma^T - \lambda_2 AA^T)$, i.e. P is composed by the l eigenvectors associated with the first l smallest eigenvalues of $(\Gamma\Gamma^T - \lambda_2 AA^T)$.

Step 4. Go back to step 2 until J converges or the maximum number iteration is reached. Finally output P .

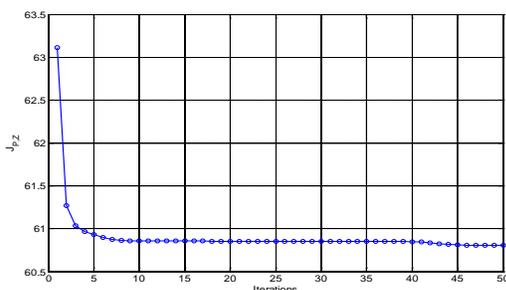


Figure 1. Example of the convergence.

The convergence of the proposed algorithm can be easily proved because in each iteration the objective function J will decrease. Here we give an example. Fig. 1 shows the curve of J versus the iteration number for 240 face images (6 images per subject) in the ORL face database. We see that J converges rapidly.

4. Experimental results

We evaluate the performance of the proposed DR method on three representative facial image databases: ORL, AR [11] and Extended Yale B [9-10]. We compare the proposed DR method with Eigenfaces and Randomfaces, which were employed and discussed in the SRC based FR in [7].

1) *ORL database:* The ORL database (<http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>) contains images from 40 individuals, each providing 10 different images. In the experiment, the images are converted into the size of 56×46 , with the first 6 images of each class for training and the remaining 4 images for testing. The parameters λ_1 and λ_2 in DR learning process and λ in SRC were selected to gain the best performance for each method. In the proposed DR learning process, $\lambda_1=0.05$ and $\lambda_2=0.9$. In the SRC process, $\lambda=0.2, 0.15$ and 0.1 for the proposed method, Eigenfaces and Randomfaces, respectively. The curves of recognition rate versus the dimension of features are illustrated in Fig. 2, and the maximal recognition rate of each method with associated feature dimension is listed in Table 1. From Fig. 2, it can be seen that the proposed DR method achieves the best

rate in all dimensions and Randomfaces is not better than Eigenfaces. In addition, Table 1 shows that proposed method has the highest recognition rate (0.9625) with the least dimension (50).

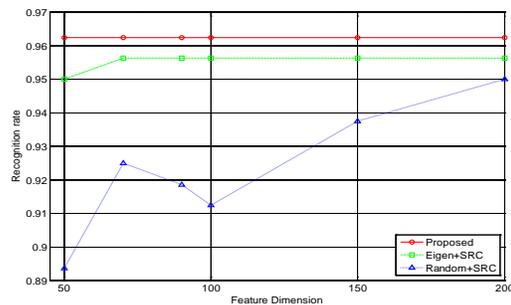


Figure 3. Recognition rates by different methods versus feature dimension on the ORL database.

Table 2. The top recognition rates of different methods on the ORL database and the associated dimension of features.

Feature	Proposed	Eigenfaces	Randomfaces
Rate	0.9625	0.9563	0.9500
Dimension	50	70	200

2) *AR database:* The AR database consists of over 4,000 frontal images from 126 individuals [11]. For each individual, 26 pictures were taken in two separate sessions. In the experiment, we resize the images to the size of 60×43 and choose a subset of the dataset consisting of 50 male subjects and 50 female subjects. For each subject, the seven images with only illumination change and expressions from Session 1 were used for training, and the other seven images with only illumination change and expression from Session 2 were used for testing. Parameters λ in SRC and λ_1 and λ_2 in DR learning process were selected to gain the best performance for each of the competing methods. In the proposed DR learning process, $\lambda_1=0.03$ and $\lambda_2=1.5$. In the SRC process, $\lambda=0.005, 0.007$ and 0.003 for the proposed method, Eigenfaces and Randomfaces, respectively. Fig. 3 demonstrates the curves of recognition rate versus the dimension. The maximal recognition rate of each method with associated feature dimension is listed in Table 2, which shows the proposed DR methods can reach the highest rate (0.9099) with the least dimension (300). From Fig. 3, we can again see that the proposed DR method has the highest recognition rate in the entire dimension while Randomfaces has the lowest recognition rate in the entire dimension.

3) *Extended Yale B Database:* The Extended Yale B database consists of 2,414 frontal-face images of 38 individuals, captured under various laboratory-controlled lighting conditions [9-10]. For each subject,

we randomly selected half of the images for training (i.e. 32 images per subject), and used the other half for testing, where the size of image is converted into 54×48 . In the SRC, $\lambda=0.001$ for eigenfaces and $\lambda=0.005$ for randomfaces with their best performance respectively. In the proposed DR training, we choose $\lambda_1=0.0005$ and $\lambda_2=2$, and in the SRC process we set $\lambda=0.0005$. Fig. 4 shows the three methods' performance in SRC and Table 3 gives their best recognition rates. Similar conclusion to that in the previous experiments can be got from Fig. 4 and Table 3, i.e. the proposed DR method can achieve the highest recognition rate with the least dimension.

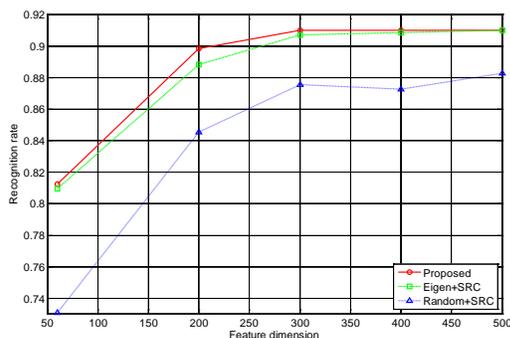


Figure 3. Recognition rates by different features versus feature dimension on the AR database.

Table 2. The top recognition rates of different methods on the AR database and the associated dimension of features.

Feature	Proposed	Eigenfaces	Randomfaces
Rate	0.9099	0.9099	0.8827
Dimension	300	500	500

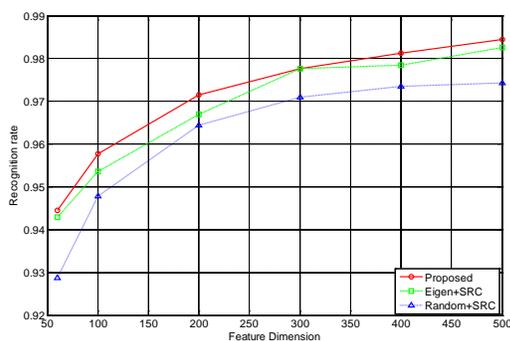


Figure 4. Recognition rates by different features versus feature dimension on the YaleB database.

Table 3. The top recognition rates of different methods on the YaleB database and the associated dimension of features.

Feature	Proposed	Eigenfaces	Randomfaces
Rate	0.9844	0.9826	0.9743
Dimension	500	500	500

5. Conclusion

In this paper, we discussed the dimensionality reduction (DR) of face images when using sparse representation based classifier (SRC) for classification. Our experiments on Extended Yale B, AR and ORL face databases demonstrated that the proposed DR algorithm has better performance than Eigenfaces and Randomfaces. It can achieve higher recognition rate under the same dimensionality than Eigenfaces and Randomfaces. The proposed DR algorithm is an unsupervised learning method. In the future, we will investigate how to introduce the class label information into the DR learning so that a supervised DR algorithm can be developed to learn a set of discriminative projection matrix under the framework of SRC.

Acknowledgments

This paper is supported by the Hong Kong RGC General Research Fund (PolyU5351/08E).

References

- [1] M. Turk and A. Pentland. Eigenfaces for recognition. *J. Cognitive Neuroscience*, 3(1):71-86, 1991.
- [2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *PAMI*, 19(7):711-720, 1997.
- [3] J. B. Tenenbaum, V. deSilva, and J. C. Langford. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*, 290(5500): 2319-2323, 2000.
- [4] S. T. Roweis and L. K. Saul. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*, 290(5500): 2323-2325, 2000.
- [5] X. He, S. Yan, Y. Hu, P. Niyogi, and H. J. Zhang. Face recognition using laplacianfaces. *PAMI*, 27(3): 328-340, 2005.
- [6] J. Yang, D. Zhang, J.-Y. Yang, B. Niu. Globally Maximizing, Locally Minimizing: Unsupervised Discriminant Projection with Applications to Face and Palm Biometrics. *PAMI*, 29(4):650-664, 2007.
- [7] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust Face Recognition via Sparse Representation. *PAMI*, 31(2): 210-227, 2009.
- [8] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. A method for large-scale l_1 -regularized least squares. *IEEE Journal on Selected Topics in Signal Processing*, 1(4):606-617, 2007.
- [9] A. Georghiadis, P. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *PAMI*, 23(6):643-660, 2001.
- [10] K. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *PAMI*, 27(5):684-698, 2005.
- [11] A. M. Martinez and R. Benavente. The AR Face Database. *CVC Technical Report No. 24*, 1998.