Dimension reduction using collaborative representation reconstruction based projections

Juliang Hua [a,b], Huan Wang [a,b,*], Mingwu Ren [a,b], Heyan Huang [a,c]

[a] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
[b] Jiangsu Key Laboratory of Image and Video Understanding for Social Safety, Nanjing University of Science and Technology, Nanjing 210094, China
[c] School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China

ARTICLE INFO

ABSTRACT

This paper develops a collaborative representation reconstruction based projections (CRRP) method for dimension reduction. Collaborative representation based classification (CRC) is much faster than sparse representation based classification (SRC) while owning the similar recognition performance to SRC. Both CRC and SRC utilize the class reconstruction error for classification. First, CRRP characterizes the between-class/within-class reconstruction error using collaborative representation; Second, CRRP seeks the projections by maximizing the between-class reconstruction error to the within-class reconstruction error. So the proposed method is called CRRP. The experimental results on AR, Yale B and CMU PIE face databases demonstrate that CRRP is an effective dimension reduction method.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

The data appears in a high dimensional form in many applications (e.g. image recognition, information retrieval, etc.). Dimension reduction could get the efficient low dimensional representation of these data [1,2], which helps to visualization, classification, calculation and storage. The typical dimension reduction methods could be divided into two categories: unsupervised methods and supervised methods [1,2].

Principal components analysis (PCA) [3] and kernel PCA (KPCA) [4] are the two most popular unsupervised methods. PCA is simple and effective in applications. PCA could not effectively handle the data with nonlinear variants (e.g. illumination, view and noise, etc.). Fisher LDA (FLDA) [5] and Kernel FLDA (KFLDA) [6] are the two most famous supervised methods. FLDA calculates the projection matrix by maximizing the ratio of the between class scatter to the within class scatter. PCA is optimal in reconstruction view and FLDA is optimal in classification view. FLDA often meets small sample size (SSS) problem [7] in applications since the within class scatter matrix is singular. Many FLDA variants [8–14] are given to overcome the SSS problem. KPCA is nonlinear version of PCA via kernel tricks and could handle implicitly deal with the nonlinear data. KFLDA is the non-linear version of FLDA via kernel tricks. J. Yang et al. [15] demonstrate that KFLDA=KPCA+FLDA. It is very difficult to select the kernel

functions and the kernel parameters in kernel methods. Another nonlinear dimension reduction methods are manifold learning based methods (e.g. ISOMAP [16], LLE [17] LE [18], LPP [19], NPE [20], UDP [21], MFA [22], MMDA [23], etc. [24–26]). The manifold learning based methods could explicitly discover the nonlinear manifold structure that embedding in the data. Most of these manifold learning based methods could surpass PCA and LDA on the real applications due to the simplicity and the property to preserve the local manifold structure. These methods could be unified in graph embedding framework [22]. The key problem in graph embedding framework is how to construct the affine graph. It is not clear how to set the neighbor size and how to set the parameters in the graph embedding framework.

Recently, sparse representation [27,28] and collaborative representation [29] have attracted many attentions in pattern recognition and machine learning. Wright et al. [27] developed a sparse representation based classification (SRC) methods. SRC sparsely constructed a test sample by all the training samples and classify the test sample into the class with minimal reconstruction residual. SRC bridged sparse presentation and pattern recognition. Qiao et al. [30] constructed a graph (i.e. L1-norm graph [31]) using sparse representation to develop a sparse preserving projections (SPP) method for dimension reduction. The L1-norm graph could be automatically constructed and does not need to explicitly set the neighbor size and the weight parameter in SPP. Zhang et al. [32] gave a method to simultaneously seek the projection matrix and the sparse representation coefficients (i.e. L1-norm graph), called a graph optimization for dimensionality reduction with

* Corresponding author.
  E-mail address: wanghuanphd@foxmail.com (H. Wang).

sparsity constraints (GODRSC). Yang et al. [33] calculated the projection matrix by maximizing the ratio of the between-class reconstruction residual to the within class reconstruction residual according to SRC classification rules and gave a SRC steered discriminant projections method (SRC-DP) for dimension reduction. There are many sparse related dimension reduction methods: sparse PCA [34], sparse Fisher LDA (SLDA) [35] and sparse tensor discriminant analysis [36,37] and group sparse multiview patch alignment framework (GSM-PAF) [38], etc. Sparse representation is a regularized least square regression with L1-norm constraint. To calculate L1-norm minimizing is very slow and many methods have been developed [39]. These optimized methods for L1-norm may have different performances in recognition tasks. Some researchers study what is the key in SRC [29,40]. Zhang et al. [29] pointed out that collaborative representation (i.e. the least square part) is the key of SRC and gave a collaborative representation based classification (CRC). CRC is a regularized least square regression with L2-norm constraint. Ren et al. [41] proposed a L21-norm-based regression based classification method. Shen et al. [42,43,44] developed a series of regression methods for feature extraction and face recognition. CRC is much faster than SRC and has the similar performance. A nonlinear CRC was given via kernel tricks [45,46]. Yang et al. [47] gave a collaborative representation based projections (CRP) for dimension reduction. CRP utilized collaborative representation to construct the affine graph and sought the projection matrix according to the graph embedding framework. Both CRP and SPP are unsupervised methods. Here, we develop a supervised dimension reduction method using CRC, named collaborative representation reconstruction based projections (CRRP). CRRP utilizes the collaborative representation reconstruction to characterize the between-class scatter and within-class scatter and seeks the optimal projection matrix by maximizing the ratio of the between-class scatter to within-class scatter. CRRP is more consistent with CRC than CRP since CRRP utilizes the classify rules in CRC to design the scatters.

The rest of the paper is organized as follows: in Section 2, CRC is reviewed; in Section 3, CRRP is formulated in detail; in Section 4, the experiments are constructed to evaluate CRRP; in Section 5, the conclusions are given.

2. Collaborative representation based classification

Denote the ith class training sample set by $X_i = [x_{i,1}, x_{i,2}, ..., x_{i,n_i}], x_{i,n_i} \in R^d$, denote the total training sample set by $X = [X_1, X_2, ..., X_c]$. Given a probe y, CRC classify y as follows:

Step1. Calculate the collaborative representation coefficients w of y over X:

$$J(w) = \arg\min_{w} \{\|y - Xw\|_2^2 + \lambda\|w\|_2^2\} \qquad (1)$$

where $\lambda > 0$ is the balance parameter. $w_{i,j}$ is the contribution of $x_{i,j}$ to reconstruct y. w could be efficiently calculated as $w = (X^T X + \lambda I)^{-1} X^T y$ [50].

Step 2. Calculate the regularized residual over each class,

$$r_i = \|y - X\delta_i(w)\|_2 / \|\delta_i(w)\|_2 \qquad (2)$$

where $\delta_i(w)$ is a vector whose only nonzero entries are the entries in w that are corresponding to class i.

Step3. Assign the label of y as

$$\text{label}(y) = \arg\min_{i}\{r_i\} \qquad (3)$$

As pointed in [29], CRC is much faster than SRC while owning a similar performance. More information could be found in [29].

3. Collaborative representation reconstruction based projections

3.1. Formulation

Many dimension reduction methods could be unified in a graph embedding framework [22]. It is the key to construct a specific graph in a graph embedding framework. The traditional graph construction method consists of two steps: (1) select the nearest neighbors as the graph vertices, and (2) set weight between the graph vertices. There are two typical methods k-nearest neighbor based method and r-ball based method to select the nearest neighbors. It is difficult to manually set the neighbor size, the radius and the weight parameters in high dimensional space. Qiao et al. utilized sparse representation to automatically construct L1-norm graph and avoided manually setting the graph. It is slow to optimize L1-norm minimizing. Inspired by the superiority of CRC, we develop a collaborative representation based dimension reduction method, named collaborative representation reconstruction based projections (CRRP).

First, we construct the affine graph using the training samples via collaborative representation. Denote the ith class training sample set by $X_i = [x_{i,1}, x_{i,2}, ..., x_{i,n_i}]$, $x_{i,j} \in R^d$, $n = \sum n_i$ denote the total training sample set by $X = [X_1, X_2, ..., X_c]$. For any training sample $x_{i,j}$, it could be collaborative representated using all the remaining samples by following equation,

$$w_{i,j} = \arg\min\{\|x_{i,j} - Xw_{i,j}\|_2^2 + \lambda\|w_{i,j}\|_2^2\} \qquad (4)$$

where $w_{i,j} = [w_{1,1}, ..., w_{i,j-1}, 0, w_{i,j+1}, ..., w_{c,n_c}]^T \in R^n$ is the collaborative representation reconstruction coefficients. When we get $W = [w_{1,1}, w_{i,2}, ..., w_{c,n_c}]$ for each $x_{i,j}$, the $G(X, W)$ is constructed where the sample set X is the vertices and W is the weight matrix, which could be called L2-norm graph.

Second, we analyze the classification rule of CRC. $X\delta_i(w)$ in Eq. (2) could be seen as the reconstruction of the probe y over class i. So CRC utilizes the residual between the probe and the reconstruction of the probe over each class to classify. The difference between the probe y and the reconstruction of the probe y over the same class of the probe y could be used to characterize the compactness. The difference between the probe y and the reconstruction of the probe y over the different classes of the probe y could be used to characterize the seperability. According to the classification rule of CRC, we design the between-class scatter matrix and the within-class scatter matrix to characterize the separability and the compactness using the reconstruction residual.

We expect that the classification rule of CRC could be preserved in the reduced subspace. Denote the projection matrix by P. We design the between-class scatter and the within-class scatter in the reduced subspace using the classification rule of CRC.

The within-class compactness in the reduced subspace could be defined as:

$$\begin{aligned} J_w &= \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n_i}\|P^T x_{i,j} - P^T X\delta_i(w_{i,j})\|^2 \\ &= tr\left\{P^T\left[\frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n_i}(x_{i,j} - X\delta_i(w_{i,j}))(x_{i,j} - X\delta_i(w_{i,j}))^T\right]P\right\} \\ &= tr\{P^T S_w P\} \end{aligned} \qquad (5)$$

where $S_w = \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n_i}(x_{i,j} - X\delta_i(w_{i,j}))(x_{i,j} - X\delta_i(w_{i,j}))^T$ is the within-class scatter matrix in the original high dimensional space.

The between-class separability in the reduced subspace could be defined as:

$$J_b = \frac{1}{n(c-1)} \sum_{i=1}^{c} \sum_{j=1}^{n_i} \sum_{k \neq i} \|P^T x_{i,j} - P^T X \delta_k(w_{i,j})\|^2$$

$$= tr\left\{ P^T \left[\frac{1}{n(c-1)} \sum_{i=1}^{c} \sum_{j=1}^{n_i} \sum_{k \neq i} \left(P^T x_{i,j} - P^T X \delta_k(w_{i,j}) \right) \right.\right.$$

$$\left.\left. \times \left(P^T x_{i,j} - P^T X \delta_k(w_{i,j}) \right)^T \right] P \right\}$$

$$= tr\left\{ P^T S_b P \right\} \qquad (6)$$

where $s_b = \frac{1}{n(c-1)} \sum_{i=1}^{c} \sum_{j=1}^{n_i} \sum_{k \neq i} (x_{i,j} - X\delta_k(w_{i,j}))(x_{i,j} - X\delta_k(w_{i,j}))^T$ is the between-class scatter matrix in the original high-dimensional space.

According to the classification rule, CRRP expects to maximize the between-class separability and minimize the within-class compactness, i.e

$$\begin{cases} \arg \min_P tr\{P^T S_w P\} \\ \arg \max_P tr\{P^T S_b P\} \end{cases} \qquad (7)$$

Finally, the objective function of CRRP could be defined as:

$$J(P) = \arg \max_P \frac{tr\left\{ P^T S_b P \right\}}{tr\left\{ P^T S_w P \right\}} \qquad (8)$$

Now, we analyze how to calculate the projection matrix P. Obviously, S_b and S_w could be directly computed in the original high dimensional space. If S_w is nonsingular, the projection matrix P consists of the generalized eigenvectors corresponding to the largest eigenvalues of $S_b P = \lambda S_w P$. If S_w is singular, two strategies could be adopted: (1) PCA is utilized to reduce the dimension of the original data so that S_w is nonsingular; and (2) a small disturbing matrix is added into S_w so that S_w is nonsingular.

3.2. Algorithm

The CRRP based dimension reduction algorithm can be listed as follows:

Step 1. Perform PCA to reduce the original sample into a low dimensional subspace. Denote by P_{PCA} the projection matrix of PCA.

Step 2. In the PCA subspace, calculate the collaborative reconstruction coefficients using Eq. (4).

Step 3. Build the between-class scatter matrix and the within-class scatter matrix using Eqs. (5) and (6)

Step 4. Calculate the projection matrix using the eigenvectors $P = [p_1, p_2, ..., p_d]$ of $(S_w)^{-1} S_b$ corresponding to the first d largest nonzero eigenvalues.

Step 5. Get the final projection matrix $P_f = P^T * W_{PCA}^T$.

3.3. Analysis

CRRP belongs to the graph embedding framework. CRRP omits to manually set the neighbor size and the weight parameters in the graph, which significantly increases the application scope. SPP and CRP have the same merit. The main difference between SPP and CRP lies on the graph construction method. SPP built the L1-norm graph using sparse representation. CRP built the L2-norm graph using collaborative representation. It is slow to build L1-norm graph. To calculate L2-norm graph is very fast. The difference between CRRP and CRP is the definition the scatter matrix. CRP defines the scatter matrix using the standard matrix in graph

embedding framework. CRRP utilizes the reconstruction residual to define the scatter matrix, which is consistent with the classification rule of CRC.

SRC-DP aims to simultaneously learn the projection matrix and the sparse representation based affine graph. SRC-DP is iteratively calculated, so SRC-DP is much slower than SPP. CRRP has analytic solution. CRRP could extract more than $c - 1$ (c is the class number) dimensional features.

4. Experiments

The experiments on Yale B, AR and CMU_PIE face databases are performed to compare PCA, LDA, LPP, NPE, SPP, CRP and CRRP. The classification method is the nearest neighbor classifier with cosine distance. The experiments are implemented on Intel i5 2430+Lenovo notebook with 8 G RAM and programmed in the MATLAB language (Version R2013b). Here, we implement the code in all experiments. In all the experiments, the parameters are set by grid search to get the optimal parameter for the best performance for each method.

4.1. Data sets

AR: a subset of the AR face [48,49] is selected. Twenty face images (each session contains 10) of these 120 individuals were selected. These images contains variations of neutral expression, smiling, angry, screaming, left light on, right light on, all sides light on, wearing sun glasses, wearing sun glasses with left light on, and wearing sun glasses with the right light on. The face portion of each image was manually cropped and was resized to 50 by 40 pixels. The example images of one person are shown in Fig. 1(a).

Yale B: the extended Yale B database [50] consists of 2414 images of 38 individuals, with images of each person taken under 64 different controlled lighting conditions. The images are all of frontal pose and are cropped to the actual facial area. All images have a resolution of 192×168 pixels. In the experiments, the image is resized to 48*42. The example images of one person are shown in Fig. 1(b).

CMU PIE: there are 68 subjects with 41368 face images in the CMU PIE face database [51]. The images of each person were taken across 13 different poses, under 43 different illumination conditions, and with 4 different expressions. We choose the subset of PIE database (Pose29), which containing 68 persons and each has 24 face images. All images have been cropped and resized to be 64*64 pixels. The example images of one person are shown in Fig. 1(c).

4.2. Experimental results

We chose the first l images as training set and the rest as testing samples. In the PCA stage, we keep nearly 98 percent image energy and select the number of principal components. Then in PCA subspace do LDA, LPP, NPE, SPP, CRP and CRRP. On the AR database, $l=10$. On the Yale B database, $l=4, 8, 12, 16, 20, 24, 28,$ and 32. On the CMU PIE database, $l=2, 4, 6, 8, 10,$ and 12. The experimental result on AR face database is shown in Table 1. The results on Yale B are shown in Table 2. The results on CMU PIE are shown in Table 3.

Tables 1–3 demonstrate that CRRP has the top performance in most cases. The recognition result vs. the different dimensional feature on AR face database is shown in Fig. 2. From Fig. 2, we can find that CRRP steadily surpass other methods. The influence of the regularized parameter λ in CRRP on AR is shown in Fig. 3. Fig.3 demonstrates that CRRP is robust to variant λ.
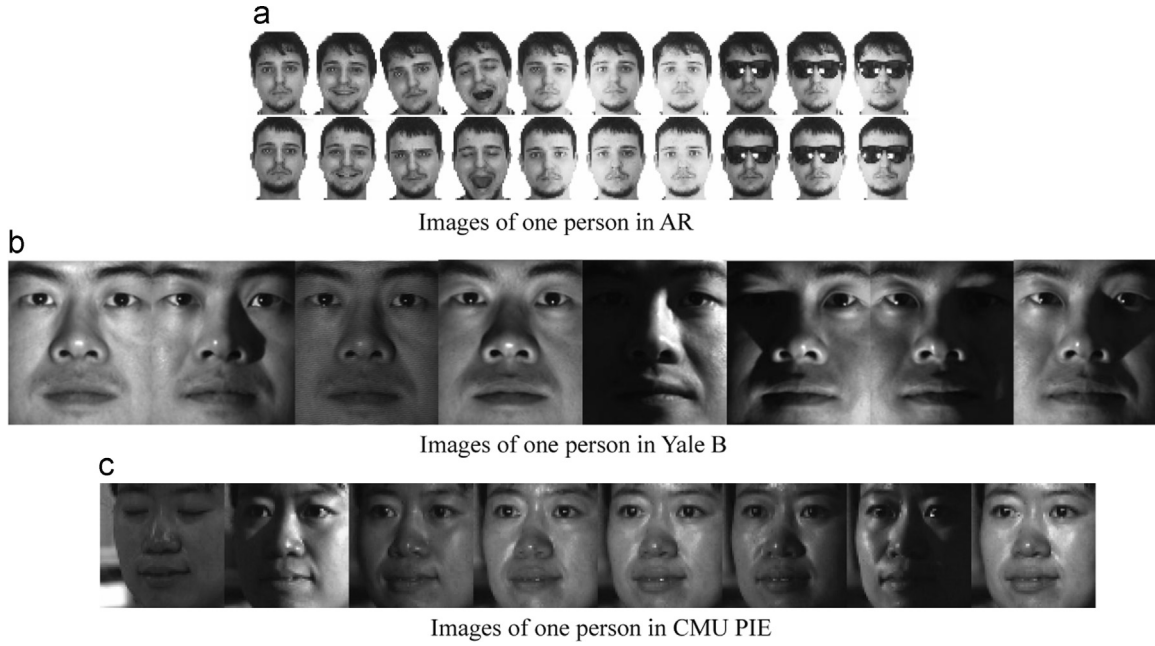
Fig. 1. Some image samples of data set.

Table 1
recognition results on AR face databases by different methods.

	PCA	LDA	LPP	NPE	SPP	CRP	Proposed
Result	0.6367	0.6942	0.5767	0.6542	0.6375	0.7125	**0.7425**

Table 2
recognition results on Yale B face databases by different methods.

	PCA	LDA	LPP	NPE	SPP	CRP	Proposed
$l=32$	0.7969	**0.9539**	0.9021	0.8717	0.9211	0.9293	0.9375
$l=28$	0.6996	0.8501	0.8253	0.8268	0.8692	0.8765	**0.8896**
$l=24$	0.5993	0.8013	0.8059	0.7888	0.8388	0.8408	**0.8533**
$l=20$	0.4671	0.6663	0.7488	0.7243	0.7841	0.7871	**0.7996**
$l=16$	0.4539	0.5795	0.6787	0.6716	0.7336	0.7209	**0.7248**
$l=12$	0.3902	0.4975	0.878	0.6933	0.6285	0.6513	**0.6538**
$l=8$	0.3816	0.5132	0.6250	0.6523	0.5935	**0.6631**	0.6208
$l=4$	0.3789	0.5123	0.5750	0.5140	0.5746	**0.5816**	**0.5816**

Table 3
recognition results on CMU PIE face databases by different methods.

	PCA	LDA	LPP	NPE	SPP	CRP	Proposed
$l=12$	0.5355	0.9118	0.9498	0.8468	0.9706	**0.9853**	0.9804
$l=10$	0.4800	0.7952	0.8498	0.7172	0.8330	**0.8971**	0.8950
$l=8$	0.5074	0.7197	0.6700	0.5588	0.7381	0.7877	**0.8575**
$l=6$	0.2377	0.5408	0.3873	0.3962	0.5523	0.5482	**0.5997**
$l=4$	0.2566	0.3735	0.2140	0.2397	0.4581	0.3956	**0.4699**
$l=2$	0.0936	0.1270	0.0802	0.1043	0.1484	0.1404	**0.2025**

The experimental results show that CRRP is an effective method. The reconstruction sample could be seen as the projection of the probe onto the class space. The distance (i.e. residual) between the probe and the projection (i.e. reconstruction sample) could be seen as the distance between the probe and the class space. It is reasonable to characterize the scatter information using the residual between the probe and the reconstruction sample in CRRP. CRRP fully bridges between CRC and feature extractor.
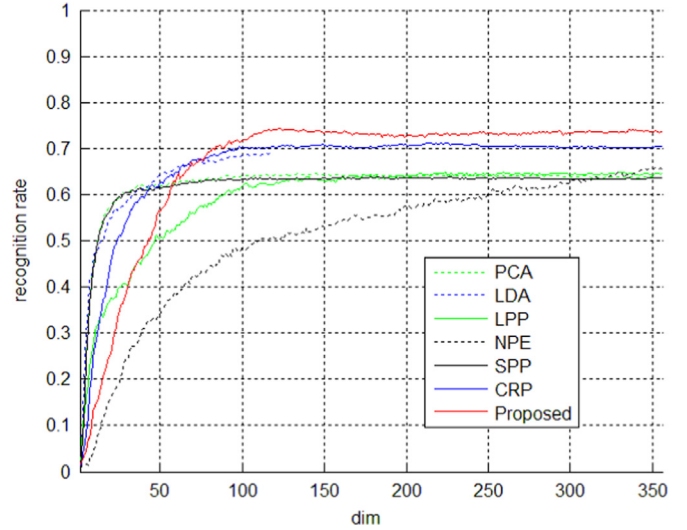

Fig. 2. Recognition rates vs. feature dimensionality on the AR face database.

5. Conclusion

In this paper, we develop a dimension reduction method, named CRRP. CRRP is fully formulated from CRC. First, a L2-norm graph from CRC is built. Second, the between-class separability and the within-class compactness are characterized using the reconstruction residual that is the classification rule of CRC. Third, according to the classification idea, the projection matrix is sought by maximizing the ratio of the between-class separability to the within-class compactness. CRRP fully inherits CRC. CRRP is a supervised method. CRRP is much faster. The experiments on the AR, Yale B and CMUPIE face databases show that CRRP is an effective dimension reduction method. In the future, we will research how to automatically generate the class label and develop an adaptive CRRP.
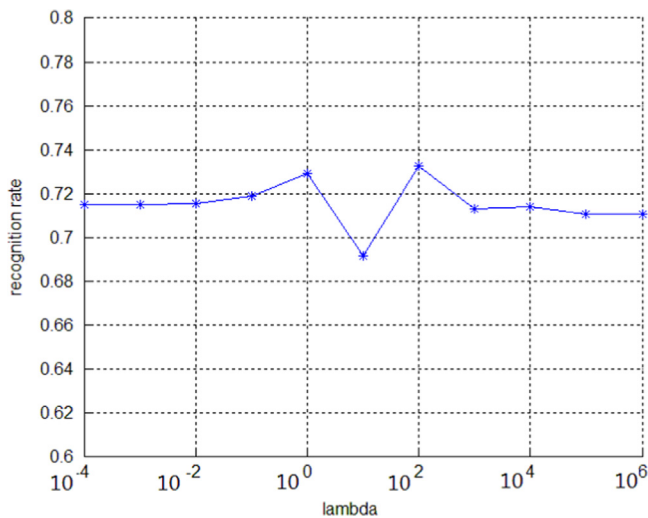
Fig. 3. Recognition rates vs. variant λ on the AR face database.

Acknowledgment

References

[1] A.K. Jain, R.P.W. Duin, J. Mao, Statistical pattern recognition: a review, IEEE Trans. PAMI 22 (1) (2000) 4–37.
[2] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, 2nd ed., John Wiley & Sons, New York, 2000.
[3] M. Turk, A. Pentland, Eigenfaces for recognition, J. Cogn. Neurosci. 3 (1) (1991) 71–86.
[4] B. Schölkopf, A. Smola, K.R. Mü ller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Comput. 10 (5) (1998) 1299–1319.
[5] V. Belhumeur, J. Hespanha, D. Kriegman, Eigenfaces vs fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.
[6] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, and K.-R.Mü ller, "Fisher Discriminant Analysis with Kernels, in: Proc. IEEE Int Workshop Neural Networks for Signal Processing IX, pp. 41–48, Aug. 1999.
[7] S.J. Raudys, A.K. Jain, Small sample size effects in statistical pattern recognition: recommendations for practitioners, IEEE Trans. Pattern Anal. Mach. Intell. 13 (3) (1991) 252–264.
[8] J.H. Friedman, Regularized discriminant analysis, J. Am. Stat. Assoc. 84 (1989) 165–175.
[9] T. Hastie, R. Tibshirani, Penalized discriminant analysis, Ann. Stat. 23 (1995) 73–102.
[10] J. Yang, J. Yang, Why can LDA be performed in PCA transformed space? Pattern Recognit. 36 (2) (2003) 563–566.
[11] L.F. Chen, H.Y.M. Liao, M.T. Ko, G.J. Yu, A new LDA-based face recognition system which can solve the small sample size problem, Pattern Recognit. 33 (1) (2000) 1713–1726.
[12] H. Yu, J. Yang, A direct LDA algorithm for high dimensional data – with application to face recognition, Pattern Recognit. 34 (10) (2001) 2067–2070.
[13] H. Li, T. Jiang, K. Zhang, Efficient and robust feature extraction by maximum margin criterion, IEEE Trans. Neural Netw. 17 (1) (2006) 1157–1165.
[14] W. Yang, J. Wang, M. Ren, J. Yang, Feature extraction based on laplacian bidirectional maximum margin criterion, Pattern Recognit. 42 (11) (2009) 2327–2334.
[15] J. Yang, A.F. Frangi, J.Y. Yang, D. Zhang, KPCA Plus LDA: acomplete Kernel Fisher Discriminant Frame work for Feature Extraction and Recognition, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2) (2005) 230–244.
[16] J.B. Tenenbaum, V. de. Silva, J.C Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (2000) 2319–2323.
[17] S.T. Roweis, L.K. Saul, Nonlinear dimension reduction by locally linear embedding, Science 290 (2000) 2323–2326.
[18] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. 15 (6) (2003) 1373–1396.
[19] X. He, S. Yan, Y. Hu, P. Niyogi, H. Zhang, Face recognition using laplacianfaces, IEEE Trans. Pattern Anal. Mach. Intell. 27 (3) (2005) 328–340.
[20] X. He, D. Cai, S. Yan, and H. Zhang, Neighborhood preserving embedding, in: Proceedings of IEEE Conference on Computer Vision (ICCV2005), 2005, 1208–1213.
[21] J. Yang, D. Zhang, J. Yang, B. Niu, Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics, IEEE Trans. Pattern Anal. Mach. Intell. 29 (4) (2007) 650–664.
[22] S. Yan, D. Xu, B. Zhang, H. Zhang, Graph embedding and extensions: a general framework for dimensionality reduction, IEEE Trans. Pattern Anal. Mach. Intell. 29 (1) (2007) 40–51.
[23] W. Yang, C. Sun, L. Zhang., A multi-manifold discriminant analysis method for image feature extraction, Pattern Recognit. 44 (8) (2011) 1649–1657.
[24] Y. Xu, A. Zhong, J. Yang, D. Zhang, LPP solution schemes for use with face recognition, Pattern Recognit. 43 (12) (2010) 4165–4176.
[25] Q. Gao, J. Liu, K. Cui, H. Zhang, X. Wang., Stable locality sensitive discriminant analysis for image recognition, Neural Netw. 54 (2014) 49–56.
[26] J. Gui, C. Wang, and L. Zhu, Locality preserving discriminant projections," International Conference on Intelligent Computing, 2009, pp.566–572.
[27] J. Wright, Yi Ma, Mairal, J. Sapiro, G. Huang, T.S. Shuicheng Yan, Sparse representation for computer vision and pattern recognition, Proc. IEEE 98 (6) (2010) 1031–1044.
[28] Z. Zhang, Y. Xu, J. Yang, X. Li, D. Zhang, A survey of sparse representation: algorithms and applications, IEEE Access 3 (2015) 490–530.
[29] L. Zhang, M. Yang, X. Feng. Sparse representation or collaborative representation: which helps face recognition. ICCV2011, 2011.
[30] L. Qiao, S. Chen, X. Tan, Sparsity preserving projections with applications to face recognition, Pattern Recognit. 43 (1) (2010) 331–341.
[31] B. Cheng, J. Yang, S. Yan, Y. Fu, T. Huang, Learning with L1-graph for image analysis, IEEE Trans. Image Process. 19 (4) (2010) 858–866.
[32] L. Zhang, S. Chen, L. Qiao, Graph optimization for dimensionality reduction with sparsity constraints, Pattern Recognit. 45 (2012) 1205–1210.
[33] J. Yang, D. Chu, Y. Xu, J. Yang, Sparse representation classifier steered discriminative projection with applications to face recognition, IEEE Trans. NNLS 24 (7) (2013) 1023–1035.
[34] H. Zou, T. Hastie, R. Tibshirani, Sparse principal component analysis, J. Comput. Graph. Stat. 15 (2) (2006) 265–286.
[35] L. Clemmensen, T. Hastie, D. Witten, B. Ersboll, Sparse discriminant analysis, Technometrics 53 (4) (2011) 406–413.
[36] Z. Lai, Y. Xu, J. Yang, J. Tang, D. Zhang, Sparse tensor discriminant analysis, IEEE Trans. IP 22 (10) (2013) 3904–3915.
[37] S. Wang, J. Yang, M. Sun, X. Peng, M. Sun, C. Zhou, Sparse tensor discriminant color space for face verification, IEEE Trans. NNLS 23 (6) (2012) 76–888.
[38] J. Gui, D. Tao, Z. Sun, Y. Luo, X. You, Y.Y. Tang, Group sparse multiview patch alignment framework with view consistency for image classification, IEEE Trans. Image Process. 23 (7) (2014) 3126–3137.
[39] Allen Y. Yang, S. Shankar Sastry, Arvind Ganesh, Yi. Ma. Fast L1-minimization algorithms and an application in robust face recognition: a review. ICIP2010.
[40] Q. Shi, A. Eriksson, A. Hengel, C. Shen. Is face recognition really a Compressive Sensing problem? in: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 11), Colorado Springs, USA, 21–23 June 2011.
[41] C. Ren, D. Dai, H. Yan, Robust classification using L2,1-norm based regression model, Pattern Recognit. 45 (7) (2012) 2708–2718.
[42] F. Shen, W. Yang, H. Li, H. Zhang, H. Shen, Robust regression based face recognition with fast outlier removal, Multimed. Tools Appl. (2014) 1–12.
[43] F. Shen, C. Shen, R. Hill, A. Hengel, Z. Tang, Fast approximate L∞ minimization: speeding up robust regression, Comput. Stat. Data Anal. 77 (2014) 25–37.
[44] F. Shen, C. Shen, A. Hengel, Z. Tang, Approximate least trimmed sum of squares fitting and applications in image analysis, IEEE Trans. IP 22 (5) (2013) 1836–1847.
[45] W. Yang, Z. Wang, J. Yin, C. Sun, K. Ricanek, Image classification using Kernel collaborative representation with regularized least square, Appl. Math. Comput. 222 (2013) 13–28.
[46] D. Wang, H. Lu, M. Yang, Kernel collaborative face recognition, Pattern Recognit. 48 (10) (2015) 3025–3237.
[47] W. Yang, Z. Wang, C. Sun, A collaborative representation based projections method for feature extraction, Pattern Recognit. 48 (1) (2015) 20–27.
[48] A.M. Martinez, R. Benavente, The AR Face Database ⟨http://cobweb.ecn.pur due.edu/~aleix/ aleix_face_DB.html⟩.
[49] A.M. Martinez and R. Benavente, The AR Face Database, CVC Technical Report #24, June 1998.
[50] K. Lee, J. Ho, D. Kriegman., Acquiring linear subspaces for face recognition under variable lighting, IEEE Trans. Pattern Anal. Mach. Intell. 27 (5) (2005) 684–698.
[51] [49] T. Sim, S. Baker, M. Bsat., The CMU pose, illumination, and expression database, IEEE Trans. Pattern Anal. Mach. Intell. 25 (12) (2003) 1615–1618.

Juliang Hua received his B.S. and M.S. degrees at Nanjing University of Science and Technology (NUST), PR China, 1995 and 1999, respectively. Now he is a Ph.D. student at School of Computer Science and Technology, Nanjing University of Science and Technology. His research interests include pattern recognition, computer vision.

Huan Wang is an assistant professor at the School of Computer Science and Technology, Nanjing University of Science and Technology, PR China. He received Ph.D. degree in Pattern Recognition and Intelligent Systems from Nanjing University of Science and Technology, PR China, 2009. His research interests include computer vision and pattern recognition.

Mingwu Ren is a professor at the School of Computer Science and Technology, Nanjing University of Science and Technology, PR China. He received M.S. and Ph.D. degrees in Pattern Recognition and Intelligent Systems from Nanjing University of Science and Technology, PR China. His research interests include computer vision, digital image processing, and pattern recognition.

Heyan Huang is a professor at the School of Computer Science and Technology, Beijing Institute of Technology, PR China, a professor at the School of Computer Science and Technology, Nanjing University of Science and Technology, PR China. Her research interests include natural language processing and pattern recognition.