Pattern Recognition 48 (2015) 20–27

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

A collaborative representation based projections method for feature extraction

Wankou Yang [a,b,c,*], Zhenyu Wang [a,b], Changyin Sun [a,b]

[a] School of Automation, Southeast University, Nanjing 210096, China
[b] Key Lab of Measurement and Control of Complex Systems of Engineering, Ministry of Education, Southeast University, Nanjing 210096, China
[c] Jiangsu Key Laboratory of Image and Video Understanding for Social Safety, Nanjing University of Science and Technology, Nanjing 210094, China

A R T I C L E I N F O

Article history:
Received 18 October 2013
Received in revised form
27 May 2014
Accepted 4 July 2014
Available online 23 July 2014

Keywords:
Sparse representation
Collaborative representation
Feature extraction
Image recognition

A B S T R A C T

In graph embedding based methods, we usually need to manually choose the nearest neighbors and then compute the edge weights using the nearest neighbors via L2 norm (e.g. LLE). It is difficult and unstable to manually choose the nearest neighbors in high dimensional space. So how to automatically construct a graph is very important. In this paper, first, we give a L2-graph like L1-graph. L2-graph calculates the edge weights using the total samples, avoiding manually choosing the nearest neighbors; second, a L2-graph based feature extraction method is presented, called collaborative representation based projections (CRP). Like SPP, CRP aims to preserve the collaborative representation based reconstruction relationship of data. CRP utilizes a L2 norm graph to characterize the local compactness information. CRP maximizes the ratio between the total separability information and the local compactness information to seek the optimal projection matrix. CRP is much faster than SPP since CRP calculates the objective function with L2 norm while SPP calculate the objective function with L1 norm. Experimental results on FERET, AR, Yale face databases and the PolyU finger-knuckle-print database demonstrate that CRP works well in feature extraction and leads to a good recognition performance.

1. Introduction

Recently, sparse representation is very hot in computer vision and pattern recognition [1]. It is widely applied to image super-resolution [2], motion segmentation [3] and supervised denoising [4]. Wright et al. [5] developed a sparse representation based classification (SRC) and got promising results in face recognition. In SRC, a testing sample was coded by sparse linear combination of all the training samples and classified into the class with minimum sparse reconstruction residual. Xu et al. [6] presented a two-phase test sample sparse representation method. He et al. [7] gave a two-stage sparse representation for robust recognition on large-scale databases. He et al. [8,9] proposed iteratively robust sparse representation methods for pattern recognition. Yang et al. [10] presented a robust sparse coding method for face recognition. All of these sparsity based methods get good performance in real-world face recognition. In sparse representation, the dimensionality of the dictionary is always smaller than the number of elements of the dictionary to get a sparse solution. To deal with small sample size problem like face recognition, the dimensionality of the image is larger than the training sample size, the dimensionality reduction (i.e. feature extraction) is necessary before implementing SRC.

Dimensionality reduction is to get a meaningful low-dimensional representation of high dimensional data. Since there are large volumes of high-dimensional data in numerous real-world applications, dimensionality reduction is a fundamental problem in many scientific fields such as visualization, computer vision and pattern recognition. With respect to pattern recognition, dimensionality reduction is an effective approach to reveal the distinctive features from the original data for pattern matching [11]. Many dimensionality reduction algorithms have been developed in the past decades and the most popular techniques include principal component analysis (PCA) [11] and linear discriminant analysis (LDA) [1]. Many improved PCA and LDA are proposed in the past several decades [12–34].

Many studies have shown that the face images possibly reside on a nonlinear sub-manifold [35–37]. Many manifold learning algorithms have been proposed to discover the intrinsic low-dimensional embedding of the original data, among which the most well-known ones are isometric feature mapping (ISOMAP) [35], local linear embedding (LLE) [36], and Laplacian Eigenmap [37]. Experiments have shown that these methods can find perceptually meaningful

* Corresponding author.

http://dx.doi.org/10.1016/j.patcog.2014.07.009
0031-3203/© 2014 Elsevier Ltd. All rights reserved.

embedding for artificial and real-world datasets such as facial or digit images. However, how to evaluate the maps remains unclear. He et al. [38] developed the Locality Preserving Projections (LPP), which was a linear subspace learning method derived from Laplacian Eigenmap. The objective function of LPP is to minimize the local scatter of the projected data. In contrast to most manifold learning algorithms, LPP possesses a remarkable advantage that it can generate an explicit map, which is linear and can be easily computed like PCA and LDA. Yang et al. [39] gave an Unsupervised Discriminant Projection (UDP) technique for dimensionality reduction. Xu et al. [40] proposed some solutions to overcome SSS problem when LPP is applied in face recognition. Neither LPP nor UDP uses the class label information and they are un-supervised methods in nature. In [41], Yan et al. proposed the Marginal Fisher Analysis (MFA) for feature extraction and recognition by combining locality and class label information to represent the intra-class compactness and interclass separability. MFA takes advantage of the partial structural information of classes and neighborhoods of samples; however, it is difficult to decide the number of nearest neighbors of each sample and the number of shortest pairs from different classes in MFA. Wang et al. [42] proposed a locality-preserved maximum information projection (LPMIP) method. LPMIP and UDP are very similar. All these methods can be unified in graph embedding framework [41]. The key point in graph embedding framework is to construct the graph. When construct a graph, typical graph embedding based methods consist of two steps: (1) manually choose the nearest neighbors (i.e. choose the nearest neighbor number, ball radius size and Gaussian-kernel parameter); (2) compute the edge weights using all the nearest neighbors via L2 norm. But it is difficult and unstable to manually choose the nearest neighbors. How to automatically construct a graph is important.

Recently, some feature extraction methods based on the idea of SRC are presented. Qiao et al. [43] gave a sparsity preserving projections (SPP) for dimensionality reduction. In SPP, the graph (i.e. reconstruction relationships) was automatically constructed based on a modified sparse representation. Chen et al. [44] also gave a L1-graph construction method. Their ideas are very similar. Zhang et al. [45] gave a graph optimization for dimensionality reduction with sparsity constraints (GODRSC). GODRSC aims to simultaneously seek the sparse representation coefficients and the projection matrix. GODRSC can be seen as extension of SPP. Yang et al. [46] gave a SRC steered discriminative projections (SRC-DP) method, which seeks a linear transformation such that in the transformed low dimensional space, to simultaneously minimizing the within-class reconstruction residual and maximizing the between-class reconstruction residual. Clemmensen et al. [47] gave a sparse linear discriminant analysis (SLDA), which impose a sparseness constraint on projection vectors. The sparse projection vector yields a set of interpretable feature for classification. Lai et al. [48,49] gave a sparse 2D local discriminant projections (S2DLDP).

The L1-norm minimization makes the sparsity based classification schemes such as SRC very expensive. So both SPP and L1 norm graph are slow. Though SRC emphasizes much on the role of L1 norm sparsity of the representation coefficients, it has been shown in [50] that the collaborative representation mechanism is more important to the success of SRC. And then a collaborative representation based classification (CRC) scheme is given for image recognition. CRC had significantly less complexity than SRC but leaded to very competitive classification performance. CRC coded a testing sample by linear combination of all the training samples with regularized least square and classified the testing sample into the class with the minimum reconstruction error. The main difference between SRC and CRC is the regularized term.

Although SPP gets encouraging results on face recognition tasks, SPP has following weakness: (1) SPP is very slow due to

SPP 的缺点

optimize the objective function with L1 norm is slow, (2) like many manifold learning methods (e.g. LPP and NPE), SPP mainly characterizes “locality” like LPP, which is not enough for classification. L2 的优点

To overcome these shortcomings, first, a L2 norm graph is given. The L2 norm graph could be constructed in one step and avoid manually choosing the nearest neighbors. L2 norm graph calculate the edge weights by using all the samples to represent the pointed sample, which is collaborative representation in essence; Second, a L2 norm graph based feature extraction method is presented, called collaborative representation based projections (CRP). In CRP, for classification, we consider two quantities: local information and global information in the modeling process. The local quantity is characterized via the L2 norm graph. The global quantity is simply characterized by the total scatter matrix like PCA. Based on the local and global quantities, we give a criterion which maximizes the global scatter and minimizes the local scatter simultaneously. The proposed criterion, similar to the classical Fisher criterion, is a Rayleigh quotient form and can be calculated via generalized eigenvalue decomposition.

The rest of the paper is organized as follows: In Section 2, we briefly review SPP. In Section 3, we describe CRP in detail. In Section 4, experiments are conducted to evaluate the proposed method. Conclusions are given in Section 5.

2. SPP

SPP builds the sparse weight matrix W (i.e. the L1-graph) based on a modified sparse representation. Denote by $X=[x_1,x_2,\ldots,x_n]$ the total training samples, where $x_i(i=1,2,\ldots,n)$ is a m-dimensional vector stretched by the ith training samples. The sparse reconstruction weight w_i for of each x_i is obtained through the following objective function:

$$\min ||w_i|| \quad \text{s.t. } x_i = Xw_i \text{ and } 1 = e^T w_i \tag{1}$$

where $w_i=[w_{i,1},\ldots,w_{i,j-1},0,w_{i,j+1},\ldots,w_{i,n}]^T \in R^n$, $w_{i,j}(i \neq j)$ denotes the contribution of each x_j to reconstruct x_i. $e=[1,1,\ldots,1]^T \in R^n$.

The objective function of SPP is as follows:

$$\min \frac{P^T X(I-W-W^T+W^T W)X^T P}{P^T XX^T P} \propto \max \frac{P^T X(W+W^T-W^T W)X^T P}{P^T XX^T P} \tag{2}$$

The optimal projection matrix P is the generalized eigenvectors corresponding to the largest eigenvalues of $X(W+W-W^T W)X^T P=\lambda XX^T P$.

3. Our proposed work

3.1. Motivation

Many subspace learning methods (e.g. PCA, LDA, LPP, NPE and UDP) could be unified in a graph embedding framework [41]. The key in the graph embedding framework is to connect vertices and set graph edge weights. The typical graphs are constructed in two steps: choose the nearest neighbors and set the edge weights. There are two classical methods for graph construction. One is k-nearest-neighbors method, and the other is r-ball based method, where, for each datum, the samples within its surrounding r-ball are connected, and then various approaches, e.g. binary, Gaussian-kernel [38] and L2-reconstruction [36], can be used to further set the graph edge weights. As we know, it is very difficult and unstable to choose the nearest neighbors in high-dimensional space.

SPP has the following drawbacks. The objective function of SPP consists of two parts: the numerator $P^T X(I-W-W^T+W^T W)X^T P$ which aims to mainly characterize the local quantity and the denominator P^TXX^TP which aims to avoid degenerate solutions. In SPP, the weight matrix W is calculated by sparse representation with L1 norm (i.e. L1-graph [38]). L1-graph is constructed via SRC. SRC represents a sample using the total training samples and the most representation coefficients are zero due to the sparse constraint. So L1-graph is like the graph in LPP and L1-graph mainly characterizes the locality information like LPP. L2-graph also mainly characterizes the locality information like L1-graph.

The advantages of L1-graph include: (1) great robustness to data noise, (2) automatic sparsity, and (3) adaptive neighborhood for individual datum. Many optimization methods are proposed to solve these objective functions with L1 norm. Although many algorithms are proposed, to calculate the objective functions with L1 norm is still slow and the performances of different optimization algorithms are different.

To address the aforesaid shortcomings, we give a feature extraction method. First, L2-graph is given. L2-graph could be constructed in one step and avoids choosing the nearest neighbors. In L2-graph, we represent the pointed sample using the samples from all the classes instead of the nearest neighbors (i.e. collaborative representation). Second, based on L2-graph, a collaborative representation based projections (CRP) method is proposed for feature extraction. For classification tasks, it is necessary to simultaneously maximize the nonlocal quantity (or the total quantity) and minimize the local quantity. In SPP, P^TXX^TP could be also seen as the term to characterize the global quantity between each point and the origin. A reasonable method is to characterize the global quantity according to the distance between each point and the mean. So we can choose the total scatter to characterize the total quantity in CRP via the total scatter matrix like PCA. In the following, we formulate CRP.

3.2. Fundamentals

First, we construct graph based on collaborative representation. And the resulted graph could be called L2-graph.

In the following, we formulate L2-graph. Denote by $X=[x_1, x_2, \ldots, x_n]$ the total training samples, where $x_i (i=1,2,\ldots,n)$ is a m-dimensional vector stretched by the ith training samples. L2 norm based reconstruction weight w_i for of each x_i is obtained through the following objective function:

$$w_i = \arg\min \{\|x_i - Xw_i\|_2^2 + \lambda \|w_i\|^q\} \tag{3}$$

where $w_i = [w_{i,1}, \ldots, w_{i,j-1}, 0, w_{i,j+1}, \ldots, w_{i,n}]^T \in R^n$, $w_{ij} (i \neq j)$ denotes the contribution of each x_j to reconstruct x_i. w_i could be efficiently calculated [50].

After computing w_i for each $x_i (i=1,2,\ldots,n)$, we can construct the $G(X, W)$ as the graph with the sample set X as the vertices and W as the weight matrix. When $q=1$, our constructed graph is L1-graph. When $q=2$, our constructed graph is L2-graph.

L2-graph characterizes the reconstruction relationships with weaker sparse constrains than L1-graph. It is nature that the sparse reconstruction weights are expected to be preserved in the low dimensional space. According to the ideas of CRC (i.e. to minimize the reconstruction residual) and classification, in the CRP subspace, we aim to: (1) preserve the sparsity reconstruction relationship (i.e. minimize the local compactness); (2) maximize the total separability information.

According to the L2-graph, the local compactness could be defined by

$$J_L = \sum_{i=1}^{n} \left\| P^T x_i - \sum_{i=1}^{N} w_{ij} P^T x_j \right\|^2 = P^T S_L P \tag{4}$$

where $S_L = X(I - W - W^T + WW^T)X^T$ is the local scatter matrix.

The total separability could be defined by

$$J_T = \sum_{i=1}^{n} \|P^T x_i - P^T \tilde{x}\|^2 = P^T S_T P \tag{5}$$

where $S_T = \sum_{i=1}^{n}(x_i - \tilde{x})(x_i - \tilde{x})^T$ is the total scatter matrix of the data.

For classification, it is natural to simultaneously maximize the total sparability and minimize the local compactness like Fisher discriminant analysis. So the final objective function of CRP is defined as:

$$J(P) = \arg\min_P \frac{P^T S_L P}{P^T S_T P} = \arg\max_P \frac{P^T S_T P}{P^T S_L P} \tag{6}$$

The projection matrix P is the generalized eigenvectors corresponding to the largest eigenvalues of $S_T P = \lambda S_L P$. In small sample size cases, S_L is often singular because the training sample size is smaller than the dimensionality of the image vector space. We can utilize PCA to reduce the dimensionality of the original image vectors so that S_L is nonsingular in the PCA-transformed subspace. CRP is then used for feature extraction.

3.3. The algorithm

The proposed CRP based feature extraction algorithm can be summarized as follows:

Step 1. Use PCA to transform the original image into a lower dimensional subspace. Denote by P_{PCA} the transformation matrix of PCA.
Step 2. In the PCA subspace, construct the reconstruction coefficient matrix W using L2-graph (Eq. (3)).
Step 3. Construct the local scatter matrix S_L using Eq. (4) and the total scatter matrix S_T using Eq. (5), and then calculate the eigenvectors $P=[p_1, p_2, \ldots, p_d]$ of $(S_L)^{-1}S_T$ corresponding to the first d largest nonzero eigenvalues.
Step 4. The final projection matrix is $W = P^T * W_{\text{PCA}}^T$.

3.4. Analysis

Graph embedding framework unified many subspace learning methods (e.g. PCA, LDA, LPP et al.). Graph embedding framework based dimensionality reduction transforms into how to construct the graph. The typical methods are k-nearest-neighbor and r-ball (e.g. LPP, MFA and UDP). The two graphs in MFA are constructed based on the nearest neighbors in the same class and between classes, respectively. The issues of the classical graph construction methods are that the parameters cannot be adaptively set and are sensitive to the noise and outliers. With the wide research of sparse representation, L1-graph is given. L1-graph could automatically choose the neighbors and are data adaptive. Although many optimization algorithms are provided to increase the speed of solving L1 norm, the speed of constructing L1-graph is still slow and L1-graphs produced by different optimization algorithms have different performances. Both L1-graph and L2-graph utilize all the samples to calculate each edge weight. L1-graph is constructed via SRC. SRC represents a sample using the total training samples, but the most representation coefficients are zero in theory. So L1-graph is like the graph in LPP and NPE. L1-graph mainly characterizes the locality information like LPP. So L2-graph also mainly characterizes the locality information like L1-graph. L2-graph has less sparsity constraint than L1-graph. According to CRC, the key in SRC is the collaborative representation. Our L2-graph inherits some merits of L1-graph, such as datum-adaptive, robust to noise and outliers, and sparsity. Specially, to construct L2-graph is much faster than to construct L1-graph.

图嵌入方法的缺点
1. 有参数
2. 对噪声敏感

SPP is based on L1-graph and aims to preserve the sparsity coefficients in the low dimensional space. Although SPP has many merits, SPP is very slow. SPP aims to preserve the sparsity reconstruction relationship. According to the graph embedding framework, SPP seeks to find a linear subspace which can preserve the local structure of data like LPP. The objective of SPP is actually to minimize the local scatter. Obviously, SPP mainly considers the local information. Therefore, SPP does not necessarily yield a good projection suitable for classification. In other words, SPP uses the distance between samples and original to characterize the global information like LPP, which is not suitable for classification.

LPP and LPMIP are unsupervised subspace learning methods. The objective of LPP is to minimize the local scatter of the projected data. In some case, this criterion cannot guarantee to yield a good projection matrix for classification purpose. The objective of LPMIP is to maximize the difference between the nonlocal scatter of the projected data and the local scatter of the projected data. The class label information is ignored in LPP and LPMIP, which is important for recognition and classification. LPP and LPMIP are unsupervised methods.

Our CRP is under graph embedding framework. CRP not only inherits many merits of SPP, but also is much faster than SPP. CRP aims to preserve the reconstruction coefficients by minimizing the local compactness while maximizing the total separability, which is consist with classification tasks. In a word, CRP uses the L2-graph to characterize the local scatter information and the total scatter matrix to characterize the between-class scatter information, seeks the projections by simultaneously maximizing the total separability and minimizing the local compactness, which is consistent with the ideas of collaborative representation and classification. Although SPP and CRP also ignore the class label information, the sparsity characteristic of SPP and CRP implicitly contains supervised information.

4. Experiments

We conduct experiments on FERET, AR, Yale face databases and the PolyU finger-knuckle-print (FKP) database to evaluate PCA, LDA, LPP, LPCA [52], LPMIP, NPE [51], SPP, LPMIP and CRP. A nearest neighbor classifier with cosine distance is employed. Justin Romberg's l1qc_logbarrier is chosen to calculate SPP with L1 norm. In the experiments, the regularized parameter λ in SPP and CRP is determined by grid search. In LPP and LPMIP, the nearest neighbor number is set as *p-1* (*p* is the number of training samples of per class). Here NPE is the supervised version. The experiments are implemented on Intel i5 2430+Lenovo notebook with 8 G RAM and programmed in the MATLAB language (Version R2011b).

4.1. Experiments on the FERET database

The FERET face image database is a result of the FERET program, which was sponsored by the US Department of Defense through the DARPA Program [53,54]. It has become a standard database for testing and evaluating state-of-the-art face recognition algorithms.

The proposed algorithm was tested on a subset of the FERET database. This subset includes 1400 images of 200 individuals (each individual has seven images). This subset has variations of facial expression, illumination, and pose. In our experiment, the facial portion of each original image was automatically cropped based on the location of eyes and the cropped images were resized to 40 by 40 pixels. Some example images of one person are shown in Fig. 1.

Fig. 1. Images of one person in FERET.

In the first experiment, we use the first 6 images per class for training and the remaining one image for testing. In PCA stage of LPP, NPE, LPMIP, SPP and CRP, we keep nearly 98 percent image energy and select the number of principal components $m=289$. The final recognition rates are given in Table 1. The regularized

Table 1
Recognition rates on the FERET database by different methods.

Method	Recognition rate	Dimensionality
PCA	*0.2950*	*289*
LDA	*0.4900*	*199*
LPP	*0.4050*	*269*
NPE	*0.7800*	*276*
LPCA	*0.7600*	*172*
LPMIP	*0.3050*	*127*
SPP($\lambda=10^9$)	*0.7700*	*279*
Proposed($\lambda=10^4$)	***0.7900***	**281**

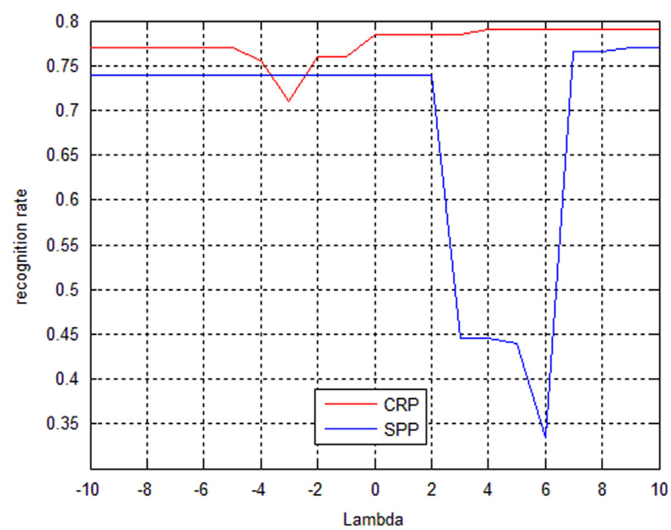

Fig. 2. Recognition rates vs. Lambda on FERET database.

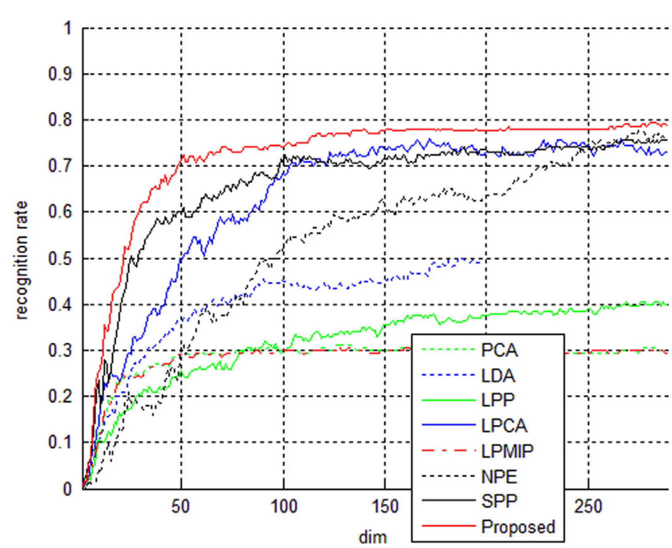

Fig. 3. Recognition rates vs. feature dimensionality on FERET database.

parameter λ in SRC and CRC could impact the performance of L1-graph and L2-graph. In the experiment, we show the influence of λ in Fig. 2. Fig. 2 shows that CRP is more robust than SPP with different λ. Fig. 3 shows the relationship between recognition rates and feature dimensionality.

We further perform experiments by 50-run tests. In each run, we randomly choose 6 images from each class for training, and the remaining images for test. The average recognition results are shown in Table 2. The average experimental time of SPP and CRP is shown in Table 3. Table 3 shows that CRP is much faster than SPP.

4.2. Experiments on the AR database

The AR face [55,56] contains over 4000 color face images of 126 people (70 men and 56 women), including frontal views of faces with different facial expressions, lighting conditions, and occlusions. The pictures of 120 individuals (65 men and 55 women) were taken in two sessions (separated by two weeks) and each section contains 13 color images. Twenty face images (each session contains 10) of these 120 individuals were selected and used in our experiments. These images have 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. 4.

Table 2
The average recognition rates of competing methods on FERET database.

Method	Mean	Std
PCA	*0.6620*	*0.1579*
LDA	*0.8087*	*0.1690*
LPP	*0.6529*	*0.1442*
NPE	*0.6982*	*0.1756*
LPCA	*0.7499*	*0.1144*
LPMIP	*0.6603*	*0.1609*
SPP	*0.7471*	*0.1339*
Proposed	***0.8118***	***0.0862***

Table 3
The average experimental time of SPP and CRP on FERET database (s).

	SPP	*CRP*
Mean	*1289.7*	*478.4*
Std	*52.6*	*22.4*

In the experiment, the first 10 images per class were used for training and the remaining images for testing. In the PCA phase of LDA, LPP, NPE, SPP and CRP, we selected the first 356 principal components. The final recognition rate of each method and the corresponding dimension are given in Table 4. The relationship between recognition rates and feature dimensionality is shown in Fig. 5. We further perform experiments by 50-run tests. In each run, we randomly choose 6 images from each class for training, and the remaining images for test. Based on the optimal feature dimensionality in Table 4, the average recognition results are shown in Table 5.

4.3. Experiments on the Yale database

The Yale face database contains 165 images of 15 individuals (each person providing 11 different images) under various facial

Table 4
The average recognition rates of competing methods on AR database.

Method	Recognition rate	Dimensionality
PCA	*0.6367*	*356*
LDA	*0.6942*	*119*
LPP	*0.5767*	*356*
NPE	*0.6542*	*352*
LPCA	*0.6982*	*146*
LPMIP	*0.6450*	*131*
SPP($\lambda = 10^0$)	***0.7175***	*349*
Proposed($\lambda = 10^{-2}$)	***0.7125***	***214***

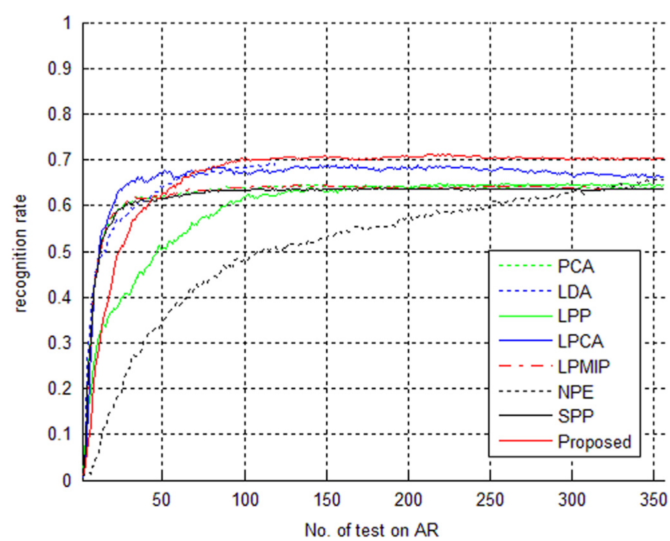

Fig. 5. Recognition rates vs. feature dimensionality on AR database.

Fig. 4. Samples of the cropped images of one person in AR database.

expressions and lighting conditions. The face portion of each image was manually cropped and was resized to 100 by 80 pixels. Fig. 6 shows sample images of one person.

In the first experiment, we use the first 4, 5, 6 images per class for training and the remaining images for testing. For feature extraction, we used, respectively, PCA, LDA, LPP, LPCA, NPE, LPMIP, SPP and CRP. In the PCA phase of LDA, LPP, NPE, SPP and CRP, we keep nearly 98% image energy and select the number of principal components, $m=43$, 50, 53, respectively. The recognition rates are given in Table 6. As can be seen, our proposed method has the good performance.

In the second experiment, 50-fold random tests are performed to reevaluate the performance of PCA, LDA, LPP, LPCA, LPMIP, SPP and CRP. In each test, l images per class are randomly chosen for training, while the remaining images are used for testing. We set the parameters according to the foresaid experiments, and the best results are reported in Table 7 shows the maximal average recognition rates across 50 runs of each method under nearest neighbor classifier with cosine distance merits. From Table 7, it can be seen that proposed method outperforms other methods.

4.4. Experiments on the PolyU FKP database

In the last experiments, we used the PolyU finger-knuckle-print (FKP) database [57,58] to evaluate the performance of the proposed method. The PolyU FKP database was collected from 165 volunteers, including 125 males and 40 females. Among them, 143 subjects are 20–30 years old and the others are 31–50 years old. The images were collected in two separated sessions. In each session, the subject was asked to provide 6 images for each of the left index finger, the left middle finger, the right index finger and the right middle finger. In total, the database contains 7920 images from 660 different fingers. The size of the ROI (region of interest) is 110×200. Fig. 7 shows twelve sample images of one left index finger. In the experiments, we choose a subset of the PolyU FKP database, the images of the left index finger, to evaluate the performance of the proposed method. In the experiments, we resize the image size to 55×110. Please note that in this paper the FKP database is used to illustrate the effectiveness of the proposed RLSDP method and other feature extraction methods. Our goal is not to compare the FKP recognition accuracy of RLSDP with the other methods proposed in [57] because they use totally different categories of techniques.

According to the protocol of this database, the images captured in the first session are used for training and the images captured in the second session for testing. Thus, for each class, there are six training samples and six testing samples. PCA, LDA, LPP, LPCA, LPMIP, NPE, SPP and the proposed CRP are used for FKP feature extraction. In the PCA phase of LDA, LPP, NPE, SPP and CRP, the number of principal components are set as 200. The maximal recognition rate of each method and the corresponding dimension are listed in Table 8. As on the face databases, we see that CRP achieves the highest recognition rate. Fig. 8 shows the recognition rate curve versus the variation of the dimensions and the proposed method consistently outperforms over the other methods.

4.5. Discussion

From the experimental results, we can find that: (1) the supervised methods usually surpass the unsupervised methods (e.g. LDA outperforms PCA in our experiments); (2) the locality preserving properties could increase the recognition performance; (3) the sparsity properties could improve the recognition performance; (4) in the cases of invariant illumination, expression and gesture, the supervised methods are too rigid to get robust performance while the locality preserving and sparsity preserving based methods could get good performance (e.g. experiments on FERET face database). (5) Our CRP methods could get more robust performance than SPP with faster speed. (6) In most cases, our CRP method gets the top performance.

5. Conclusions

In this paper, first, a L2-graph is given like L1-graph. L2-graph could be constructed in one step and avoid manually choosing the

Table 5
The average recognition rates of competing methods on AR database.

Method	mean	std
PCA	*0.8954*	*0.0415*
LDA	***0.9911***	***0.0044***
LPP	*0.9537*	*0.0235*
NPE	*0.9638*	*0.0122*
LPCA	*0.9797*	*0.0100*
LPMIP	*0.9171*	*0.0371*
SPP	*0.9765*	*0.0169*
Proposed	***0.9815***	***0.0105***

Table 6
Recognition rates on the Yale database by different methods.

	l=4	l=5	l=6
PCA	*0.8762*	*0.8778*	*0.8533*
LDA	***0.9429***	***0.9222***	***0.9467***
LPP	*0.8467*	*0.9000*	***0.9200***
NPE	*0.8571*	*0.8778*	*0.8933*
LPCA	*0.8857*	*0.9111*	*0.8933*
LPMIP	*0.8857*	*0.9000*	*0.8800*
SPP	*0.8857* ($\lambda=10^{-2}$)	*0.9111* ($\lambda=10^{4}$)	*0.8933* ($\lambda=10^{-3}$)
Proposed	***0.9333*** ($\boldsymbol{\lambda=10^{-3}}$)	***0.9333*** ($\boldsymbol{\lambda=10^{-1}}$)	***0.9200*** ($\boldsymbol{\lambda=10^{-2}}$)

Table 7
The average recognition rates of competing methods on Yale database.

	l=4	l=5	l=6
PCA	*0.8767*	*0.8858*	*0.9064*
LDA	***0.9078***	***0.9032***	***0.9528***
LPP	*0.8762*	*0.8951*	*0.9200*
NPE	*0.9390*	*0.9562*	*0.9704*
LPCA	*0.8724*	*0.8898*	*0.9163*
LPMIP	*0.8623*	*0.8820*	*0.8960*
SPP	*0.8709*	*0.8893*	*0.9120*
Proposed	***0.8952***	***0.9102***	***0.9248***

Fig. 6. Eleven images of one person in Yale.

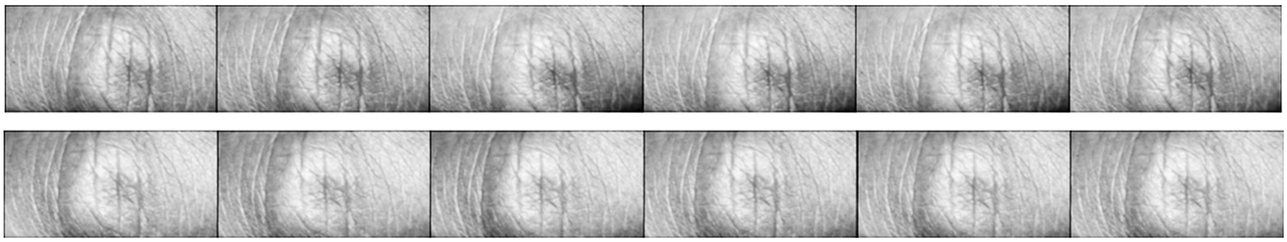

Fig. 7. Samples of the cropped images in the PolyU FKP database.

Table 8
Recognition rates on the PolyU FKP database by different methods.

Method	Recognition Rate	Dimensionality
PCA	*0.3324*	*200*
LDA	*0.6303*	*164*
LPP	*0.6949*	*155*
NPE	*0.4333*	*176*
LPCA	*0.6515*	*190*
LPMIP	*0.6899*	*179*
SPP($\lambda = 10^{-2}$)	*0.6657*	*197*
Proposed($\lambda = 10^{-3}$)	***0.7121***	***162***

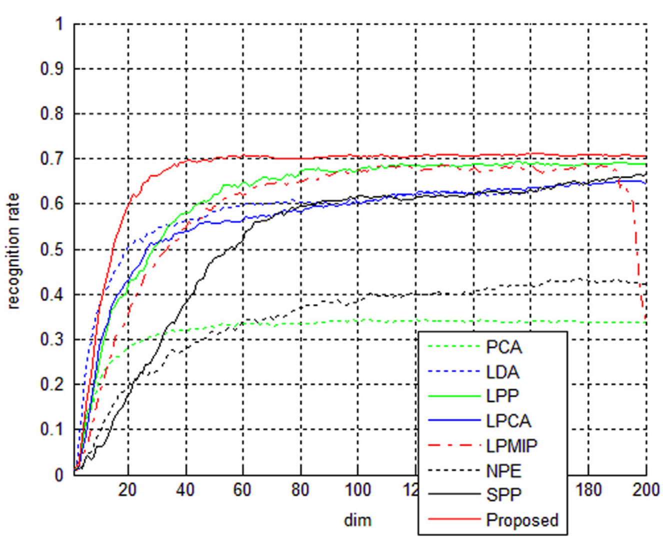

Fig. 8. Recognition rates vs. feature dimensionality on the PolyU FKP database.

nearest neighbors in traditional graph construction (e.g. LLE). L2-graph is much faster than L1-graph. Second, for feature extraction, we characterize the global quantity by the total scatter matrix like PCA. Based on the local and global quantities, we develop a criterion for feature extraction which maximizes the global scatter and minimizes the local scatter simultaneously for classification. Finally, the objective function for feature extraction is a Rayleigh quotient form and can be calculated via generalized eigenvalue decomposition. Since a pointed sample is represented using all the samples in L2-graph, which is like collaborative representation, our proposed feature extraction method is named collaborative representation based projections (CRP). The experimental results on benchmark face databases (FERET, AR, Yale) and the PolyU FKP database show that CRP outperforms many existing representative DR methods.

6. Conflict of interest statement

The authors declare that they have no conflict of interests.

Acknowledgments

This project is partly supported by NSF of China (61375001, 61273023), partly supported by Research Fund for the Doctoral Program of Higher Education of China (20120092110024), partly supported by the open fund of Key Laboratory of Measurement and Control of Complex Systems of Engineering, Ministry of Education (No. MCCSE2013B01), and the Jiangsu Key Laboratory of Image and Video Understanding for Social Safety (Nanjing University of Science and Technology) (No. 30920130122006).

References

[1] J. Wright, Y.I. Ma, J. Mairal, G. Sapiro, T.S. Huang, Shuicheng Yan, Sparse representation for computer vision and pattern recognition, in: Proceedings of the IEEE, 98(6), 2010, pp. 1031–1104.
[2] J. Yang, J. Wright, T. Huang, Y. Ma, Image super-resolution as sparse representation of raw patches, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR2008), 2008.
[3] S. Rao, R. Tron, R. Vidal, Y. Ma, Motion segmentation via robust subspace separation in the presence of outlying, incomplete, and corrupted trajectories, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR2008), 2008.
[4] J. Mairal, G. Sapiro, M. Elad, Learning multiscale sparse representations for image and video restoration, SIAM MMS 7 (1) (2008) 214–241.
[5] J. Wright, A. Yang, S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 210–227.
[6] Y. Xu, D. Zhang, J. Yang, J.-Y. Yang, A two-phase test sample sparse representation method for use with face recognition, IEEE Trans. Circuits Syst. Video Technol. 21 (9) (2011) 1255–1262.
[7] R. He, B.-G. Hu, W.-S. Zheng, Y. Guo, Two-stage sparse representation for robust recognition on large-scale database, in: Proceeding of 24th AAAI Conference on Artificial Intelligence, 2010, pp. 475–480.
[8] R. He, W.-S. Zheng, B.-G. Hu, X.W. Kong, A regularized correntropy framework for robust pattern recognition, Neural Comput. 23 (8) (2011) 2074–2100.
[9] R. He, W.-S. Zheng, B.-G. Hu, Maximum correntropy criterion for robust face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 33 (8) (2011) 1561–1576.
[10] M. Yang, L. Zhang, J. Yang, D. Zhang, Robust sparse coding for face recognition, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR2011), 2011, pp. 625–632.
[11] R.O. Duda, P.E. Hart, D.G. Stork., Pattern Classification, 2nd ed., John Wiley & Sons, New York, 2000.
[12] S.J. Raudys, A.K. Jain, Small sample size effects in statistical pattern recognition: recommendations for practitioners, IEEE Trans. Pattern Ana1. Mach. Intel1. 13 (3) (1991) 252–264.
[13] D.L. Swets, J. Weng, Using discriminant eigenfeatures for image retrieval, IEEE Trans Pattern Anal. Mach. Intell. 18 (8) (1996) 831–836.
[14] V. Belhumeur, J. Hespanha, D. Kriegman, Eigenfaces vs Fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Ana1. Mach. Intel1. 19 (7) (1997) 711–720.
[15] J. Yang, J. Yang, Why can LDA be performed in PCA transformed space? Pattern Recog. 36 (2) (2003) 563–566.
[16] Z. Hong, J. Yang, Optimal discriminant plane for a small number of samples and design method of classifier on the plane, Pattern Recog. 24 (4) (1991) 317–324.
[17] J. Yang, H. Li, PCA based Sequential Feature Space Learning for Gene Selection, ICMLC2010, pp. 3079–3084.
[18] T. Hastie, R. Tibshirani, Penalized discriminant analysis, Ann. Stat. 23 (1995) 73–102.
[19] 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 Recog. 33 (1) (2000) 1713–1726.
[20] H. Yu, J. Yang, A direct LDA algorithm for high dimensional data – with application to face recognition, Pattern Recog. 34 (10) (2001) 2067–2070.
[21] Y. Pang, D. Tao, Y. Yuan, X. Li, Binary two-dimensional PCA, IEEE Trans. Syst. Man, Cybern. B, Cybern. 38 (4) (2008) 1176–1180.

[22] X. Zhuang, D. Dai, Inverse fisher discriminant criteria for small sample size problem and its application to face recognition, Pattern Recog. 38 (11) (2005) 2192–2194.
[23] Z. Jin, J. Yang, Z. Hu, Z. Lou, Face recognition based on the uncorrelated discrimination transformation, Pattern Recog. 34 (7) (2001) 1405–1416.
[24] X. Li, S.S. Lin, D. Yan, Xu, Discriminant locally linear embedding with high-order tensor data, IEEE Trans. Syst. Man, Cybern. B, Cybern. 38 (2) (2008) 342–352.
[25] M. Loog, Approximate pairwise accuracy criterion for multiclass linear dimension reduction: generalization of the Fisher criterion, IEEE Trans. Pattern Anal. Mach. Intell. 26 (7) (2001) 762–766.
[26] R. Lotlikar, R. Kothari, Fractional-step dimensionality reduction, IEEE Trans. Pattern Recog. Mach. Intell. 22 (6) (2000) 623–627.
[27] J. Lu, K.N. Plataniotic, A.N. Venetsanopouls, Face recognition using LDA-based algorithm, IEEE Trans. Neural Netw. 14 (1) (2003) 195–200.
[28] D. Tao, X. Li, X. Wu, S.J. Maybank, Geometric mean for subspace selection, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 260–274.
[29] Y. Xu, D. Zhang, J. Yang, J.-Y. Yang, A two-phase test sample sparse representation method for use with face recognition, IEEE Trans. Circuits Syst. Video Technol. 21 (9) (2011) 1255–1262.
[30] W. Yang, J. Wang, M. Ren, J. Yang, Feature extraction based on laplacian bidirectional maximum margin criterion, Pattern Recog. 42 (11) (2009) 2327–2334.
[31] C. Ren, D. Dai, Incremental learning of bidirectional principal components for face recognition, Pattern Recog. 43 (1) (2010) 318–330.
[32] X.D. Jiang, Asymmetric principal component and discriminant analyses for pattern classification, IEEE Trans. Pattern Anal. Mach. Intell. 31 (5) (2009) 931–937.
[33] X.D. Jiang, B. Mandal, A. Kot, Eigenfeature regularization and extraction in face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 30 (3) (2008) 383–394.
[34] X.G. Wang, X.O. Tang, A unified framework for subspace face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 26 (9) (2004) 1222–1228.
[35] J.B. Tenenbaum, V.d.e. Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (2000) 2319–2323.
[36] S.T. Roweis, L.K. Saul, Nonlinear dimension reduction by locally linear embedding, Science 290 (2000) 2323–2326.
[37] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. 15 (6) (2003) 1373–1396.
[38] X. He, S. Yan, Y. Hu, P. Niyogi, H. Zhang, Face recognition using laplacianfaces, IEEE Trans. Pattern Ana1. Mach. Intel1. 27 (3) (2005) 328–340.
[39] 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 Ana1. Mach. Intel1. 29 (4) (2007) 650–664.
[40] Y. Xu, A. Zhong, J. Yang, D. Zhang, LPP solution schemes for use with face recognition, Pattern Recog. 43 (12) (2010) 4165–4176.
[41] S. Yan, D. Xu, B. Zhang, H. Zhang, Graph embedding and extensions: a general framework for dimensionality reduction, IEEE Trans. Pattern Ana1. Mach. Intell. 29 (1) (2007) 40–51.
[42] H. Wang, S. Che, Z. Hu, W. Zheng, Locality-preserved maximum information projection, IEEE Trans. Neural Netw. 19 (4) (2008) 571–585.
[43] L. Qiao, S. Chen, X. Tan, Sparsity preserving projections with applications to face recognition, Pattern Recog. 43 (1) (2010) 331–341.
[44] 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.
[45] L.M. Zhang, S. Chen, L. Qiao, Graph optimization for dimensionality reduction with sparsity constraints, Pattern Recog. 45 (3) (2012) 1205–1210.
[46] J. Yang, D. Chu, L. Zhang, Y. Xu, J. Yang, Sparse representation classifier steered discriminative projection with application to face recognition, IEEE Trans. Neural Netw. Learn. Syst. 24 (7) (2013) 1023–1034.
[47] L. Clemmensen, T. Hastie, D. Witten, b. Ersboll, Sparse discriminant analysis, Technometrics 53 (4) (2011) 406–413.
[48] Z. Lai, M. Wan, Z. Jin, J. Yang, Sparse two-dimensional local discriminant projections for feature extraction, Neurocomputing 74 (4) (2011) 629–637.
[49] Z. Lai, W.K. Wong, Z. Jin, J. Yang, Y. Xu, Sparse approximation to the eigen subspace for discrimination, IEEE Trans. Neural Netw. Learn. Syst. 23 (12) (2012) 1948–1960.
[50] L. Zhang, M. Yang, X. Feng, Sparse representation or collaborative representation: which helps face recognition, in: Proceedings of IEEE Conference on Computer Vision, 2011 (ICCV2011).
[51] X. He, D. Cai, S. Yan, and H. Zhang, Neighborhood preserving embedding, in Proceedings of IEEE Conference on Computer Vision (ICCV2005), 2005, pp. 1208–1213.
[52] D. Zhao, Z. Lin, and X. Tang. Laplacian PCA and its applications, in: Proceedings of IEEE Conference on Computer Vision (ICCV2007), 2007.
[53] P.J. Phillips, H. Moon, S.A. Rizvi, P.J. Rauss, The FERET Evaluation Methodology for Face Recognition Algorithms,, IEEE Trans. Pattern Ana1. Mach. Intel1. 22 (10) (2000) 1090–1104.
[54] P.J. Phillips, The facial recognition technology (FERET) database. (2004).
[55] A.M. Martinez, R. Benavente, The AR Face Database ⟨http://cobweb.ecn.purdue.edu/~aleix/ aleix_face_DB.html⟩.
[56] A.M. Martinez, R. Benavente, The AR Face Database, CVC Technical Report #24, June 1998.
[57] L. Zhang, L. Zhang, D. Zhang, H. Zhu, Online finger-knuckle-print verification for personal authentication, Pattern Recog. 43 (7) (2010) 2560–2571.
[58] The Hong Kong PolyU FKP database: ⟨http://www4.comp.polyu.edu.hk/~biometrics/FKP.htm⟩.

Wankou Yang received his B.S., M.S. and Ph.D. degrees at the School of Computer Science and Technology, Nanjing University of Science and Technology (NUST), PR China, 2002, 2004 and 2009 respectively. Now he is an assistant professor with school of automation, Southeast University. His research interests include pattern recognition, computer vision, digital image processing.

Zhenyu Wang received the B.Eng. degree in electrica engineering from the Anhui University of Technology, Anhui, China, in 2001, and the M.E. degree in control theory from Kunming University of Science an Technology, Yunnan, China, in 2008.He is currently pursuing the Ph.D. degree at Southeast University, China. His research interests include image/video processing, computer vision, and super-resolution.

Changyin Sun is a professor in School of Automation at Southeast University, China. He received the M.S. and Ph.D. degrees in Electrical Engineering from Southeast University, Nanjing, China, respectively in 2001 and 2003. His research interests include Intelligent Control, Neural Networks, SVM, Pattern Recognition, Optimal Theory, etc. He has received the First Prize of Nature Science of Ministry of Education, China. He has published more than 40 papers. He is an associate editor of IEEE Transactions on Neural Networks, Neural Processing Letters, International Journal of Swarm Intelligence Research, Recent Patents on Computer Science. He is an IEEE Member.