

A dimensionality reduction method based on structured sparse representation for face recognition

Guanghua Gu¹ · Zhichao Hou¹ ·
Chunxia Chen¹ · Yao Zhao²

Published online: 7 March 2016
© Springer Science+Business Media Dordrecht 2016

Abstract Face recognition (FR) has been one of the most fundamental problems in computer vision. Two issues are always concerned in a FR task: one is the dimensionality reduction (DR) of the features, and the other is the sparse representation for the samples. DR is an important step because it can not only reduce the storage space of face images, but also enhance the discrimination of the features. Meanwhile, sparse representation based classification (SRC) has been proved a powerful method to solve the problem of dimensionality. It simply considers the training samples as the dictionary to represent the testing samples. However, most of the SRC algorithms do not consider the structure of the dictionary. To consider these two aspects, in this paper, we proposed a FR method by combining a new DR model with the structured sparse representation (SSR). The key idea is projecting the images on a learned projection matrix, and performing the face classification by the SSR considering the structure information of the dictionary. The validity of the proposed method is verified by the evaluations on three databases.

Keywords Face recognition · Dimensionality reduction · Structured sparse representation · Projection matrix

1 Introduction

Face recognition (FR) has been one of the most challenging tasks in computer vision because of the wide applications. Many techniques have already been proposed to solve the different issues in face recognition. Due to the variety changes of faces, such as expression, pose, occlusion (wearing glasses or hats), there are still many challenges for the FR tasks. [Chen et al. \(2014\)](#) pointed that the face alignments to the samples were helpful to provide better features

✉ Guanghua Gu
guguanghua@ysu.edu.cn

¹ School of Information Science and Engineering, YanShan University, Qinhuangdao, China

² Institute of Information Science, Beijing Jiaotong University, Beijing, China

for face classification. However, the features of face images always have high dimensions, so the dimensionality reduction has been one of the most essential issues in FR. The face images are always represented by low dimensional subspaces. Therefore, the subspace learning methods have been widely discussed in FR (Azeem et al. 2014; Shah et al. 2013; Zhao et al. 2003).

There are some classical dimensionality reduction methods such as principal component analysis (PCA), and linear discriminant analysis (LDA). The PCA method finds the best projection subspace to represent the data. The LDA method finds a linear combination of features to classify two or more classes of objects. Based on the PCA algorithm, Turk and Pentland (1991) proposed the classical Eigenfaces algorithms. In the method, some eigenvectors were derived from the covariance matrix of the probability distribution over the high-dimensional vector space of face images. The Eigenfaces formed a basis set of all images used to construct the covariance matrix. This produced dimensionality reduction by allowing the smaller set of basis images to represent the original training images. Besides, based on the LDA algorithm, Belhumeur et al. (1997) proposed another representative Fisherfaces algorithms. This method was a derivative of Fisher's Linear Discriminant (FLD). It found a linear projection of the faces from the high-dimensional image space to a significantly lower dimensional feature space which was insensitive both to variation in lighting direction and facial expression, and maximized the ratio of between-class scatter to that of within-class scatter. Both the two liner methods PCA and LDA consider the global structure dimensionality reduction. However, the actual data, such as face images fail to be well represented by the liner combination. Recently, to overcome the limitations, the manifold learning methods have been developed rapidly (Tenenbaum et al. 2000; Roweis and Saul 2000; He et al. 2005; Yang et al. 2007). The manifold learning is a nonlinear dimensionality reduction method. It assumes that the high dimensioning data can be represented by a low dimensional manifold space to solve the corresponding embedding projection. There are some representative manifold learning methods such as the complete isometric feature mapping (Isomap) Tenenbaum et al. (2000), locally linear embedding (LLE) Roweis and Saul (2000), locality preserving projection (LPP) He et al. (2005) and unsupervised discriminant projection (UDP) (Yang et al. 2007). Although these methods have the good self-adaption, how to project the new testing data into a low dimensional subspace is still a difficult problem.

Olshausen and Field (1997) proposed the sparse representation (SR) theory, which was a more effective representation method for nature images. Then Wright et al. (2009) used the SR technique for the robust face recognition tasks. The training samples were directly regarded as the dictionary to represent the testing samples. A new testing sample could be represented or coded by the liner combination of the dictionary via l_1 -norm. The results clearly demonstrated the success of sparse representation based classification (SRC), and it could even handle the problem of the face occlusion well. The SRC method attempts to find the minimized sparse representation for a testing sample by all the training samples in the dictionary. Researchers proposed many optimized algorithms based on SRC (Yang et al. 2011; Zhang et al. 2015; Zhuang et al. 2013, 2015; Gao et al. 2015). Yang et al. (2010) proposed the Metaface learning frame, a more robust and compact dictionary learning method. Zhang et al. (2010) proved that the dimensionality reduction (DR) could impact the performance of FR before using SRC, so they proposed the optimal unsupervised DR matrix of the given training dataset based on the framework of SRC. Feng et al. (2013) learned the dictionary by jointing the projection matrix for DR and Metaface. The SRC method took the atom sparse representation as the classification criteria. However, these methods based on SRC failed to consider the structure information in the dictionary.



Fig. 1 The dictionary has a block structure where the training images of each subject form a few blocks

The data containing structure information always affect the performance of classification. Taking face images for example, as shown in Fig. 1, datasets from the same class product a few blocks, and each block has the structure information. The images in each block include different expressions. All the blocks form a new dictionary. Different from the dictionary of SRC that considers each training image as the “atom”, in the new dictionary, each block is regarded as the “atom”. Elhamifar and Vidal (2011) casted the face recognition as a structure sparse representation (SSR) problem, in which the minimized blocks sparse representations of testing samples were found from the dictionary.

Some researchers have presented some methods based on SSR. Hu (2014) combined the discriminative decomposition (DD) model with the structured sparse representation to solve the FR problem with occlusions. Chen and Su (2012) proposed a Gabor wavelet and structured sparse representation based classification (SSRC) for FR tasks. The SSRC method took the structure information of the dictionary into account for a better classification. Even though these methods mainly considered the structure sparse representation as the classifier and considered the structure information in the dictionary, they failed to reduce the dimensions of the dictionary. As well known, the features of face images always have high dimensions, so the dimensionality reduction processing must be considered.

Inspired by Zhang’s (2010) and Elhamifar and Vidal (2011) works, this paper proposed a dimensionality reduction based structured sparse representation (DR-SSR) method for face recognition. After computing the optimal unsupervised DR matrix of the training samples, we projected the training and testing samples on the DR matrix. Then the new given testing samples were classified by using the SSR method. Compared with some other dimensionality reduction methods such as Eigenface and Randomface, our method can achieve the better performance.

2 Related work

2.1 Sparse representation based classification

Wright et al. (2009) proposed the sparse representation based classification (SRC) method by using the SR technique for the robust face recognition tasks. Denote the training samples of the i th class as

$$X_i = [x_{i,1}, x_{i,2}, \dots, x_{i,n}] \in \mathfrak{R}^{m \times n} \tag{1}$$

where $x_{i,j}$ denotes a m -dimensional vector of the j th sample in the i th class. In each class, there are n training samples. Generally, in the same class, a testing sample $y \in \mathfrak{R}^m$ could be well approximated represented by the linear combination of the training samples.

$$y = \sum_{j=1}^n a_{i,j} x_{i,j} = X_i A_i \tag{2}$$

where $A_i = [a_{i,1}, a_{i,2}, \dots, a_{i,n}]^T$ are the representation coefficients.

Suppose that there are K object classes, and the training samples of all classes are used to build the sample dictionary, $X = [X_1, X_2, \dots, X_K]$. If X represents the input testing image y , we can get $y = XA$ where $A = [a_1, a_2, \dots, a_K]$ is the concatenation of the n training samples from all the K classes. Ideally, all the coefficients $a_k, k = 1, 2, \dots, K$ and $k \neq i$ are equal to zeroes or nearly zero. Then the non-zero elements in the coefficient vector could encode the identity of the testing image y . The procedure of SRC algorithm can be summarized as follows:

- a. Normalize each column of X to have unit by using l_2 -norm.
- b. Solve the l_1 -minimization problem:

$$\begin{aligned} \hat{a}_1 &= \arg \min_a \|a\|_1 \\ \text{s.t. } & \|Aa - y\|_2 \leq \varepsilon \end{aligned} \tag{3}$$

- b. Compute the residuals

$$r_i(y) = \|y - A\delta_i\|_2, \quad \text{for } i = 1, \dots, k. \tag{4}$$

where $\delta_i \in \mathfrak{R}^n$ is the coefficient of the i th class.

- d. Output the result of classification by

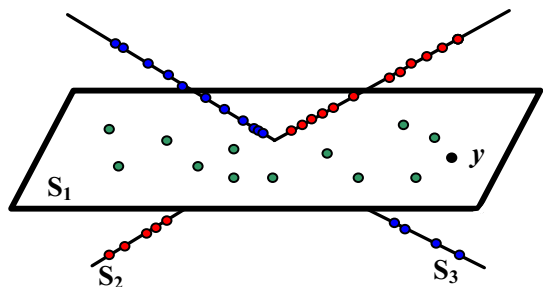
$$\text{class}(y) = \arg \min r_i(y) \tag{5}$$

2.2 Structured sparse representation

The structured sparse representation (SSR) is the expansion of the atom sparse representation (SR). In the SRC, the testing samples can be directly represented by the linear combination of atoms in the training samples, and the coefficient of representation is sparse. However, most of the existing methods based on SR do not take into account the structure of the data. Elhamifar and Vidal (2011) argued that looking for the minimized sparse representation of a testing sample might not be the best classification criterion. They proposed the idea that structure sparse representation finds the minimized sparse block representation to be the criterion for classification.

Figure 2 explains the advantages of SSR. Assume there are three classes of training samples, which are in the subspace $\{S_1, S_2, S_3\}$ respectively as shown in Fig. 2. S_1 is a

Fig. 2 Sparsest representation of a testing sample does not result in a correct classification



2-dimensional subspace, and S_2, S_3 are 1-dimensional subspace. From Fig. 2, the testing sample y belonging to class 1 can be represented by a linear combination of any two data points from class 1. Meanwhile, it can also be represented by a linear combination of one point from class 2 and one point from class 3. From the sparse representation perspective, there is no difference between the two representations because both of them have two nonzero atoms. Thus it can lead to a wrong classification. If we want to find the minimum number of block representations, the expected result of classification will be achieved by choosing only the subspace S_1 to represent the testing samples.

Assume that there are n classes and the i th class contains m_i training samples. Define $B[i] \in \mathfrak{R}^{D \times m_i}$ as the collection of the training data in the i th class and B as the collection of all the training data:

$$B[i] = [b_{i1}b_{i2} \cdots b_{im_i}] \in \mathfrak{R}^{D \times m_i} \tag{6}$$

$$B = [B[1]B[2] \cdots B[n]] \tag{7}$$

Given a testing sample $y \in \mathfrak{R}^D$, then it can be represented as follows:

$$y = [B[1]B[2] \cdots B[n]]c \tag{8}$$

where $c = [c[1]c[2] \cdots c[n]]^T$ is the coefficient vector.

The SRC finds the sparsest representation of the testing examples by the dictionary of all the training data. The SSR transforms the representation of the testing examples to a few blocks of the dictionary corresponding to its class. Then the problem of classification will be turned to look for the minimum number of blocks from the dictionary. The idea of SSR can be formulated using the following non-convex optimization programs

$$P_{\ell_q/l_0} : \min \sum_{i=1}^n I(\|c[i]\|_q > 0) \tag{9}$$

s.t. $y = Bc$

where $I(\cdot)$ is an indicator function, and $c[i] \in \mathfrak{R}^{m_i}$ are the entries of c corresponding to the i th block of the dictionary. The indicator function in Eq. (9) seeks the minimum number of nonzero coefficient blocks. Because the optimization program P_{ℓ_q/l_0} is also a NP-hard problem, it finds all possible blocks of B and check whether they can represent the given y . Eldar and Mishali (2009) proved Eq. (9) can be written by a ℓ_1 relaxation formulation if Eq. (9) meets the specific geometric continuity.

$$P_{\ell_q/l_1} : \min \sum_{i=1}^n \|c[i]\|_q \tag{10}$$

s.t. $y = Bc$

If $q \geq 1$, Eq. (10) is a convex program.

Elhamifar and Vidal (2011) also proposed other optimization program for the classification task, which can be formulated as

$$P'_{\ell_q/l_0} : \min \sum_{i=1}^n I(\|B[i]c[i]\|_q > 0) \tag{11}$$

s.t. $y = Bc$

In Eq. (11), if $q \geq 1$, the ℓ_1 relaxation can be written as

$$\begin{aligned}
 P'_{\ell_q/l_0} : \min & \sum_{i=1}^n \|B[i]c[i]\|_q \\
 \text{s.t. } & y = Bc
 \end{aligned}
 \tag{12}$$

Unlike P_{ℓ_q/l_0} that finds the minimized number of nonzero coefficient blocks $c[i]$, the optimization program P'_{ℓ_q/l_0} minimizes the number of nonzero reconstructed vectors $B[i]c[i]$. Based on Elhamifar’s experiments of P_{ℓ_q/l_0} and P'_{ℓ_q/l_0} processing Elhamifar and Vidal (2011), it proved that P'_{ℓ_q/l_0} performs better than P_{ℓ_q/l_0} , that is, the classification objective function $\|y - B[i]c[i]\|_2$ can result in correct classifications when the vector $B[i]c[i]$ is minimized.

2.3 Sparse dimensionality reduction

The SRC has been proved to be a powerful classifier. It is insensitive to DR or feature extraction. Hence SRC can obtain an acceptable robust recognition result with the DR method such as random projection or direct down-sampling. Moreover, a DR method to reduce the training samples on a low dimension space will lead to the better FR performance. For example, researchers always reduce training samples on a much lower dimension subspace by the popular method PCA, and a better FR result can be obtained by using the training samples on the SRC framework. Zhang et al. (2010) proposed a new unsupervised dimensionality reduction algorithm to compute the desired projection matrix. The DR algorithm achieved a higher recognition rate under the framework of SRC.

The main idea of DR algorithm is as follows. For the training samples S , denote $\mathbf{z}_k \in \mathbb{R}^{m \times 1}$ is the k th training sample, and $D_k = [\mathbf{z}_1, \dots, \mathbf{z}_{k-1}, \mathbf{z}_{k+1}, \dots, \mathbf{z}_n] \in \mathbb{R}^{m \times (n-1)}$ is the collection of training samples without the k th sample. We learn a projection matrix P with the dimension $l \times m$, and $l \ll m$. P is orthogonal, i.e. $PP^T = I$. P can be determined by the following objective function:

$$\begin{aligned}
 J_{P, \{\beta_k\}} = \arg \min_{P, \{\beta_k\}} & \left\{ \sum_{k=1}^n \left(\|P\mathbf{z}_k - PD_k\beta_k\|_F^2 + \lambda_1 \|\beta_k\|_1 \right) + \lambda_2 \|S - P^T P S\|_F^2 \right\} \\
 \text{s.t. } & PP^T = I
 \end{aligned}
 \tag{13}$$

where β_k is the coefficient vector of \mathbf{z}_k over D_k , and λ_1, λ_2 are the scalar parameters of regularization. The minimization of J can be implemented by optimizing P and $\{\beta_k\}$ alternatively.

3 Dimensionality reduction based on structure sparse representation

3.1 Model

Mentioned in Sect. 2, the DR and SSR are always applied separately. Usually, the PCA and LDA methods are firstly used to reduce the dimensionality of training samples, and then the training samples are taken for dictionary directly. Sometimes a discriminate dictionary can be learned from the training samples after the dimensionality reduction. To take advantages

of the new DR method and the structure information of SSR dictionary, in this paper, the DR-SSR method was proposed to learn the DR matrix \mathbf{P} and SSR jointly.

This paper expects to learn a projection matrix \mathbf{P} that keeps the energy of dictionary. Meanwhile the projection matrix should be orthogonal that could maximize the total scatter of the dictionary and the between-class scatter of the dictionary. Inspired the idea of Eq. (13), we improve the method to learn the projection matrix \mathbf{P} as follows:

$$J_{P, \{\beta_i\}} = \arg \min_{P, \{\beta_i\}} \left\{ \sum_{i=1}^K \left(\|P\mathbf{z}_i - PD_i\beta_i\|_F^2 + \lambda_1 \|\beta_i\|_1 \right) + \lambda_2 \|S - P^T P S\|_F^2 \right\} \quad (14)$$

s.t. $PP^T = I$

Suppose there are K classes in the training samples S , and there are n samples in each class. Different from the definition of \mathbf{z}_k , D_k and β_k in Eq. (13), denote that $\mathbf{z}_i \in \mathbb{R}^{m \times n}$ is the training samples in the i th class, $D_i = [\mathbf{z}_1, \dots, \mathbf{z}_{i-1}, \mathbf{z}_{i+1}, \dots, \mathbf{z}_K] \in \mathbb{R}^{m \times (K-1)}$ is the collection of training samples without the i th class and β_i is the coefficient vector of \mathbf{z}_i over D_i in Eq. (14). The projection matrix \mathbf{P} can be obtained by minimizing J .

On the SSR process, the training samples are projected on the projective matrix \mathbf{P} to obtain the dictionary with lower dimension. Then the dictionary was divided into some blocks by using the SSR frame. The dictionary includes the structure information. Summarize our proposed DR-SSR approach in Algorithm 1.

Algorithm 1. Algorithm for DR-SSR model

Input: Data S , parameters λ_1, λ_2 , testing example y

Output: $class(y)$

1. Learn the dimensionality projection matrix \mathbf{P}

Before learning the dimensionality projection matrix, the columns of D_i should be normalized with l_2 -norm. By optimizing Eq. (14), we can obtain the projection matrix \mathbf{P} .

2. Obtain dictionary D

Project all the training samples on the matrix \mathbf{P} and obtain the dictionary D .

3. Classification

Compute the optimal solution via Eq. (12). Project the testing samples on the matrix \mathbf{P} and get a dimensionality reduction testing sample y , which is determined according to

$$class(y) = \arg \min_i \|y - D[i]c[i]\|_2 \quad (15)$$

3.2 Optimization

This paper improved the objection function Eq. (13) to increase the discrimination capability of D . The function Eq. (14) was proposed to learn the projection matrix \mathbf{P} . To minimize the objective function J , we can optimize \mathbf{P} and $\{\beta_i\}$ alternatively. Summarize the optimization algorithm in detail in Algorithm 2.

In Eq. (14), β_i can be solved by the method in Kim et al. (2007).

In Eq. (17), let $N = [\eta_1, \dots, \eta_K]$, $\eta_i = \mathbf{z}_i - D_i\beta_i$, then Eq. (17) becomes

$$J_P = \arg \min_P \left\{ \|PN\|_F^2 + \lambda_2 \|S - P^T P S\|_F^2 \right\}$$

s.t. $PP^T = I$ (18)

Algorithm 2. Learning the dimensionality projection matrix P

Input: Data S , parameters λ_1, λ_2

Output: P

Initialize: To obtain a robust solution, the frame obtains the initialized P by applying PCA to S .

While not converged, $iteration \leq maxIter$ **do**

 Fix P and compute each β_i . Now the objective function J will be reduced to

$$J_{\beta_i} = \arg \min_{\beta_i} \left\{ \|Pz_i - PD_i\beta_i\|_F^2 + \lambda_1 \|\beta_i\|_1 \right\} \quad (16)$$

 Fix z , update P . Now the objective function J will be reduced to

$$J_P = \arg \min_P \left\{ \sum_{i=1}^K \left(\|Pz_i - PD_i\beta_i\|_F^2 \right) + \lambda_2 \|S - P^T P S\|_F^2 \right\} \quad (17)$$

s.t. $PP^T = I$

$iteration = iteration + 1$

end while

Then Eq. (18) can be optimized by

$$\begin{aligned} J_P &= \arg \min_P \left\{ \|PN\|_F^2 + \lambda_2 \|S - P^T P S\|_F^2 \right\} \\ &= \arg \min_P tr \left\{ PNN^T P^T + \lambda_2 (S^T - S^T P^T P)(S - P^T P S) \right\} \\ &= \arg \min_P tr \left\{ PNN^T P^T - \lambda_2 P S S^T P^T + \lambda_2 S^T S \right\} \\ &= \arg \min_P tr \left\{ P(NN^T - \lambda_2 S S^T) P^T + \lambda_2 S^T S \right\} \end{aligned} \quad (19)$$

Seen from the expression in Eq. (19), the term $S^T S$ has no contribution to solve the projection matrix P due to it is generally a constant. Because the term $NN^T - \lambda_2 S S^T$ is a square matrix, we can first apply the singular value decomposition (SVD) technique for the term $NN^T - \lambda_2 S S^T$ to optimize and learn the projection matrix P .

3.3 Convergence of the DR-SSR model

The proposed algorithm to obtain the projection matrix P in Eq. (14) is a convex function, and thus the optimization algorithm in algorithm 2 can reach a local minimum of J . The convergences of the objective function Eq. (14) can guarantee t/o have a stable solution. Figure 3 shows the convergence of the objective function J versus the iteration number for 700 face images in the AR face database. From the Fig. 3 we can see the J decreases and converges rapidly.

4 Experiments

4.1 Datasets

This paper evaluated the performance of the proposed method on three representative face image databases. The first one is the AR database [Martinez \(1998\)](#). It contains over 4000 images for 100 individuals. For each person, there are 26 images taken in two different sessions. In the first session, the images are taken under the same pose with 3 different

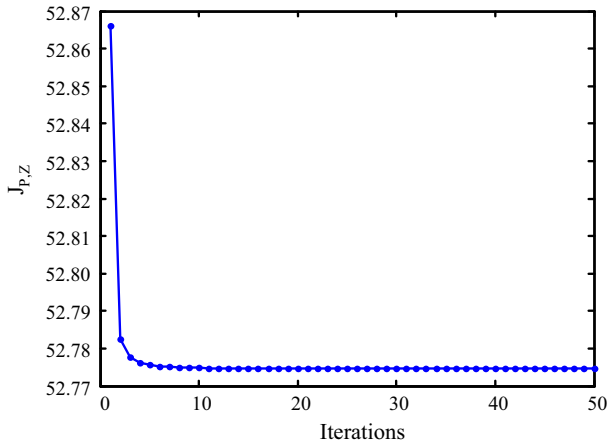


Fig. 3 Example of the convergence

illumination conditions, and 3 different expressions. In the second session, the images contain 2 different facial disguises, wearing a scarf or sunglasses. The original sizes of the images are 165×120 pixels. Resize the images to 60×43 in our experiments.

The second one is the Extended Yale Face Database B (Kuang-Chih et al. 2005). It contains 2414 frontal face images of 38 human subjects under 9 poses and 64 illumination conditions. For each subjects, there are approximately 64 images. The original sizes of images are 192×168 pixels. In the experiments, the sizes of the images are converted to 54×48 .

The last one is Face Recognition Grand Challenge (FRGC) version 2.0 database (Phillips et al. 2005). It is a large-scale face database that contains 7318 images in total of 316 subjects. This database has large lighting, accessory (e.g. glasses) and expression variations. Feng et al. (2013) applied this database and modified it in their experiments. In their work, the images were cropped to 32×42 . This paper followed their setup in experiments.

4.2 Settings

The performance of the proposed DR-SSR method was evaluated on the three databases. The representative dimensionality reduction algorithms under the SRC framework are used for comparison. Those methods include dimensionality reduction for SRC (DR-SRC), SRC with PCA (PCA+SRC), SSR with PCA (SSR+PCA), the dictionary learning method FDDL (Fisher Discrimination Dictionary Learning) Yang et al. (2011) and JDDRDL (Joint Discriminative Dimensionality Reduction and Dictionary Learning) (Feng et al. 2013). On each database, the paper reported the results with different dimensionalities of the features.

For the AR database, the seven images with illumination changes and expressions from session 1 were used for training, and the other seven images with illumination change and expressions from session 2 were used for testing. In the DR learning process, take $\lambda_1 = 0.03$ and $\lambda_2 = 1.5$. In the SSR process, select P'_{ℓ_q/l_0} and $q = 1$. For the comparison experiment, the parameters in the SRC process are $\lambda = 0.005, 0.007$ for the DR and PCA respectively.

For the Extended Yale B database, we randomly choose 20 images per subject for training and the rest for testing. In DR process, select $\lambda_1 = 0.0005$ and $\lambda_2 = 2$. In the SSR process, select P'_{ℓ_q/l_0} and $q = 1$. For the comparison experiment, the paper chooses $\lambda = 0.0005$ for the DR, $\lambda = 0.001$ for PCA in the SRC process.

For the FRGC version 2.0 database, 3 images of each subject were randomly selected for training samples and the remaining as the testing samples. In the DR process, take $\lambda_1 = 0.03$ and $\lambda_2 = 1.5$. In the SSR process, select P'_{ℓ_q/l_0} and $q = 2$. For the comparison experiment, $\lambda = 0.001$ is chosen for DR and PCA.

Note that we choose $q = 1$ for the first two datasets and $q = 2$ for the last one due to the computational complexity of the different datasets. In addition, 5 different feature dimensions (100, 200, 300, 400, 500) are selected to implement the experiment for the three databases.

4.3 FR results

Table 1 shows the recognition results on the AR database. The recognition rates are related to the feature dimensions. From the table, most methods have an increasing recognition rate when the feature dimension takes from 100 to 400. Our DR-SSR method receives better recognition rate than other methods on almost each dimension as shown in Table 1. Though the state-of-the-art JDDRDL method has a powerful recognition result, our DR-SSR outperforms it except for when the feature dimension is 400. Other, the two methods based on PCA, that is PCA+SSR and PCA+SRC, achieve almost the same recognition rates on each dimension. It is clear to see from Table 1 that the methods based on DR (DR-SSR and DR-SRC) present a higher recognition rate than those based on PCA (PCA+SSR and PCA+SRC).

The performance of the face recognition on the Extended Yale B database is demonstrated in Table 2. For DR-SSR, DR-SRC, JDDRDL and FDDL, these four methods have the close recognition rates, and our DR-SSR method has a little superiority to other methods. However, DR-SSR surpasses other methods on almost all the dimensions. From Table 2, we can make

Table 1 Recognition rates on the AR database under different feature dimensions

Dimension	100	200	300	400	500
DR-SSR	0.9000	0.9243	0.9271	0.9343	0.9314
DR-SRC Gao et al. (2015)	0.8729	0.9143	0.9229	0.9300	0.9243
PCA+SSR	0.8171	0.8571	0.8671	0.8643	0.8743
PCA+SRC	0.8214	0.8557	0.8686	0.8700	0.8743
JDDRDL Yang et al. (2010)	0.8729	0.9057	0.9271	0.9400	0.9286
FDDL Yang et al. (2011)	0.8486	0.8886	0.9071	0.9143	0.9200

Bold values indicate the best rate per column

Table 2 Recognition rates on the Extended Yale B database under different feature dimensions

Dimension	100	200	300	400	500
DR-SSR	0.9387	0.9553	0.9652	0.9644	0.9677
DR-SRC Zhang et al. (2010)	0.9363	0.9487	0.9561	0.9603	0.9586
PCA+SSR	0.8940	0.9147	0.9247	0.9247	0.9222
PCA+SRC	0.9015	0.9189	0.9296	0.9197	0.9106
JDDRDL Feng et al. (2013)	0.9346	0.9586	0.9586	0.9611	0.9586
FDDL Yang et al. (2011)	0.9363	0.9512	0.9561	0.9619	0.9603

Bold values indicate the best rate per column

Table 3 Recognition rates on the FRGC 2.0 database under different feature dimensions

Dimension	100	200	300	400	500
DR-SSR	0.7451	0.8196	0.8452	0.8537	0.8375
DR-SRC Zhang et al. (2010)	0.7410	0.7975	0.8133	0.8201	0.8217
PCA+SSR	0.5611	0.6243	0.6493	0.6526	0.6520
PCA+SRC	0.6176	0.6425	0.6369	0.6424	0.6323
JDDRDL Feng et al. (2013)	0.7187	0.8166	0.8385	0.8501	0.8454
FDDL Yang et al. (2011)	0.7312	0.8061	0.8253	0.8342	0.8375

Bold values indicate the best rate per column

the same conclusion as from Table 1 that the methods based on DR achieve the higher recognition rates than those based on PCA.

Table 3 lists the results of the recognition rates on the FRGC 2.0 database. Seen from the results, our DR-SSR method almost achieves better recognition performance than the other methods. Especially, the DR-SSR method outperforms the PCA+SSR method nearly 20% on each feature dimension. For this database, 3 samples are randomly selected per subject from the training samples. The methods based on PCA (PCA+SSR and PCA+SRC) are sensitive to the problem of the small samples, so the two methods have the lower recognition rates. Additional, the JDDRDL method gets a good recognition performance, but is lower than our DR-SSR method. When the feature dimension is 100, the DR-SSR method outperforms the JDDRDL method by about 3%. For the other feature dimensions, the DR-SSR method has almost the same recognition performance as the JDDRDL method.

5 Discussions

In Sect. 4, our method was evaluated on three popular datasets. It demonstrates that the proposed DR-SSR method can improve the classification performance for the FR task. However, not only the different classification methods influence the classification results of face images, but also the classification performances are related to the different datasets. In this section, we discuss the results on the method and datasets.

- (1) It is obvious that the proposed DR-SSR method is very useful for face recognition tasks as shown in the Tables 1, 2 and 3. On one hand, the methods based on DR perform better results than that based on PCA. It is because that the classical PCA method mainly considers the dimensionality reduction of global pictures but not the discrimination between each class. In addition, different from the Zhang's (2010) work which learned the projection matrix P sample by sample, this paper improve the learning of projection matrix P class by class in this paper. According to this way, the projection matrix P increases the discrimination capability of D directly. On the other hand, the structure sparse representation method plays a role on the classification. In this paper, we perform the optimization program P'_{ℓ_q/l_0} to minimize the number of nonzero reconstructed vectors $B[i]c[i]$, which uses the structure information on the data to enhance the efficiency of classification.
- (2) Three representative databases are used to verify the performance of our proposed DR-SSR method. Each database has its own trait, for example, the expression changes are

very large in AR database; the Extend Yale B database has strong variation in illumination; and for FRGC 2.0 database, the images are taken under the uncontrolled conditions. For different complexity of the databases, our method performs better recognition performance than other methods. Notice that only 3 samples were selected per subject for the FRGC 2.0 database to be the training samples in the experiments, which indicates that the DR-SSR method has the superiorities when the number of training samples is inadequate.

- (3) Seen from Tables 1, 2 and 3, it is clear to know that the proposed DR-SSR method has a little better performance than that of the state-of-the-art method JDDRDL, even though the two methods have the similar performance. In the FR experiments, five different feature dimensions have been settled for three datasets. The JDDRDL method has a little superiority to our proposed method only for one out of five dimensions (dimension is 400 in Table 1, 200 in Table 2 and 500 in Table 3). The DR-SSR method combined the dimensionality reduction model with the structured sparse representation frame, and the JDDRDL method jointed the dimensionality reduction model and the Metaface frame. Both of two methods utilized the DR approach to learn the dictionary, but the proposed method also considered the structure information in the dictionary. Overall, the DR-SSR method outperforms the JDDRDL method.

6 Conclusion

This paper proposed a dimensionality reduction method based on structure sparse representation. After improving the processing of the projection matrix learning, we project the training samples into the projection matrix to construct a lower dimension dictionary. The well-learned projection matrix can result in higher face recognition rates at a lower dimensionality. By applying structure sparse representation on the classification process, the structure information is contained in the dictionary, which leads to the better recognition results. The experimental results on the different databases displayed that the proposed DR-SSR method achieved the better performance even when the dictionary dimensionality is relatively small.

However, if the datasets are highly complicated, there is still challenging for the DR method to remove the influence of the illumination and poses. In addition, the DR approach is an unsupervised learning method. The label information can be added into the DR algorithm to obtain a discriminative projection matrix in the next work.

Acknowledgments This work was partly supported by Natural Science Foundation of China (No. 61303128), Natural Science Foundation of Hebei Province (Nos. F2013203220, F2014203132), Key Foundation of Hebei Educational Committee (ZD2015095) and Youth Foundation of Hebei Educational Committee (Q2012047).

References

- Azeem A, Sharif M, Raza M et al (2014) A survey: face recognition techniques under partial occlusion. *Int Arab J Inf Technol* 11:1–10
- Belhumeur P, Hespanha J, Kriegman D (1997) Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Trans Pattern Anal Mach Intell* 19:711–720
- Chen D, Ren S, Wei Y et al (2014) Joint cascade face detection and alignment. *ECCV* 8694:109–122
- Chen T, Su F (2012) Facial expression recognition via gabor wavelet and structured sparse representation. *IC-NIDC*, pp 420–424

- Eldar YC, Mishali M (2009) Robust recovery of signals from a structured union of subspaces. *IEEE Trans Inf Theory* 55:5302–5316
- Elhamifar E, Vidal R (2011) Robust classification using structured sparse representation. *CVPR*, pp. 1873–1879
- Feng Z, Yang M, Zhang L et al (2013) Joint discriminative dimensionality reduction and dictionary learning for face recognition. *Pattern Recognit* 46:2134–2143
- Gao S, Jia K, Zhuang L et al (2015) Neither global nor local: regularized patch-based representation for single sample per person face recognition. *Int J Comput Vis* 111:365–383
- He X, Yan S, Hu Y et al (2005) Face recognition using laplacianfaces. *IEEE Trans Pattern Anal Mach Intell* 27:328–340
- Hu ZP (2014) Discriminative decomposition structure-sparse representation for face recognition with occlusion. *J Signal Process* 30:214–220
- Kim S, Koh K, Lustig M et al (2007) An interior-point method for large-scale l_1 -regularized least squares problems with applications in signal and statistics. *IEEE J Sel Top Signal Process* 1:606–617
- Kuang-Chih L, Ho J, Kriegman D (2005) Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans Pattern Anal Mach Intell* 27:684–698
- Martinez AM (1998) The AR face database. CVC technical report
- Olshausen BA, Field DJ (1997) Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vis Res* 37:3311–3325
- Phillips PJ, Flynn PJ, Scruggs T et al (2005) Overview of the face recognition grand challenge. *CVPR*, pp 947–954
- Roweis ST, Saul LK (2000) Nonlinear dimensionality reduction by locally linear embedding. *Science* 290:2323–2326
- Shah JH, Sharif M, Raza M et al (2013) A survey: linear and nonlinear pca based face recognition techniques. *Int Arab J Inf Technol* 10:536–545
- Tenenbaum JB, De Silva V, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. *Science* 290:2319–2323
- Turk M, Pentland A (1991) Eigenfaces for recognition. *J Cognit Neurosci* 3:71–86
- Wright J, Yang AY, Ganesh A et al (2009) Robust face recognition via sparse representation. *IEEE Trans Pattern Anal Mach Intell* 31:210–227
- Yang J, Zhang D, Yang JY et al (2007) Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics. *IEEE Trans Pattern Anal Mach Intell* 29:650–664
- Yang M, Zhang D, Feng X et al (2011) Fisher discrimination dictionary learning for sparse representation. *ICCV*, pp 543–550
- Yang M, Zhang L, Yang J et al (2010) Metaface learning for sparse representation based face recognition. *ICIP*, pp 1601–1604
- Zhang L, Zhou W, Li F (2015) Kernel sparse representation-based classifier ensemble for face recognition. *Multimed Tools Appl* 74:123–137
- Zhang D, Meng Y, Zhao F et al (2010) On the dimensionality reduction for sparse representation based face recognition. *ICPR*, pp 1237–1240
- Zhao W, Chellappa R, Phillips PJ et al (2003) Face recognition: a literature survey. *ACM Comput Surv (CSUR)* 35:399–458
- Zhuang L, Chan TH, Yang A et al (2015) Sparse illumination learning and transfer for single-sample face recognition with image corruption and misalignment. *Int J Comput Vis* 114:272–287
- Zhuang L, Yang A, Shankar SS et al (2013) Single-sample face recognition with image corruption and misalignment via sparse illumination transfer. *CVPR*, pp 3546–3553