

Enhancing sparsity via full rank decomposition for robust face recognition

Yuwu Lu · Jinrong Cui · Xiaozhao Fang

Received: 30 August 2013 / Accepted: 24 March 2014 / Published online: 16 April 2014
© Springer-Verlag London 2014

Abstract In this paper, we propose a fast and robust face recognition method named enhancing sparsity via full rank decomposition. The proposed method first represents the test sample as a linear combination of the training data as the same as sparse representation, then make a full rank decomposition of the training data matrix. We obtain the generalized inverse of the training data matrix and then solve the general solution of the linear equation directly. For obtaining the optimum solution to represent the test sample, we use the least square method to solve it. We classify the test sample into the class which has the minimal reconstruction error. Our method can solve the optimum solution of the linear equation, and it is more suitable for face recognition than sparse representation classifier. The extensive experimental results on publicly available face databases demonstrate the effectiveness of the proposed method for face recognition.

Keywords Sparse representation · Face recognition · Full rank decomposition · Generalized inverse

1 Introduction

For the importance of face recognition, most techniques of face recognition have been used in pattern recognition [1, 2], image processing [3], and machine learning [4, 5]. Over the last decade, many algorithms have been proposed for face recognition, such as Fisherface [6], Eigenface [7], and graph manifold learning [8]. In face recognition,

appearance-based methods represent an image of size $n \times m$ by a vector in the $n \times m$ dimensional space and then produce lower-dimensional features of the face image for better classifying it [9]. In the application of face recognition, these $n \times m$ dimensional spaces are too large to allow robust and fast face recognition. Dimensionality reduction is a common way to attempt to resolve this problem. Many dimensionality reduction methods have been proposed to produce lower-dimensional features, such as principal component analysis (PCA) [10], linear discriminant analysis (LDA) [11], and 2D-PCA [12]. PCA performs dimensionality reduction by projecting the original data on a much lower-dimensional linear subspace spanned by the leading eigenvectors of the covariance matrix of the origin data. LDA searches for the project axes on which data points of the same class are close to each other and data points of different classes are far from each other. The basic idea of 2D-PCA is to directly use 2D matrices to represent face image which improves the computational efficiency and increases face recognition rate because of the preservation of image structural information.

Sparsity could be a useful principle in neuroscience, information theory, and signal processing over the past few decades [13–15]. Sparse representation encodes an image using only a small number of atoms parsimoniously chosen from an overcomplete dictionary. It has been developed in computer vision and pattern recognition with promising results [16]. Sparse representation classifier (SRC) [17] has been successfully used for robust face recognition. SRC is a nonparametric learning method similar as nearest neighbor (NN) [18] and nearest subspace (NS) [19].

The basic idea of SRC is to represent a test sample as a linear combination of all training samples and then classify the test sample into the class which has the minimal

Y. Lu (✉) · J. Cui · X. Fang
Shenzhen Graduate School, Harbin Institute of Technology,
Shenzhen, China
e-mail: luyuwu2008@163.com

reconstruction error. SRC uses l^1 minimization instead of the l^0 minimization to seek for the sparse solution for its computational convenience and efficiency.

For the efficiency of SRC in face recognition, many extended work have been proposed. In [20], Gao et al. proposed kernel sparse representation, which is the sparse coding in the mapped high-dimensional feature space. In [21], the authors indicated a more general case where features lie in a union of low-dimensional linear subspaces. He et al. [22] incorporated the non-negativity constraint into sparse graph to learn the probabilistic latent clustering relationship between data. Yang et al. proposed robust sparse representation [23].

Although there are so many applications of SRC, l^1 minimization does not always yield sufficiently sparse solution. In [24], the authors proposed l^p ($0 < p < 1$) sparse representation-based classification (l^p -SRC) to seek for the optimal sparse representation of a test image. However, how to get the optimal parameter p is still an open problem. If we can solve the solution of the sparse system directly without iterative steps, we can save much time and improve the accuracy of classification.

In this paper, we propose a novel method named enhancing sparsity via full rank decomposition (ES-FRD) for face recognition, which first represents the test sample as a linear combination of the training data as the same as sparse representation, then make a full rank decomposition of the training data matrix. Since, we can obtain the generalized inverse of the training data matrix and solve the general solution of the linear equation directly. For obtaining the optimum solution to represent the test sample, we use the least square method to solve it. We classify the test sample into the class which has the minimal reconstruction error. The contributions of the proposed method are as follows:

1. Computational efficiency. By introducing the general solution of linear equation, we can obtain the optimum solution. Without counting the norm minimization, the proposed method is more efficient than SRC.
2. Open solution. As we know, the state-of-the-art algorithms of SRC have no open solution to solve norm minimization problem. The proposed method has an open solution.
3. More sparse. Our method introduces full rank decomposition, which factorizes the image database into two low-rank matrices. It is helpful for obtaining much sparser solution.

The rest of this paper is organized as follows: We review the related works on sparse representation in Sect. 2. In

Sect. 3, we give the details of the proposed method. Section 4 gives the experimental results on three public face image data sets. Finally, we conclude the paper in Sect. 5.

2 Related work

In this section, we briefly review SRC and l^p ($0 < p < 1$)-SRC.

2.1 Sparse representation classification (SRC)

Assuming that there are n training samples from c object classes, $A = [A_1, A_2, \dots, A_c] \in R^{m \times n}$ denotes the entire training set, where m is the dimension of sample, and A_i ($i = 1, 2, \dots, c$) is the set of training samples from the i th object class. Given a test sample $y \in R^m$ of the i th class, the goal is exactly to predict the label of y from the given c class training samples. The linear representation of y can be written in terms of all training samples A as [17]

$$y = Ax \quad (1)$$

where x is the vector of coefficients. If the test sample y belongs to the i th class, then the entries of x are expected to be zero except some of those associated with this class.

In SRC, the problem of finding the coefficient vector is formulated as a convex programming problem

$$\min_x \|x\|_1 \quad \text{subject to} \quad y = Ax \quad (2)$$

where $\|\cdot\|$ denotes the l^1 norm. The sparsity of the coefficient vector can be measured by l^0 norm. However, l^0 norm minimization is a NP-hard problem. Some recent studies [25, 26] show that if the solution of x is sparse enough, the l^1 minimization can be employed to seek for the sparse solution for its computational convenience and efficiency.

After obtaining x of Eq. (2), the class reconstruction residual is used to design a sparse representation-based classification (SRC). For each class i , let $\delta_i : R^n \rightarrow R^n$ be the characteristic function which selects the coefficients associated with the i th class. For $x \in R^n$, $\delta_i \in R^n$ is a new vector whose nonzero entries are the entries in x that are associated with class i . The test sample y can be approximated by $\hat{y}_i = A\delta_i(x)$ which uses the vector δ_i from each class. The reconstruction residual for class i is defined as:

$$r_i(y) = \|y - \hat{y}_i\|_2 \quad (3)$$

We classify the given test sample y to the class i associated with the minimal reconstruction residual. We give the algorithm of SRC as follows.

Algorithm of SRC

-
1. Input: The training data matrix $A = [A_1, A_2, \dots, A_c] \in R^{m \times n}$ for c classes, a test sample y
 2. Solve the l^1 -minimization:

$$\min_x \|x\|_1 \quad \text{subject to} \quad Ax = y$$
 3. Then compute the residuals $r_i(y) = \|y - \hat{y}_i\|_2$
 4. Output the identify of y as: Identity $(y) = \arg \min_i \{r_i(y)\}$
-

2.2 $l^p(0 < p < 1)$ sparse representation for face recognition

The optimization problem of l^1 minimization cannot always obtain the sparsest solution. In recent, $l^p(0 < p < 1)$ -norm has been used as an alternative to l^0 norm for sparse signal recovery. The $l^p(0 < p < 1)$ sparse representation-based classification seeks for the optimal sparse representation of a test image by choosing the most suitable parameter p . The following is the l^p minimization problem:

$$\min_x \|x\|_p \quad \text{subject to} \quad Ax = y \tag{4}$$

The authors first proposed an iterative algorithm for solving the non-convex system (4) in [24].

An iterative algorithm for l^p minimization ($0 < p < 1$)

- Step Initialize the iteration count $t = 0$ and coding coefficients $x_i^0 = 1, i = 1, 2, \dots, n$
- Step 1: Update the coding vector x^{t+1} by solving the weighted l^1 minimization problem
- Step 2:

$$x^{t+1} = \arg \min \sum_i^n \frac{|x_i|}{(|x_i^t| + \mu_t)^{1-p}} \quad \text{subject to} \tag{5}$$

$$Ax = y$$

- Step 3: Terminate on convergence or when reaches the maximal number of iteration t_{\max} . Otherwise, let $t = t + 1$, and go to step 2

The solution of l^p minimization is sought by the iterative l^1 minimization algorithm. In the first step, the coding coefficients are initialized by $x_i^0 = 1, i = 1, 2, \dots, n$. Step 2 is to solve a weighted l^1 minimization where the weights $w_i^{t+1} = 1/(|x_i^t| + \mu_t)^{1-p}$ ($i = 1, 2, \dots, n$) depend on the solution of the previous iteration. The weights in Eq. (5) relate inversely to the magnitude of the coefficients, so l^p minimization can partially counteract the influence of the coefficient magnitude on the l^1 penalty function [27].

The residuals are computed by

$$r_i = \|y - A_i \hat{x}_i\|_2 \tag{6}$$

The test sample y is classified to the object class that has the minimize residual.

The state-of-the-art algorithms of sparse representation use norm minimization to get a coefficient vector. Among $l^p(0 < p < 1)$ minimization, l^1 minimization and l^2 minimization, which one is the sparsest is still unknown. Which norm minimization is best suit for sparse representation is still an open problem in theory. In the next section, by introducing the full rank decomposition of the dictionary matrix, we solve the sparse representation system directly, precisely, and computationally efficiently.

The algorithm of $l^p(0 < p < 1)$ SRC as follows.

Algorithm of $l^p(0 < p < 1)$ SRC

1. Input: The training data matrix $A = [A_1, A_2, \dots, A_c] \in R^{m \times n}$ for c classes, a test sample y and the error tolerance $\epsilon \geq 0$
 2. Solve the following l^p minimization problem

$$\min_x \|x\|_p \quad \text{subject to} \quad Ax = y$$
 3. Then compute the residuals $r_i = \|y - A_i \cdot \hat{x}_i\|_2, i = 1, 2, \dots, c$.
 4. Output the identify of y as: Identity $(y) = \arg \min_i \{r_i\}$
-

3 The proposed method

In this section, we present the proposed method and give analysis of it. The main steps of the proposed method are as follows. First, we approximately obtain the full rank decomposition of the training data matrix. Second, we solve the general solution of Eq. (1) and calculate the minimum residual solution. Finally, we classify the test sample into the class that has the minimal residual.

3.1 Full rank decomposition and general solution of the linear equation in our method

In the linear equation $y = Ax$, no matter what the dimension of A is, there is always a general solution of it. We present the details to obtain the general solution of a linear equation as follows.

Definition 1 (full rank decomposition) $A \in R^{m \times n}$ is a matrix of which rank is r . If $A = FG$, $\text{rank}(F) = r$, and $\text{rank}(G) = r$, $A = FG$ is the full rank decomposition of matrix A .

By exploiting the full rank decomposition, we can obtain the generalized inverse matrix of A by Eq. (7). Definition 2 is the generalized inverse matrix of the matrix A .

Definition 2 (Generalized inverse matrix) $A \in R^{m \times n}$ is a matrix which rank is r , and its full rank decomposition is

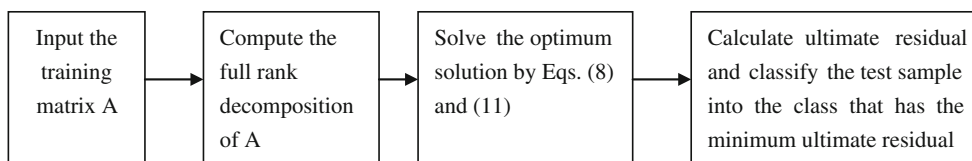


Fig. 1 Flowchart of the proposed method

$A = FG$, $F \in R^{m \times r}$ is full column rank, $G \in R^{r \times n}$ is full row rank, the generalized inverse matrix of A is

$$A^- = G^T (F^T A G^T)^{-1} F^T \tag{7}$$

We can compute a general solution of the linear equation $y = Ax$ by Eq. (8) after solving the generalized inverse matrix of A . Definition 3 defines general solutions of linear equation.

Definition 3 (General solution of linear equation) If $A^- \in R^{m \times n}$ is a generalized inverse matrix of $A \in R^{m \times n}$, then the general solution of linear equation $y = Ax$ is

$$x = A^-y + (I - A^-A)z, \tag{8}$$

where $z \in R^{n \times 1}$ is a random vector.

3.2 The optimum solution in the proposed method for face recognition

Let $A = [A_1, A_2, \dots, A_c] \in R^{m \times n}$ denote the entire training set, in which $A_i \in R^{m \times n_i}$ are training samples of the i th object classes, where c is the number of class. For any test sample $y \in R^m$, the linear representation of y can be written as

$$y = Ax \tag{9}$$

where x is the coefficient vector. By full rank decomposition, we find two matrix factors F and G whose product is an approximation of the matrix A , represented as

$$A = FG \tag{9}$$

By Eq. (7), we solve the generalized inverse matrix of A , that is $A^- = G^T (F^T A G^T)^{-1} F^T$. We obtain the general solution of linear equation $y = Ax$ by Eq. (8), i.e.,

$x = A^-y + (I - A^-A)z$. Because z is a random vector, the general solution of the linear equation $y = Ax$ is not the optimum. We use the least square to solve the optimum solution of Eq. (1). The details as follow: Let

$$W = \|y - Ax\|_2^2 \tag{10}$$

substituting Eq. (8) into (10), we obtain

$$\begin{aligned} W &= \|y - Ax\|_2^2 \\ &= (y - Ax)^T (y - Ax) \\ &= (y^T - x^T A^T) (y - Ax) \\ &= y^T y - 2x^T A^T y + x^T A^T A x \\ &= y^T y - 2(A^-y + (I - A^-A)z)^T A^T y \\ &\quad + (A^-y + (I - A^-A)z)^T A^T A (A^-y + (I - A^-A)z) \end{aligned}$$

Since Eq. (10) is convex and differentiable, any stationary point is a global minimizer of it. Requiring that the derivation of W with respect to z to be zero, we get the following equation

$$W'(z) = -2(I - A^-A)^T A^T y + 2A^T A (I - A^-A) \times (A^-y + (I - A^-A)z) = 0 \tag{11}$$

We can derive from Eq. (11)

$$z = (I - A^-A)^{-1} ((A^T A (I - A^-A))^{-1} (I - A^-A)^T A^T y - A^-y). \tag{12}$$

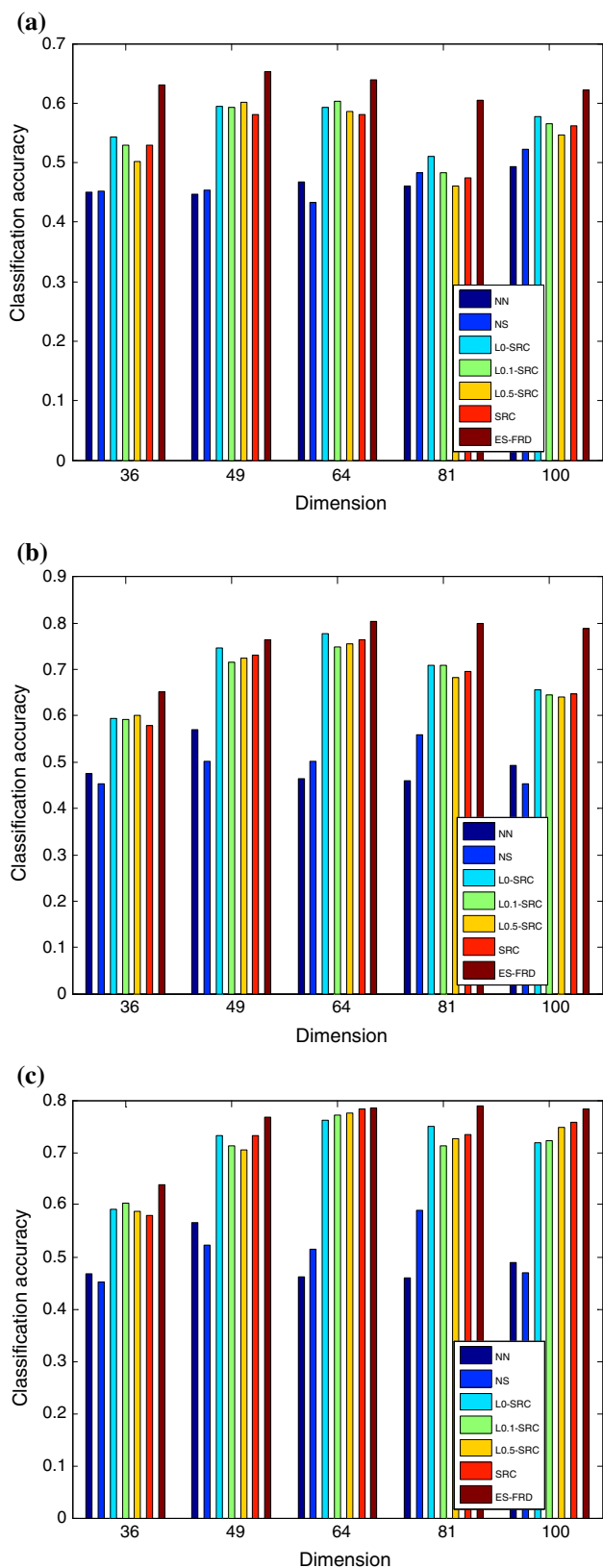
The Eq. (12) is an optimal coefficient vector for the test sample y .

Then, we calculate the residual by

$$r_i(y) = \|y - A\hat{x}\|_2 \tag{13}$$



Fig. 2 Example FERET images used in our experiments. (Images of a subject from the FERET database)



◀**Fig. 3** Face recognition rate on FERET database. We randomly select three (a), four (b), and five (c) images from each subject to construct the training set and the rest used for testing. We conduct five trials for each partition and compare the performance of different algorithms based on the averaged accuracy of the five trials on each dimension for each type of the partition

If $\hat{r}_k = \arg \min_i \{r_i\}, i = 1, 2, \dots, c$, we classify y into the k th class, c is the number of distinguished classes.

The classification procedure of the proposed method is shown in Algorithm 1.

Algorithm 1 Algorithm of the Proposed Method

1. **Input:** a set of training samples $A = [A_1, A_2, \dots, A_c] \in R^{m \times n}$ for c classes, a test sample $y \in R^m$
2. Normalize the columns of A .
3. Compute the full rank decomposition of A
4. Compute the generalized inverse matrix of A by Eq. (7)
5. Compute the general solution of linear equation $y = Ax$ (1) by Eq. (8)
6. Solve the optimal solution \hat{x} for $y = Ax$ (1) by (11)
7. Compute the residuals $r_i(y) = \|y - A\hat{x}\|_2, i = 1, 2, \dots, c$.
8. **Output:** $\text{identity}(y) = \arg \min_i r_i(y)$.

Figure 1 describes the flowchart of the proposed method.

3.3 Analysis of the proposed method

In this section, we analyze the characteristics, rationale, and potential advantages of our method. Our method differs from SRC and $l^p(0 < p < 1)$ SRC as follows: Our method uses full rank decomposition to represent the training data matrix approximately and then best represents the test sample as a linear combination of the training data. Here, the “best” means that the residual between the obtained linear combination and the test sample is the smallest. SRC and $l^p(0 < p < 1)$ SRC are all use norm minimization to obtain an approximately solution for classifying a new test sample.

The proposed method represents the training data matrix by full rank decomposition and then expresses the test sample as a linear combination of the training data. We use generalized inverse of matrix to solve the minimum norm solution of the linear equation. We classify the test sample by evaluating the reconstruction error class by class. Our method can be also viewed as a method that exploits a linear combination of all training samples to represent the



Fig. 4 Images of one subject in ORL

test sample and calculate the solution of this linear representation. The underlying rationale is that for different test samples, the coefficients of the linear representation are different. We can solve the optimal solution of Eq. (8) for the linear representation with the different test samples.

The advantages of our work are as follows:

1. The proposed method introduces the full rank decomposition, which factorizes the image database into two low-rank matrices. It is helpful for obtaining much sparser solution.
2. By introducing the general solution of the linear equation, we can find the optimum solution for the sparse system (1). There is no need to solve the norm minimization problem in the proposed method, thus leading to more efficient procedures.
3. To our knowledge, there is no method can directly solve the solution of SRC. The state-of-the-art algorithms of SRC use norm minimization to solve the coefficient vector. Collaborative representation-based classification (CRC) uses l^2 norm to solve the sparse solution of the linear equation. It obtains dominant performance for face recognition, but it should consider the distribution of data [28]. The proposed method can solve the linear representation solution directly and efficiently and not need consider the distribution of the data. That is to say, our method can directly applied data with any distribution.

4 Experimental results and analysis

We use the FERET [29], ORL [30], and AR databases [31] to evaluate the performance of the proposed method for face recognition. We compare our method with NN, NS, SRC, and $l^p(0 < p < 1)$ SRC. For all above learning algorithms, we test the classification performance with the feature subspace dimensions of 36, 49, 64, 81, and 100. The parameters set as Refs. [17] and [24].

4.1 Evaluation on the FERET database

The FERET database [29] was acquired without any restrictions imposed on facial expression and with at least two frontal images shot at different times during the same

photo session. The image sets used for evaluating face recognition algorithms displays diversity across gender, ethnicity, and age. For the FERET face database, we only use a subset made up of 1,400 images from 200 individuals with each subject providing seven images [32]. We crop and normalize all the face images of FERET in 40×40 . We random select different number (3, 4, 5) of images from each subject to construct the training set, and the rest images make up the test set. Figure 2 shows some example images used in our experiments. Figure 3 is the result of NN, NS, SRC, $l^p(0 < p < 1)$ SRC and our method conducted on FERET face database.

From Fig. 3, the classification accuracy of the proposed method is higher than other comparison methods. In particular, when the data dimension is 81 and we random choose 3 samples as training samples, our method achieves 13 % than SRC.

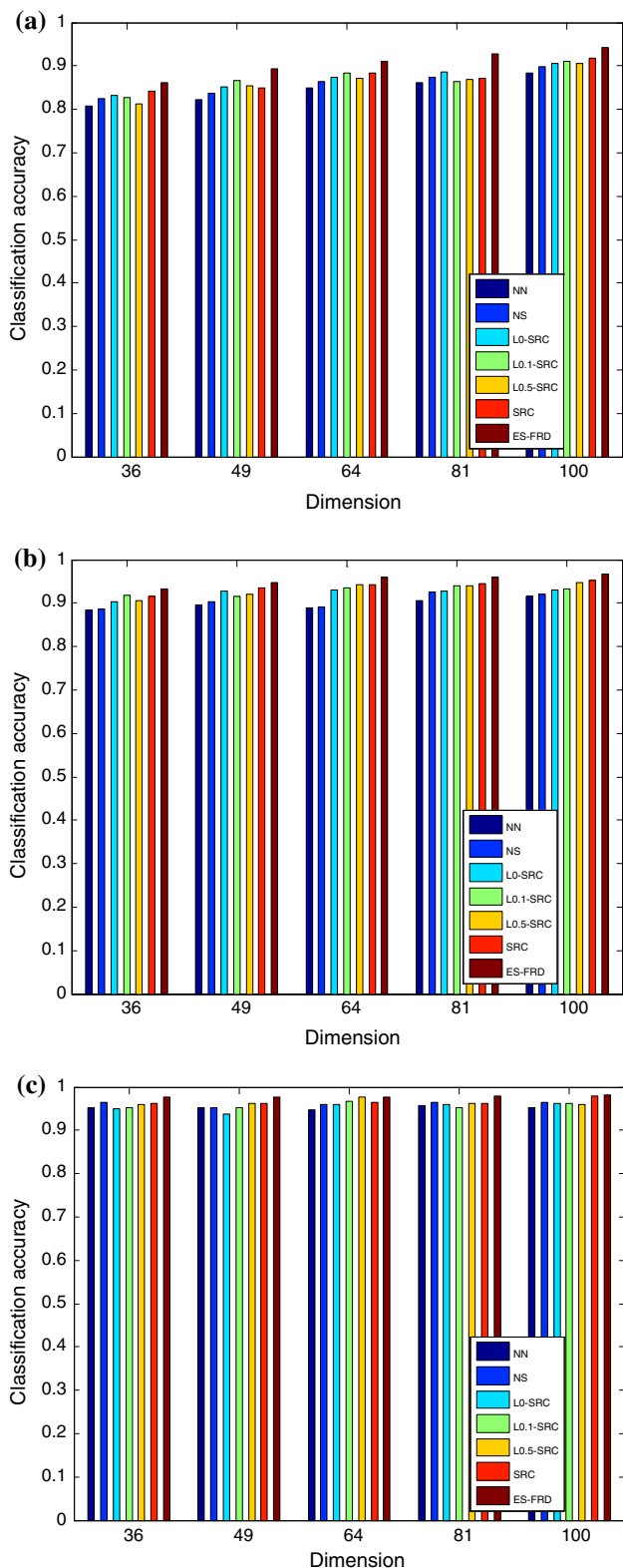
4.2 Evaluation on the ORL database

The ORL database [30] contains images from 40 individuals, each providing 10 different images. All subjects are in up-right, frontal position (with tolerance for some side movement). The size of each face image is 112×92 , and the resulting standardized input vectors are of dimensionality 10,304. In the experiment, the images are converted into the size of 40×40 . Figure 4 shows images of the same subject of ORL. Figure 5 is the result of NN, NS, SRC, $l^p(0 < p < 1)$ SRC and our method conducted on ORL face database.

From Fig. 5, the accuracy of classification of the proposed method is higher than other comparison methods. Fig. 5b shows that at the data dimension is 81 and when we random choose 5 samples as training samples, our method achieves 1.5 % than SRC.

4.3 Evaluation on the AR database

The AR [31] database consists of more than 4,000 face images of 126 subjects (70 men and 56 women). The database characterizes divergence from ideal conditions by incorporating various facial expressions (neutral, smile, anger, and scream), occlusion modes (sunglass and scarf), and luminance alterations (left light on, right light on, and



◀**Fig. 5** Face recognition rate on ORL database. We randomly select four (a), five (b), and six (c) images from each subject to construct the training set and the rest used for testing. We conduct five trials for each partition and compare the performance of different algorithms based on the averaged accuracy of the five trials on each dimension for each type of the partition

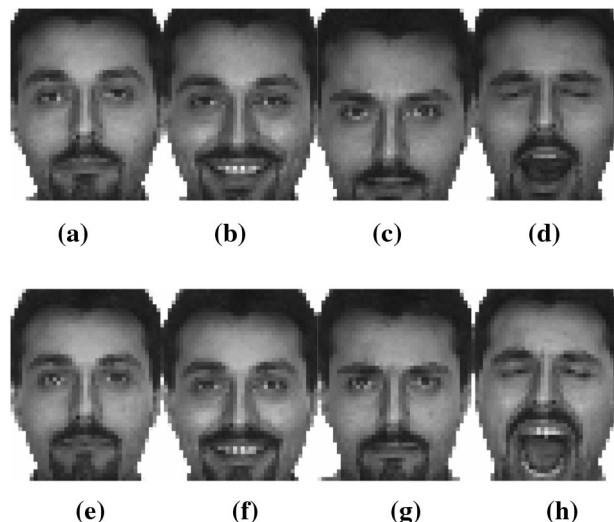


Fig. 6 Facial expression variation in the AR database. a–d and e–h correspond to two different sessions incorporating neutral, happy, angry, and screaming expressing, respectively

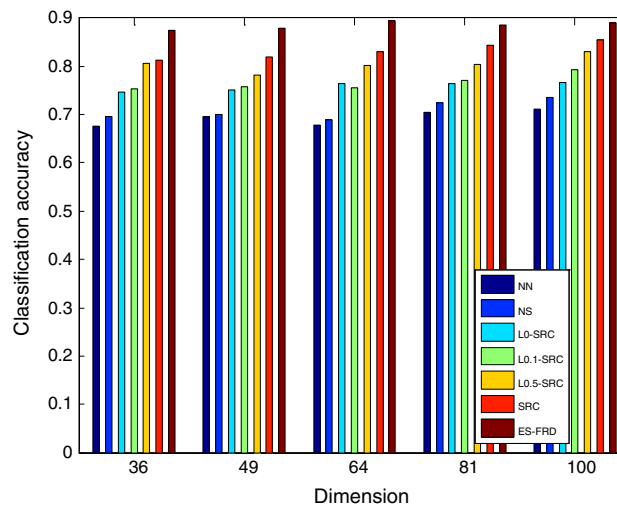


Fig. 7 Face recognition rate for testing variations in facial expression. We conduct five trials for each partition and compare the performance of different algorithms based on the averaged accuracy of the five trials on each dimension for each type of the partition

all side lights on). Each individual participated in two sessions, separated by 2 weeks (14 days). In the experiments, we exploited cropped face images of 100 subjects (50 men and 50 women). We crop and normalize all the

face images of AR in 40×40 . We test the robust performance of the proposed method on AR database. The experiments are conducted for variations in facial expression, variations in lighting condition, and contiguous occlusion.

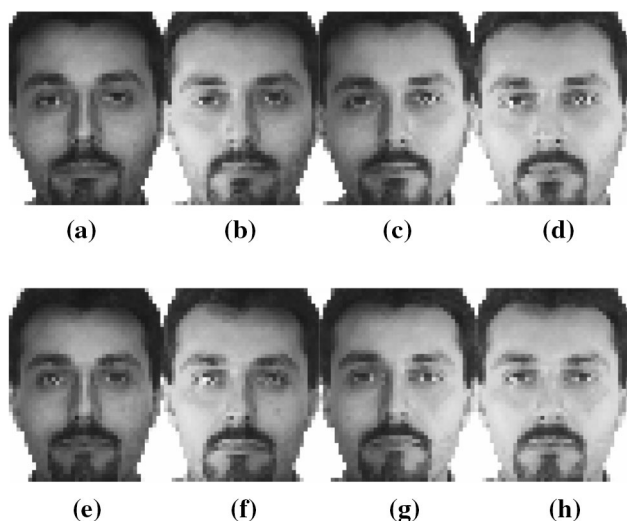


Fig. 8 Lighting variations images in the AR database. **a–d** and **e–h** correspond to two different sessions incorporating neutral, left light on, right light on, and all sides light on, respectively

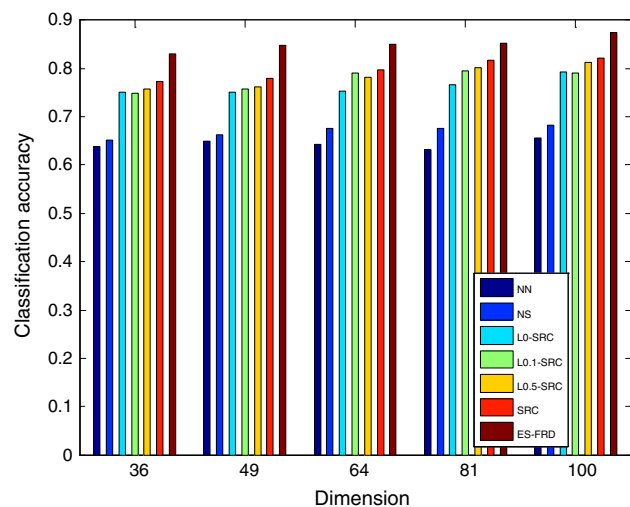


Fig. 9 Face recognition rate for testing variations in lighting condition. We conduct five trials for each partition and compare the performance of different algorithms based on the averaged accuracy of the five trials on each dimension for each type of the partition

4.3.1 Variations in facial expressions

We selected a subset database in which involves variations in facial expressions. Figure 6 shows images used for testing variations in facial expression of one subject. Figure 6a, e is used for training and the others are used for testing. The number of training sample is 240 and the number of test sample is 720. Figure 7 shows the results of variations in facial expressions.

From Fig. 7, we can see that SRC has better classification performance on facial expression than $p(0 < p < 1)$

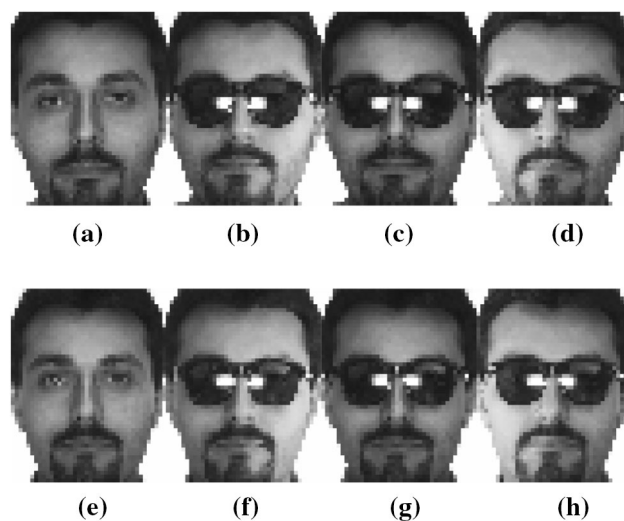


Fig. 10 Images of sunglasses occlusion in AR database. **b–d** and **f–h** are sunglasses occlusion images of one individual

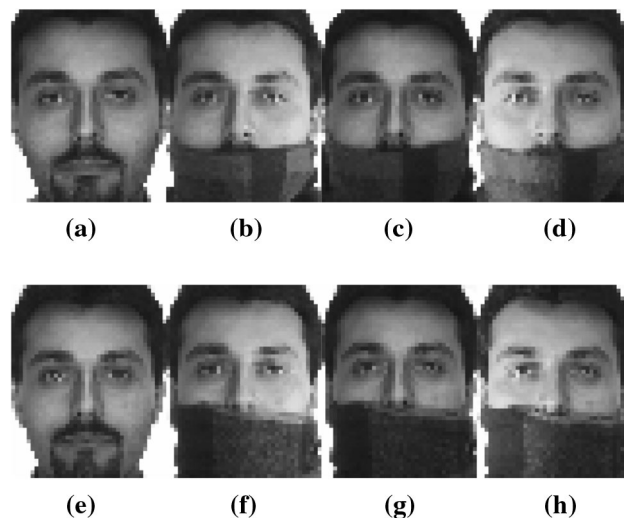


Fig. 11 Images of scarf occlusion in AR database. **b–d** and **f–h** are scarf occlusion images of one individual

SRC. Our method has the best performance among all the comparison methods.

4.3.2 Variations in lighting conditions

We selected images which involves lighting change on left, right, and all sides as a subset for test variations in lighting conditions. Figure 8a, e is used for training and the rest images of Fig. 8 are used for testing. Thus, the total number of training samples is 240 and the total number of test samples is 720. Figure 9 shows the experimental results of lighting variation tested on AR database.

From Fig. 9, we can see our method has the best performance among all the comparison methods. SRC has

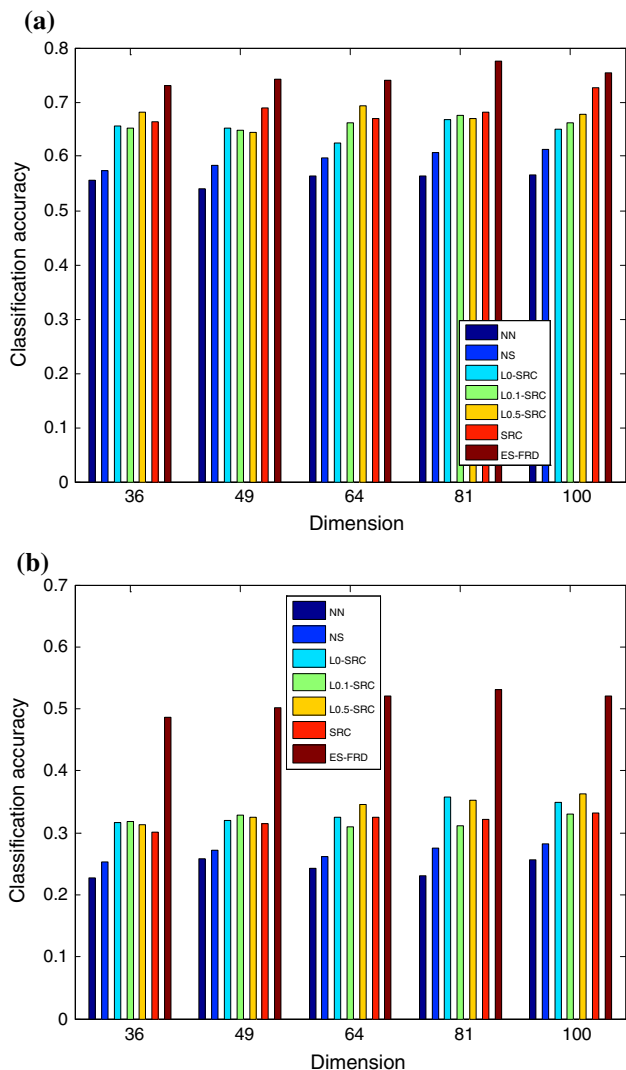


Fig. 12 Face recognition rate for testing contiguous occlusion. We conduct five trials for each partition and compare the performance of different algorithms based on the averaged accuracy of the five trials on each dimension for each type of the partition. **a** Sunglasses occlusion. **b** Scarf occlusion

Table 1 Sparseness of SRC, $l^p(0 < p < 1)$ SRC, and our method on ORL, FERET database

Database	l^0 SRC	$l^{0.1}$ SRC	$l^{0.5}$ SRC	SRC	Our method
ORL	0.6052	0.7011	0.7311	0.7602	0.7806
FERET	0.5942	0.6432	0.6128	0.7264	0.7427

better classification performance on lighting condition than $l^p(0 < p < 1)$ SRC.

4.3.3 Contiguous occlusion

The problem of face identification in the presence of contiguous occlusion is arguably one of the most challenging

paradigms in the context of robust face recognition [33]. For testing the performance of ES-FRD on the contiguous occlusion, we make two parts of experiments in this section. We test the occlusion with sunglasses and scarf, respectively. Figures 10 and 11 are used to test the occlusion with sunglasses and scarf, respectively. Figure 10a, e is used for training and the others are used for testing. Thus, the total number of training samples is 240. The experiment scheme of scarf occlusion is all the same as sunglasses occlusion.

From the Figs. 7, 9 and 12, we verify the robust performance of the proposed method. The proposed method has better classification performance for variations in facial expression, lighting variations, and contiguous occlusion in AR database.

4.4 Sparsity evaluation

In this section, we evaluate the sparsity of the proposed algorithm. According to the Ref [34], the sparseness can be calculated by the equation defined as:

$$\text{sparseness}(v) = \frac{\sqrt{t} - (\sum_i |v_i|) / \sqrt{\sum v_i^2}}{\sqrt{t} - 1}$$

where t is the dimensionality of vector v . From Table 1, we can see that the proposed method has more sparseness than other comparison methods.

5 Conclusion

Sparse representation-based classification (SRC) has been successfully applied for face recognition. SRC seeks the sparsest linear combination of training sample for any test samples, but it is time consuming for counting the norm minimization solution of the associated coding coefficients. In the paper, we propose a novel method for face recognition, which need not count norm minimization problem. The proposed method first approximately represents the training data matrix by full rank decomposition, then represents the test sample as a linear combination of the training data. The generalized inverse of matrix is used for solving the solution of linear equation. We classify the test sample into the class which has the minimum residual. The experiment results suggest that the proposed method achieves higher accuracy for face recognition.

Acknowledgments This article is partly supported by the Natural Science Foundation of China (Grant Nos. 61203376, 61375012, 61263032, and 61362031), the General Research Fund of Research Grants Council of Hong Kong (Project No. 531708), the China Postdoctoral Science Foundation under Project 2012M510958 and 2013T60370, the Guangdong Natural Science Foundation under Project S2012040007289, and Shenzhen Municipal Science and Technology Innovation Council (Nos. JC201005260122A,

JCYJ201206 13153352732 and JCYJ20120613134843060, JCYJ20130329152024199).

References

- Howland P, Wang J, Park H (2006) Solving the small sample size problem in face recognition using generalized discriminant analysis. *Pattern Recogn* 39:277–287
- Xu Y, Zhang D, Jin Z, Li M, Yang J-Y (2006) A fast kernel-based nonlinear discriminant analysis for multi-class problems. *Pattern Recogn* 39:1026–1033
- Lai ZH, Wong WK, Jin Z, Yang J, Xu Y (2012) Sparse approximation to the eigensubspace for discrimination. *IEEE Trans Neural Netw Learn Syst* 23:1948–1960
- Jiang X, Mandal B, Kot A (2008) Eigenfeature regularization and extraction in face recognition. *IEEE Trans Pattern Anal Mach Intell* 30:383–394
- Xu Y, Zhang D, Yang J-Y (2010) A feature extraction method for use with bimodal biometrics. *Pattern Recogn* 43:1106–1115
- Belhumeur PN, Hespanha J, Kriegman DJ (1997) Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Trans Pattern Anal Mach Intell* 19:711–720
- Turk M, Pentland A (1991) Eigenfaces for recognition. *J Cogn Neurosci* 3:71–86
- Roweis S, Saul L (2000) Laplacian eigenmaps and spectral techniques for embedding. *Science* 290:2323–2326
- He XF, Yan SC, Hu YX, Niyogi P, Zhang HJ (2005) Face recognition using laplacianfaces. *IEEE Trans Pattern Anal Mach Intell* 27:328–340
- Levin A, Shashua A (2002) Principal component analysis over continuous subspaces and intersection of half-spaces. In: *Proceedings of European conference computer vision*. Springer, Berlin
- Fan ZZ, Xu Y, Zhang D (2011) Local linear discriminant analysis framework using sample neighbors. *IEEE Trans Neural Netw* 22:1119–1132
- Yang J, Zhang D, Frangi AF, Yang J (2004) Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE Trans Pattern Anal Mach Intell* 26:131–137
- Zhou T, Tao D, Wu X (2011) Manifold elastic net: a unified framework for sparse dimension reduction. *Data Min Knowl Disc* 22:340–371
- Ji S, Xue Y, Carin L (2008) Bayesian compressive sensing. *IEEE Trans Signal Process* 56:2346–2356
- Poggio T, Girosi F (1998) A sparse representation for function approximation. *Neural Comput* 10:1445–1454
- Wright J, Ma Y, Mairal J, Sapiro G, Huang T, Yan S (2010) Sparse representation for computer vision and pattern recognition. *Proc IEEE* 98:1031–1044
- Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y (2009) Robust face recognition via sparse representation. *IEEE Trans Pattern Anal Mach Intell* 31:210–227
- Duda R, Hart P (2001) and D. Stork. *Pattern Classification*, second ed. John Wiley & Sons
- Ho J, Yang M, Lim J, Lee K, Kriegman D (2003) Clustering appearances of objects under varying illumination conditions. *Proceedings of the IEEE international conference computer vision and pattern recognition*, pp 11–18
- Gao S, Tsang I, Chia L-T (2010) Kernel sparse representation for image classification and face recognition. In: *ECCV*
- Elhamifar E, Vidal R (2009) Sparse subspace clustering. In: *CVPR*
- He R, Zheng W-S, Hu B-G, Kong X-W (2011) Nonnegative sparse coding for discriminative semi-supervised learning. In: *CVPR*
- Yang M, Zhang L, Yang J, Zhang D (2011) Robust sparse coding for face recognition. In: *CVPR*
- Guo S, Wang Z, Ruan Q (2013) Enhancing sparsity via $l^p(0 < p < 1)$ minimization for robust face recognition. *Neurocomputing* 99:592–602
- Donoho D (2006) For most large underdetermined systems of linear equations the minimal l_1 -norm solution is also the sparsest solution. *Commun Pure Appl Math* 59:797–829
- Candès E, Romberg J, Tao T (2006) Stable signal recovery from incomplete and inaccurate measurements. *Commun Pure Appl Math* 59:1207–1223
- Candès EJ, Wakin M, Boyd S (2008) Enhancing sparsity by reweighted l_1 minimization. *J Fourier Anal Appl* 14:877–905
- Zhang L, Yang M, Feng XC (2011) Sparse representation or collaborative representation: which helps face recognition? In: *ICCV*
- http://www.itl.nist.gov/iad/humanid/feret/feret_master.html
- Samaria S, Harter AC (1994) Parameterisation of a stochastic model for human face identification. In *Proceedings of the 2nd IEEE international workshop applications of Computer Vision*, Sarasota, FL
- Martinez A, Benavente R (1998) The AR face database, CVC technical report 24
- Phillips PJ, Moon H, Rizvi SA, Rauss PJ (2000) The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans Pattern Anal Mach Intell* 22:1090–1104
- Kim MY (2013) Accelerated max-margin multiple kernel learning. *Appl Intell* 38:45–57
- Hoyer PO (2004) Nonnegative matrix factorization with sparseness constraints. *J Mach Learn Res* 5:1457–1469