Neurocomputing 113 (2013) 213–219

Contents lists available at SciVerse ScienceDirect

Neurocomputing

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

# Optimized projections for sparse representation based classification

Can-Yi Lu [a,b], De-Shuang Huang [a,*]

[a] School of Electronics and Information Engineering, Tongji University, 4800 Caoan Road, Shanghai 201804, China
[b] Department of Automation, University of Science and Technology of China, Hefei, China

ARTICLE INFO

*Article history:*
Received 20 May 2012
Received in revised form
23 December 2012
Accepted 13 January 2013
Communicated by Shiguang Shan
Available online 14 March 2013

*Keywords:*
Dimensionality reduction
Sparse representation
Face recognition

ABSTRACT

Dimensionality reduction (DR) methods have been commonly used as a principled way to understand the high-dimensional data such as facial images. In this paper, we propose a new supervised DR method called Optimized Projections for Sparse Representation based Classification (OP-SRC), which is based on the recent face recognition method, Sparse Representation based Classification (SRC). SRC seeks a sparse linear combination on all the training data for a given query image, and makes the decision by the minimal reconstruction residual. OP-SRC is designed on the decision rule of SRC, it aims to reduce the within-class reconstruction residual and simultaneously increase the between-class reconstruction residual on the training data. The projections are optimized and match well with the mechanism of SRC. Therefore, SRC performs well in the OP-SRC transformed space. The feasibility and effectiveness of the proposed method is verified on the Yale, ORL and UMIST databases with promising results.


## 1. Introduction

In many application domains, such as appearance-based object recognition, information retrieval and text categorization, the data are usually provided in high-dimensional form. One of the problems is the so-called "curse of dimensionality" [1], which is a well known but not entirely well-understood phenomenon. Limited data lie in high-dimensional space, and important features are not so much. Moreover, it has been observed that a large number of features may actually degrade the performance of classifiers if the number of training samples is small relative to the number of features [2]. Consequently, dimensionality reduction is essential not only to engineering applications but also to the design of classifiers. In fact, the design of a classifier becomes extremely simple if all patterns of the same class hold the same feature vector while hold different feature vectors between classes.

Up to now, a large family of algorithms had been designed to provide different solutions to the problem of DR. Among them, the linear algorithms Principal Component Analysis (PCA) [3] and Linear Discriminative Analysis (LDA) [4] had been the two most popular methods due to their relative simplicity and effectiveness. However, PCA and LDA considered only the global scatter of training samples and they failed to reveal the essential data structures nonlinearly embedded in a high dimensional space. To overcome these limitations, the manifold learning methods were proposed by assuming that the data lie in a low dimensional manifold of the high dimensional space [5]. Locality Preserving Projection (LPP) [6] was one of the representative manifold learning methods. Success of manifold learning implies that the high dimensional facial images can be sparsely represented or coded by the representative samples on the manifold. Very recently, Wright et al. presented a Sparse Representation based Classification (SRC) method for face recognition [7]. The main idea of SRC is to represent a given test sample as a sparse linear combination of all training samples, the nonzero sparse representation coefficients are supposed to concentrate on the training samples with the same class label as the test sample. SRC shows that the classification performance of most meaningful features converges when the feature dimension increases if a SRC classifier is used. Although this does provide some new insights into the role of feature extraction played in a pattern classification tasks, Qiao et al. [8] argued that designing an effective and efficient feature extractor is still of great importance since the classification algorithm could become simple and tractable, and a unsupervised DR method called Sparsity Preserving Projections (SPP) was proposed, which aimed to preserve the sparse reconstructive relationship of the data in low-dimensional subspace. Yang and Chu [9] proposed a Sparse Representation Classifier steered Discriminative Projection (SRC-DP) method. It used the decision rule of SRC to steer the design of a dimensionality reduction method. SRC-DP iteratively obtained the projection matrix and spare coding coefficient of each training data. But the convergence of SRC-DP was not clear, and also it was time consuming due to the large computing cost of iterative sparse coding.

* Corresponding author. Tel.: +86 21 3351414; fax: +86 21 33514140.
*E-mail addresses:* canyilu@mail.ustc.edu.cn (C.-Y. Lu), dshuang@tongji.edu.cn, huangdeshuang@yahoo.com (D.-S. Huang).

0925-2312/$ - see front matter Crown Copyright © 2013 Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.01.009

In this paper, to enhance the recognition performance of SR, we propose a supervised DR method based on the sparse representation, which is named the Optimized Projections for Sparse Representation based Classification (OP-SRC). Similar to SRC-DP, OP-SRC aims to gain a discriminative projection such that SRC achieves the optimum performance in the transformed low-dimensional space. Since SRC predicts the class label of a given test sample based on the representational residual, OP-SRC utilizes the label information to enhance the residuals more informative. We will also show that OP-SRC is naturally orthogonal, which may help preserve the shape of the data distribution.

The remainder of this paper is organized as follows: Section 2 reviews the SRC algorithm. Section 3 presents the OP-SRC method. The experimental results are presented in Section 4 and some discussions will be presented based on the results on several databases. Finally, we conclude this paper in Section 5.

## 2. Sparse representation based classification

Given sufficient $c$ classes training samples, a basic problem in pattern recognition is to correctly determine the class which a new coming (test) sample belongs to. We arrange the $n_i$ training samples from the $i$-th class as columns of a matrix $X_i = [x_{i1}, \ldots, x_{in_i}] \in \mathbb{R}^{m \times n_i}$, where $m$ is the dimension. Then we obtain the training sample matrix $X = [X_1, \ldots, X_c] \in \mathbb{R}^{m \times n}$, where $n = \sum_{i=1}^{c} n_i$ is the total number of training samples.

Under the assumption of linear representation, a test sample $y \in \mathbb{R}^m$ will approximately lie on the linear subspace spanned by training samples

$$y = X\alpha \in \mathbb{R}^m \tag{1}$$

If $m < n$, the system of Eq. (1) is underdetermined, its solution is not unique. This motivates us to seek the sparest solution to Eq. (1), by solving the following $\ell^0$-minimization problem:

$$(\ell^0): \ \hat{\alpha}_0 = \arg\min \|\alpha\|_0 \text{ subject to } y = X\alpha, \tag{2}$$

where $\|\cdot\|_0$ denotes the $\ell^0$-norm, which counts the number of nonzero entries in a vector. However, the problem of finding the sparsest solution of an underdetermined system of linear equations is NP-hard and difficult even to approximate [10]. The theory of compressive sensing [11,12] reveals that if the solution to the $\ell^0$-minimization problem is sparse enough, then it is equal to the following $\ell^1$-minimization problem:

$$(\ell^1): \ \hat{\alpha}_1 = \arg\min \|\alpha\|_1 \text{ subject to } y = X\alpha. \tag{3}$$

In order to deal with occlusion, the $\ell^1$-minimization problem is extended to the stable $\ell^1$-minimization problem as follow:

$$(\ell^1_s): \ \hat{\alpha}_1 = \arg\min \|\alpha\|_1 \text{ subject to } \|y - X\alpha\|_2 \le \varepsilon, \tag{4}$$

where $\varepsilon$ is a given tolerance.

For a given test sample $y$, SRC first computes its sparse representation coefficient $\hat{\alpha}_1$ by solving the $\ell^1$-minimization problem (3) or (4), then determines the class of this test sample from its reconstruction error between this test sample and the training samples of class $i$,

$$r_i(\alpha) = \|y - X\delta_i(\alpha)\|_2. \tag{5}$$

For each $i$, $\delta_i(\alpha): \mathbb{R}^n \to \mathbb{R}^n$ is the characteristic function which selects the coefficients associated which the $i$-th class. Then the class $C(y)$ which the test sample $y$ belongs to is determined by

$$C(y) = \arg\min_i r_i(\alpha). \tag{6}$$

SRC is robust to noise and performs well for face recognition, it attracts much attention in recent years and boosts the research of sparsity based machine learning. Elhamifar and Vidal [13] proposed a more robust classification method using structured sparse representation, while Gao et al. [14] introduced a kernel version of SRC. Lu et al. [15] proposed a weighted sparse presentation method by utilizing the locality information. In [16], the $\ell^1$-graph was established by sparsely coding one sample over the other samples for clustering. In this paper, we focus on the sparse representation based dimensionality reduction problem, not the extension of SRC. A discriminative learning method is presented in the next section.

## 3. Optimized projections for sparse representation based classification

In this section, we consider the supervised DR problem. Considering a training sample $x$ (belonging to the $i$-th class) and its sparse representation coefficient $\alpha$ based on other training samples as a dictionary. Ideally, the entries of $\alpha$ are zero except those associated with the $i$-th class. In many practical face recognition scenarios, the training sample $x$ could be partially corrupted or occluded, or sometimes the training samples are not enough to represent the given sample. In these cases, the residual associated with the $i$-th class $r_i(\alpha)$ may be not small enough, and may produce an erroneous predict. Thus, the Optimized Projections for Sparse Representation based Classification (OP-SRC) is proposed which aims to seek a linear projection matrix such that in the transformed low-dimensional space, the within-class reconstruction residual is as small as possible and simultaneously the between-class reconstruction residual is as large as possible. SRC will perform better in projected subspace.

Let $P \in \mathbb{R}^{m \times d}$ be the optimized projection matrix with $d \ll m$. The data matrix in the original input space $\mathbb{R}^m$ are mapped into a $d$-dimensional space $\mathbb{R}^d$, that is, $Y = P^T X$. For each training sample $y_{ij} = P^T x_{ij}$ from $Y$ in the transformed $d$-dimensional space $\mathbb{R}^d$, by solving the extended $\ell^1$-minimization problem (4), we obtain its sparse coding coefficient $\alpha_{ij}$ by using the remaining training samples as a dictionary. Based on the decision rule of SRC, we define the within-class residual matrix as follows

$$\tilde{R}_W = \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n_i}(y_{ij} - Y\delta_i(\alpha_{ij}))(y_{ij} - Y\delta_i(\alpha_{ij}))^T. \tag{7}$$

The between-class residual matrix is defined as follow

$$\tilde{R}_B = \frac{1}{n(c-1)}\sum_{i=1}^{c}\sum_{j=1}^{n_i}\sum_{l \ne i}(y_{ij} - Y\delta_l(\alpha_{ij}))(y_{ij} - Y\delta_l(\alpha_{ij}))^T. \tag{8}$$

The total residual matrix is defined as follow

$$\tilde{R}_T = \frac{n\tilde{R}_W + n(c-1)\tilde{R}_B}{nc} \tag{9}$$

$$\tilde{R}_T = \frac{1}{nc}\sum_{i=1}^{c}\sum_{j=1}^{n_i}\sum_{l=1}^{c}(y_{ij} - Y\delta_l(\alpha_{ij}))(y_{ij} - Y\delta_l(\alpha_{ij}))^T. \tag{10}$$

To make SRC perform well on training data, we expect that the within-class residual is as small as possible and simultaneously the between-class residual is as large as possible. Therefore, we can choose to maximize the following criterion [17]

$$J(P) = \mathrm{tr}(\beta\tilde{R}_B - \tilde{R}_W), \tag{11}$$

where $\beta$ is the weight parameter which balances the between-class and within-class residual information. Since $P$ is a linear mapping, it is easy to show $\tilde{R}_W = P^T R_W P$ and $\tilde{R}_B = P^T R_B P$, where

$$R_W = \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n_i}(x_{ij} - X\delta_i(\alpha_{ij}))(x_{ij} - X\delta_i(\alpha_{ij}))^T, \tag{12}$$

$$R_B = \frac{1}{n(c-1)} \sum_{i=1}^{c} \sum_{j=1}^{n_i} \sum_{l \neq i} (x_{ij} - X\delta_l(\alpha_{ij}))(x_{ij} - X\delta_l(\alpha_{ij}))^T. \quad (13)$$

So, we have

$$J(P) = \mathrm{tr}(P^T(\beta R_B - R_W)P). \quad (14)$$

In order to avoid degenerate solutions, we additionally require that $P$ is constituted by the unit vectors, *i.e.* $P=[p_1,\ldots,p_d]$ and $p_k^T p_k = 1$, $k=1,\ldots,d$. One may use other constraints. For example, we can require $\mathrm{tr}(P^T R_W P)=1$ and then maximize $\mathrm{tr}(P^T R_B P)$. The motivation by using the constraint $p_k^T p_k = 1$ is that it will result to an orthogonal projection, which may help preserve the shape of the data distribution [18]. Thus, the objective function can be recast as the following optimization problem:

$$\max \sum_{k=1}^{d} p_k^T(\beta R_B - R_W)p_k$$
$$\text{subject to} \quad p_k^T p_k = 1, k = 1,\ldots,d. \quad (15)$$

We can use the Lagrange multipliers to transform the above objective function to include the constraint

$$L(p_k,\lambda_k) = \sum_{k=1}^{d} p_k^T(\beta R_B - R_W)p_k - \lambda_k(p_k^T p_k - 1). \quad (16)$$

The optimization is performed by setting the partial derivative of $L$ with respect to $p_k$ to zero

$$\frac{\partial L}{\partial p_k} = (\beta R_B - R_W - \lambda_k I)p_k = 0, \quad k=1,\ldots,d. \quad (17)$$

Then we obtain

$$(\beta R_B - R_W)p_k = \lambda_k p_k, \quad k=1,\ldots,d, \quad (18)$$

which means that the $\lambda_k$'s are the eigenvalues of $\beta R_B - R_W$ and the $p_k$'s are the corresponding eigenvectors. Thus

$$J(P) = \sum_{k=1}^{d} p_k^T(\beta R_B - R_W)p_k = \sum_{k=1}^{d} \lambda_k p_k^T p_k = \sum_{k=1}^{d} \lambda_k. \quad (19)$$

Therefore, $P$ is composed of the first $d$ largest eigenvectors of $\beta R_B - R_W$ and $J(P)$ is maximized.

The solution of the optimization problem (15) has the following property:

**Proposition 1.** *The columns of the optimal the solution $P$ to the optimization problem (15) are orthogonal, that is, $p_i^T p_j = 0$, for any $i \neq j$, and $p_i^T p_i = 1$.*

It is easy to prove the orthogonality of solution $P$ due to the symmetry of $\beta R_B - R_W$. Thus, OP-SRC is an supervised orthogonal projection method which may preserve more discriminative information for classification, especially for the SRC method.

## 4. Experimental verification

In this section, we investigate the performance of our proposed OP-SRC method for face representation and recognition. The system performance is compared with PCA [3], LDA [4], MMC [17], SPP [8] and SRC-DP [9]. PCA and LDA are two most popular linear methods in FR. MMC is a variant of LDA without dimension limitation. SPP and SRC-DP are two new methods corresponding to sparse representation. Similar to SPP and SRC-DP, we first perform PCA to reduce the dimension before implementing OP-SRC. Finally, SRC is employed for classification.

### *4.1. Data sets and experimental settings*

We test our proposed method on three popular face databases, including Yale [4], ORL [19] and UMIST [20]. There are wide-range variations, including pose, illumination, and gesture alterations existing in these databases. In our experiments, we randomly select part of the images per class for training (*i.e.*4, 5, 6, and 7 of 11 images per subject for Yale, 4, 5, 6 and 7 of 10 images per subject for ORL and 6, 8, 10 and 12 of about 29 images per subject for UMIST), and the remainder for test. In particular, with the given training set, the projection $P$ is learned by PCA, LDA, MMC, SPP, SRC-DP and OP-SRC, respectively, and the test samples are subsequently transformed by the learned projection. Then specific classifier is employed to evaluate the recognition rates on the test data, and SRC is used in this paper.

In the experiments, the images are cropped to a size of $32 \times 32$, and the gray level values of all images are rescaled to [0,1]. Twenty training/test splits are randomly generated and the average classification accuracies over these splits are reported in the tables and figures.

The SPAMS package [21,22] is used for solving the extended $\ell^1$-minimization problem (4). In our experiments, we experimentally set $\varepsilon = 0.05$ (refer to (4)) which usually leads SRC to better performance, and set $\beta = 0.25$ (refer to (14)) by searching in a large range of candidates.

### *4.2. Yale database*

The Yale database contains 165 gray scale images of 15 individuals. It was constructed at the Yale Center for Computational Vision and Control. The images demonstrate variations in lighting condition, facial expression (normal, happy, sad, sleepy, surprised, and wink). Fig. 1 shows some samples of two subjects of the Yale database. A random subset with $l$ (=4, 5, 6, 7) images per individual is taken with labels to form the training set, and the rest of the database is considered to be the test set. For each given $l$, we report the average of the recognition accuracies over 20 random splits. Notice that LDA is different from other methods because the maximal number of dimension is less than the number of class $c$ [4]. MMC is a variant of LDA without the dimension limitation.

In general, the performance of all these methods varies with the number of dimensions. We show the best results and the

**Fig. 1.** Samples of two subjects from the Yale database.

optimal dimensions obtained by PCA, LDA, MMC, SPP, SRC-DP and OP-SRC in Table 1, including the mean of accuracies as well as the standard deviations.

From Table 1, it can be found that OP-SRC obtains the highest recognition rates in all cases. Fig. 2 shows the plots of accuracy rates versus reduced dimensions. Note that, when the dimension of feature continues to increase, the performance of the OP-SRC algorithm decreases and has the same accuracy with PCA on the highest dimension. In this case, the obtained optimized projection matrix $P$ is square and orthogonal, that is $P^TP=PP^T=I$. Thus, $\|P^Tx-P^TX\alpha\|_2=\|x-X\alpha\|_2$. The sparse representation coefficient in the transformed space will be the same as in the subspace projected by PCA. Thus, they always obtain the same recognition result.

### 4.3. ORL database

The ORL database consists of 10 face images from 40 subjects for a total of 400 images, with some variations in poses, facial expressions and details. Some images were captured at different times and had different variations including expression (open or closed eyes, smiling or nonsmiling) and facial details (glasses or no glasses). The images were taken with a tolerance for some tilting and rotation of the face up to 20°. Fig. 3 shows some samples of two subjects of the ORL database. A random subset with $l$ (=4, 5, 6, 7) images per individual is taken with label to form the training set. The rest of the database is considered to be the test set. The experimental protocol is the same as that on the Yale database. The recognition results are shown in Table 2 and Fig. 4.

From Table 2 and Fig. 4, we find that most dimensionality reduction methods perform well, since the variation of faces in the ORL database is limited. PCA is even more accurate than SRC-DP which is supervised. If the number of training data is small, *i.e.*4 and 5 samples of each subject for training, OP-SRC also performs worse than PCA in low-dimensional space, but much better in high-dimensional space. The sparse representation tends to be not correct when the data are limited in low-dimensional subspace. Notice SPP and SRC-DP have similar phenomenons, but OP-SRC performs better than SPP and SRC-DP in this case. The best accuracy of MMC is obtained in the low-dimensional subspace, which is similar to LDA.

### 4.4. UMIST database

The UMIST database contains 564 images of 20 individuals, each covering a range of poses from profile to frontal views. The subjects cover a range of race, sex and appearance. We use a

**Table 1**
Mean recognition rates (%) and standard deviations on the Yale database.

| | 4 Train | 5 Train | 6 Train | 7 Train |
|---|---|---|---|---|
| PCA | 64.67 ± 4.37(52) | 67.06 ± 2.91(64) | 72.53 ± 4.25(88) | 72.08 ± 5.51(64) |
| LDA | 71.71 ± 5.88(14) | 75.28 ± 4.09(14) | 80.40 ± 4.78(14) | 81.50 ± 4.29(14) |
| MMC | 71.10 ± 5.15(21) | 75.61 ± 2.96(43) | 81.40 ± 3.90(76) | 82.08 ± 4.52(91) |
| SPP | 60.71 ± 4.86(57) | 63.83 ± 4.45(72) | 67.60 ± 4.37(88) | 70.17 ± 4.55(104) |
| SRC-DP | 70.57 ± 4.87(29) | 72.44 ± 3.49(37) | 77.07 ± 4.21(34) | 77.25 ± 4.30(43) |
| OP-SRC | **75.76 ± 4.81(48)** | **79.44 ± 3.63(62)** | **83.33 ± 3.83(74)** | **85.25 ± 4.81(88)** |

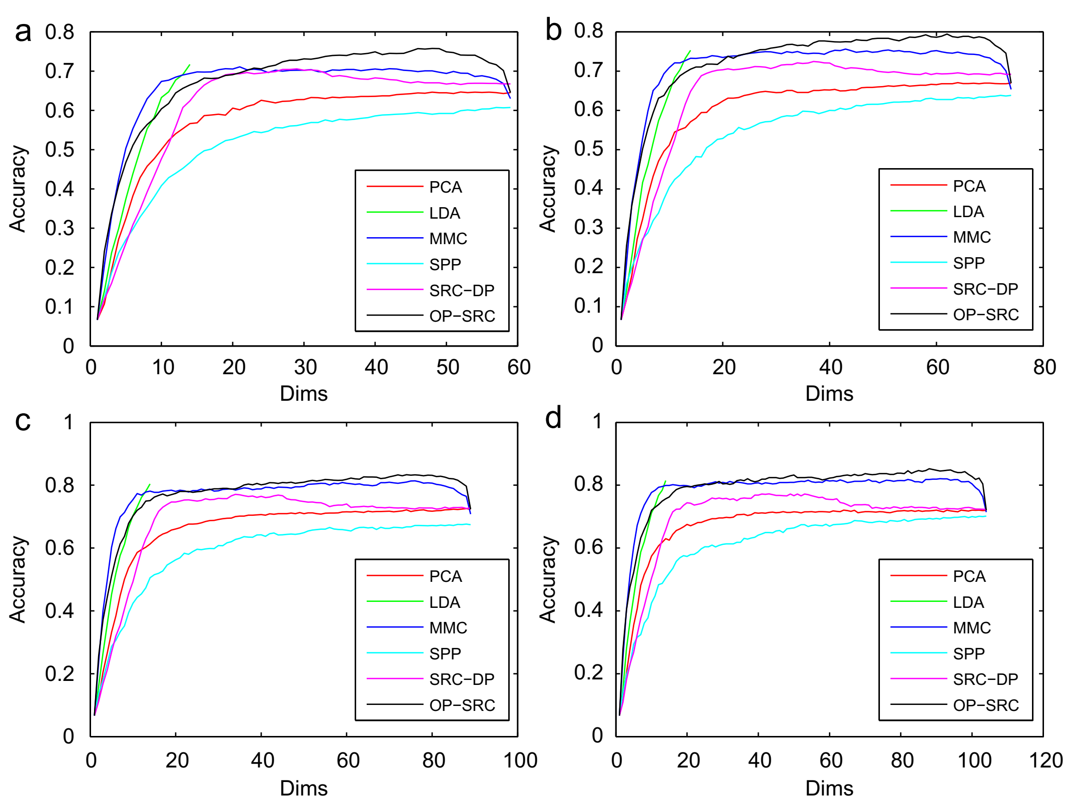


**Fig. 2.** Accuracy rates versus reduced dimensions on the Yale database: (a) 4 Train; (b) 5 Train; (c) 6 Train; and (d) 7 Train.

**Fig. 3.** Samples of two subjects from the ORL database.

**Table 2**
Mean recognition rates (%) and standard deviations on the ORL database.

| | 4 Train | 5 Train | 6 Train | 7 Train |
|---|---|---|---|---|
| PCA | 89.81 ± 1.93(127) | 92.13 ± 1.81(183) | 94.06 ± 1.82(193) | 95.42 ± 2.25(134) |
| LDA | 89.90 ± 1.95(39) | 93.03 ± 1.71(39) | 94.13 ± 1.92(39) | 95.04 ± 2.32(39) |
| MMC | 89.96 ± 1.71(37) | 92.50 ± 1.38(38) | 94.22 ± 2.07(34) | 94.47 ± 2.36(43) |
| SPP | 86.06 ± 1.84(108) | 88.70 ± 2.62(170) | 90.28 ± 3.14(180) | 92.17 ± 2.57(202) |
| SRC-DP | 88.77 ± 1.75(124) | 91.80 ± 1.75(131) | 92.88 ± 2.79(221) | 94.29 ± 2.15(190) |
| OP-SRC | **92.50** ± **1.68**(**153**) | **95.00** ± **1.75**(**195**) | **96.84** ± **1.47**(**224**) | **97.46** ± **1.28**(**255**) |

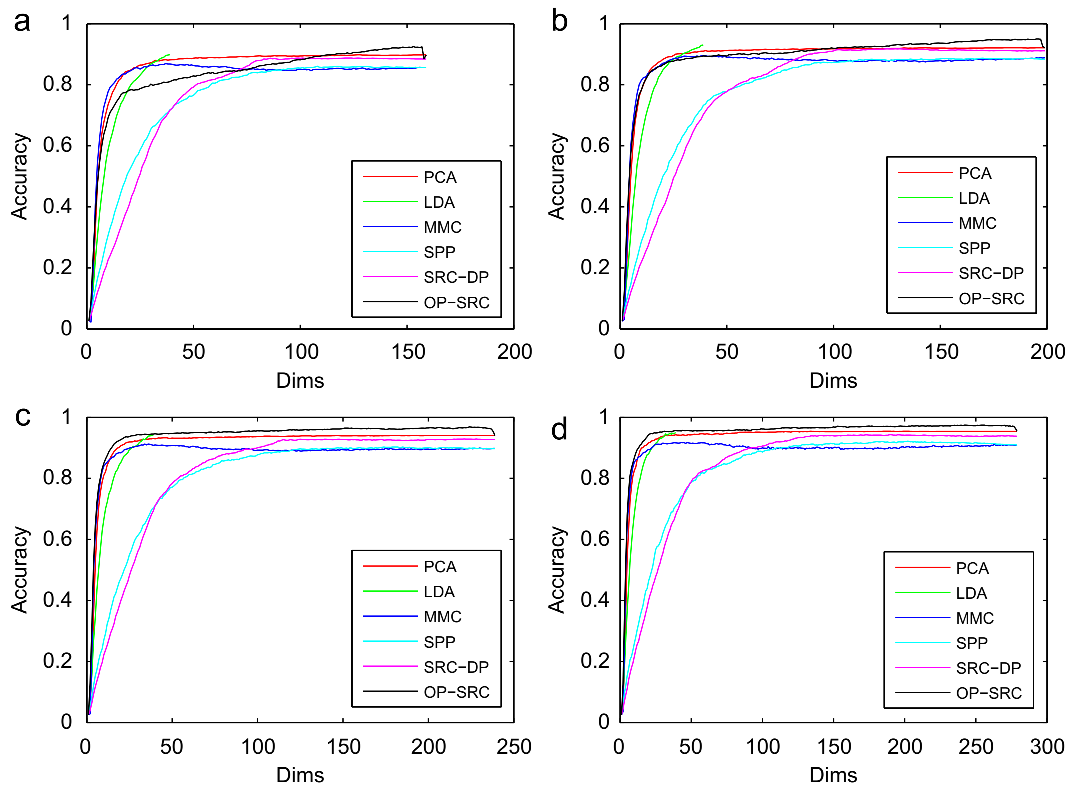


**Fig. 4.** Accuracy rates versus reduced dimensions on the ORL database: (a) 4 Train; (b) 5 Train; (c) 6 Train; and (d) 7 Train.

cropped version of the UMIST database that is publicly available at S. Roweis' web page[1]. Fig. 5 shows some images of two subjects of the UMIST database. We randomly select $l$ (=6, 8, 10, 12) images from each individual for training, and the rest for test.

Table 3 gives the best classification accuracy rates and the corresponding standard deviations of six algorithms under different sizes of the training set. Fig. 6 plots the recognition rates of these methods under different reduced dimensions when the size of training samples from each class is 6, 8, 10 and 12, respectively. From Table 3 and Fig. 6, we find that OP-SRC outperforms the other methods in different dimensions and different numbers of training data setting.

### 4.5. Discussions

Based on the results on the Yale, ORL and UMIST databases, we draw the following observations and discussions:

1. OP-SRC always outperforms PCA, SPP and SRC-DP on the Yale and UMIST databases, and also is more accurate than PCA when the subspace dimension exceeds a certain threshold on the ORL database. OP-SRC even performs better than LDA and

[1] http://cs.nyu.edu/roweis/data.html

**Fig. 5.** Samples of two subjects from the UMIST database.

**Table 3**
Mean recognition rates (%) and standard deviations on the UMIST database.

| | 6 Train | 8 Train | 10 Train | 12 Train |
|---|---|---|---|---|
| PCA | 88.35 ± 2.32(105) | 92.48 ± 3.13(125) | 95.92 ± 1.29(110) | 96.93 ± 1.84(85) |
| LDA | 83.54 ± 1.82(15) | 86.58 ± 3.17(15) | 91.15 ± 1.26(15) | 92.18 ± 1.68(15) |
| MMC | 87.52 ± 2.01(20) | 92.27 ± 3.24(15) | 95.89 ± 1.35(15) | 96.06 ± 2.45(15) |
| SPP | 83.08 ± 2.69(80) | 87.25 ± 2.64(105) | 91.17 ± 2.13(135) | 90.45 ± 2.78(155) |
| SRC-DP | 85.63 ± 2.20(75) | 89.42 ± 2.73(105) | 93.28 ± 1.50(120) | 93.07 ± 2.48(130) |
| OP-SRC | **89.41** ± **1.93**(**115**) | **93.93** ± **2.98**(**105**) | **97.44** ± **1.19**(**105**) | **98.00** ± **1.57**(**120**) |

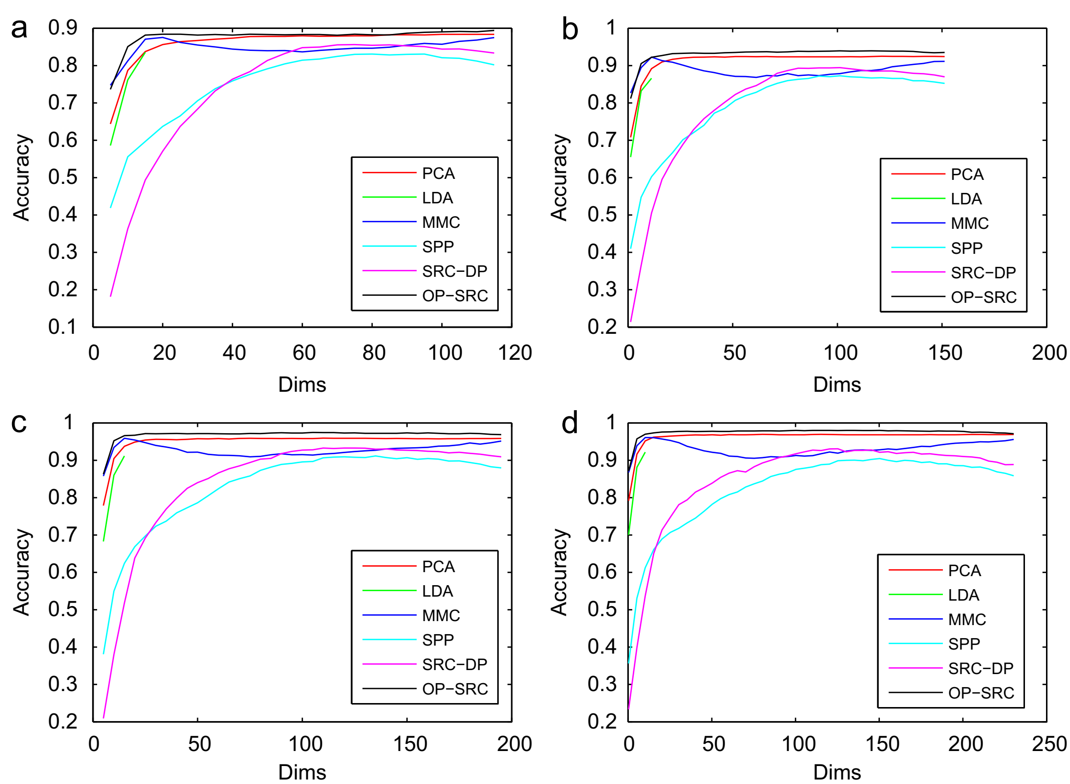


**Fig. 6.** Accuracy rates versus reduced dimensions on the UMIST database: (a) 6 Train; (b) 8 Train; (c) 10 Train; and (d) 12 Train.

MMC in the low-dimensional subspace on the ORL and UMIST databases. The top average recognition rates of OP-SRC are much higher than PCA, LDA, MMC, SPP and SRC-DP on these three databases. The superior of OP-SRC comes from its orthogonality and it matches well with the SRC algorithm.

2. Similar to other dimensionality reduction methods, the recognition accuracy of OP-SRC first increases according to the dimensions, but decreases at last and obtains the same result as PCA on the highest dimension. This is because the data is first projected onto a PCA subspace, and the $\ell^2$-norm is invariant to orthogonal projection on the highest dimension.
3. From our experiments, we also find that OP-SRC is more efficient than SPP and SRC-DP which are spare coding based methods. It is more practical for real applications.

## 5. Conclusions

In this paper, based on the sparse representation, we propose a new algorithm called Optimized Projections for Sparse Representation based Classification (OP-SRC) for supervised dimensionality reduction. The optimized projections of SRC decreases the within-class reconstruction residual and simultaneously increases the between-class reconstruction residual which matches with SRC optimally in theory. The experimental results on three face databases clearly demonstrate that the proposed OP-SRC has much better performance than PCA, LDA, MMC, SPP and SRC-DP, and also it is more effective with respect to the sparse representation based classification.

## Acknowledgment

This work was supported by the grants of the National Science Foundation of China, nos. 60975005, 61005010, 60873012, 60805021, 60905023, 31071168 and 61133010.

## References


[1] A.K. Jain, R.P.W. Duin, J.C. Mao, Statistical pattern recognition: a review, IEEE Trans. Pattern Anal. Mach. Intell. 22 (1) (2000) 4–37.
[2] S.J. Raudys, A.K. Jain, Small sample-size effects in statistical pattern-recognition – recommendations for practitioners, IEEE Trans. Pattern Anal. Mach. Intell. 13 (3) (1991) 252–264.
[3] M.A. Turk, A.P. Pentland, Face recognition using eigenfaces, in: IEEE Conference on Computer Vision and Pattern Recognition, 1991, pp. 586–591.
[4] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.
[5] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.
[6] X.F. He, S.C. Yan, Y.X. Hu, P. Niyogi, H.J. Zhang, Face recognition using laplacianfaces, IEEE Trans. Pattern Anal. Mach. Intell. 27 (3) (2005) 328–340.
[7] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 210–227.
[8] L.S. Qiao, S.C. Chen, X.Y. Tan, Sparsity preserving projections with applications to face recognition, Pattern Recognition 43 (1) (2010) 331–341.
[9] J. Yang, D. Chu, Sparse representation classifier steered discriminative projection, in: International Conference on Pattern Recognition, 2010, pp. 694–697.
[10] E. Amaldi, V. Kann, On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems, Theor. Comput. Sci. 209 (1–2) (1998) 237–260.
[11] D.L. Donoho, For most large underdetermined systems of linear equations the minimal $\ell^1$-norm solution is also the sparsest solution, Commun. Pure Appl. Math. 59 (6) (2006) 797–829.
[12] E.J. Candes, J.K. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Commun. Pure Appl. Math. 59 (8) (2006) 1207–1223.
[13] E. Elhamifar, R. Vidal, Robust classification using structured sparse representation, in: IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 1873–1879.
[14] S. Gao, I.W.-H. Tsang, L.-T. Chia, Kernel sparse representation for image classification and face recognition, in: European Conference on Computer Vision, 2010, pp. 1–14.
[15] C.-Y. Lu, H. Min, J. Gui, L. Zhu, Y.-K. Lei, Face recognition via weighted sparse representation, J. Vis. Commun. Image Represent. URL: 〈http://dx.doi.org/10.1016/j.jvcir.2012.05.003〉.
[16] B. Cheng, J. Yang, S. Yan, Y. Fu, T.S. Huang, Learning with $\ell^1$-graph for image analysis, IEEE Trans. Image Process. 19 (Compendex) (2010) 858–866.
[17] H.F. Li, T. Jiang, K.H. Zhang, Efficient and robust feature extraction by maximum margin criterion, Adv. Neural Inf. Process. Syst. 16 (2004) 97–104 1621.
[18] D. Cai, X.F. He, J.W. Han, H.J. Zhang, Orthogonal laplacianfaces for face recognition, IEEE Trans. Image Process. 15 (11) (2006) 3608–3614.
[19] F.S. Samaria, A.C. Harter, Parameterisation of a stochastic model for human face identification, in: Proceedings of the Second IEEE Workshop on Applications of Computer Vision, 1994, pp. 138–142.
[20] D.B. Graham, N.M. Allinson, Characterizing virtual eigensignatures for general purpose face recognition, in: Face Recognition: From Theory to Applications vol. 163, 1998, pp. 446–456.
[21] J. Mairal, F. Bach, J. Ponce, G. Sapiro, Online learning for matrix factorization and sparse coding, J. Mach. Learn. Res. 11 (2010) 19–60.
[22] J. Mairal, F. Bach, J. Ponce, G. Sapiro, Online dictionary learning for sparse coding, in: International Conference on Machine Learning, vol. 382, 2009, pp. 689–696.


**Can-Yi Lu** received the B.S. degree in Information and Computing Science & Applied Mathematics from Fuzhou University (FZU), Fuzhou, China, in 2009. Now, he is a candidate of the master degree in Pattern Recognition & Intelligent Systems from University of Science and Technology of China (USTC), Hefei, China. His research interests include sparse representation and low rank based machine learning and applications.

**De-Shuang Huang** received the B.Sc., M.Sc. and Ph.D. degrees all in electronic engineering from Institute of Electronic Engineering, Hefei, China, National Defense University of Science and Technology, Changsha, China and Xidian University, Xian, China, in 1986, 1989 and 1993, respectively. During 1993–1997 period he was a postdoctoral student, respectively, in Beijing Institute of Technology and in National Key Laboratory of Pattern Recognition, Chinese Academy of Sciences, Beijing, China. In September 2000, he joined the Institute of Intelligent Machines, Chinese Academy of Sciences as the Recipient of Hundred Talents Program of CAS. In September 2011, he entered into Tongji University as Chaired Professor. From September 2000 to March 2001, he worked as Research Associate in Hong Kong Polytechnic University. From August to September 2003, he visited the George Washington University as visiting professor, Washington DC, USA. From July to December 2004, he worked as the University Fellow in Hong Kong Baptist University. From March 2005 to March 2006, he worked as Research Fellow in Chinese University of Hong Kong. From March to July 2006, he worked as visiting professor in Queen University of Belfast, UK. In 2007, 2008, 2009, he worked as visiting professor in Inha University, Korea, respectively. At present, he is the head of Machines Learning and Systems Biology Laboratory, Tongji University.

Dr. Huang is currently a Senior member of the IEEE. He has published over 200 papers. Also, in 1996, he published a book entitled Systematic Theory of Neural Networks for Pattern Recognition (in Chinese), which won the Second-Class Prize of the 8th Excellent High Technology Books of China, and in 2001 & 2009 another two books entitled Intelligent Signal Processing Technique for High Resolution Radars (in Chinese) and The Study of Data Mining Methods for Gene Expression Profiles (in Chinese), respectively. In addition, he was the Ph.D. advisor in the University of Science and Technology of China. His current research interest includes bioinformatics, pattern recognition and machine learning.