Corrupted and occluded face recognition via cooperative sparse representation

Zhong-Qiu Zhao [a,b,*], Yiu-ming Cheung [b], Haibo Hu [d], Xindong Wu [c]

[a] College of Computer Science and Information Engineering, Hefei University of Technology, China
[b] Department of Computer Science, Hong Kong Baptist University, Hong Kong, China
[c] Department of Computer Science, University of Vermont, USA
[d] Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hong Kong, China

ABSTRACT

In image classification, can sparse representation (SR) associate one test image with all training ones from the correct class, but not associate with any training ones from the incorrect classes? The backward sparse representation (bSR) which contains complementary information in an opposite direction can remedy the imperfect associations discovered by the general forward sparse representation (fSR). Unfortunately, this complementarity between the fSR and the bSR has not been studied in face recognition. There are two key problems to be solved. One is how to produce additional bases for the bSR. In face recognition, there is no other bases than the single test face image itself for the bSR, which results in large reconstruction residual and weak classification capability of the bSR. The other problem is how to deal with the robustness of the bSR to image corruption. In this paper, we introduce a CoSR model, which combines the fSR and the bSR together, into robust face recognition, by proposing two alternative methods to these two key problems: *learning bases* and *unknown faces help* to enrich the bases set of the bSR. Thereby, we also propose two improved algorithms of the CoSR for robust face recognition. Our study shows that our CoSR algorithms obtain inspiring and competitive recognition rates, compared with other state-of-the-art algorithms. The bSR with the proposed methods enriching the bases set contributes the most to the robustness of our CoSR algorithm, and *unknown faces* works better than *learned bases*. Moreover, since our CoSR model is performed in a subspace with a very low dimensionality, it gains an overwhelming advantage on time consumption over the traditional RSR algorithm in image pixel space. In addition, our study also reveals that the sparsity plays an important role in our CoSR algorithm for face recognition.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Recently, sparse representation (SR) based image classification has undergone a great development [1–7]. Given some training images with class labels, sparse representation selects several training images which give the most accurate and compact representation of one test image according to visual content similarity, and then classifies the test image by the reconstruction residual associated with each object class. The sparse representation classifier can be considered as a generalization of the nearest neighbor (NN) [8,9] or the nearest subspace (NS) [10]. The NN classifier approximates the test image with a single training image which is the most similar, while the NS classifier is based on *the strategy* of the best linear representation of the test image with all of the training images in each class. However, the NN can be easily affected by noise, especially for real applications, while the NS may not work well when classes are highly correlated to each other [3]. As a tradeoff between the NN and the NS, the sparse representation (SR) [3,11] can automatically select a small number of training images and approximate the test image with a weighted sum of the selected ones. The motivation of finding a sparse solution rather than a dense one lies in that a dense solution usually results in a large number of training images selected to approximately represent the test image, while these training images may come from various classes. Therefore, this dense solution is not especially informative for classification [3].

Robustness is an important and open problem in real face recognition systems since test face images are usually partially corrupted or occluded [12,13]. One of the most successful applications of sparse representation based image classification is to deal with corruptions and occlusions in face recognition [3,14]. Face recognition can be performed in the subspaces such as principal component analysis (PCA) [15,16], linear discriminant analysis (LDA) [17], and
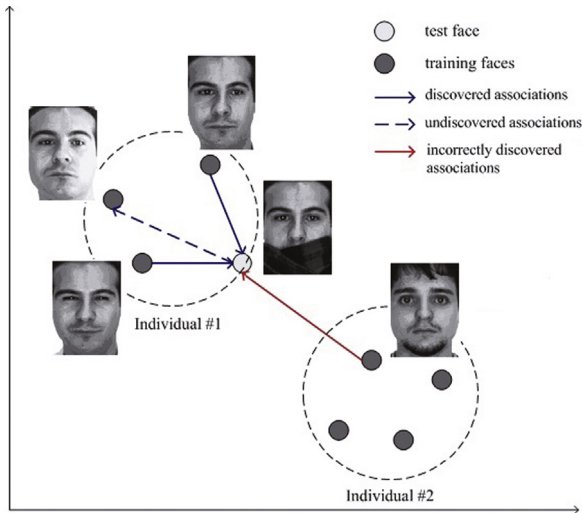
Fig. 1. A sketch of the SR bridging between test face and training faces.

locality preserving projection (LPP) [18] where the dimensionality is reduced below the number of training faces so that the approximation problem is underdetermined and simplified. However, Wright et al. [3] believed that noise on the original pixels cannot be eliminated by the subspace projections, and no bases or features are more spatially localized than the original image pixels themselves. So they proposed to perform face recognition with sparse representation in the original pixel space. In order to do that, besides training images, the bases set of robust sparse representation (RSR) is enlarged to reduce the reconstruction residual with the identity matrix which can linearly reconstruct any noise. We call this algorithm 'Pixel+RSR' in the following sections. In addition, Wang et al. proposed a sparse residue method for occluded face image reconstruction, which considers an occluded face image as the summation of a non-occluded face and a sparse occlusion [14]. Both of the above two methods can achieve satisfactory recognition accuracy for robust face recognition.

For face recognition, the SR bridges between the test face and the training faces from the same individual, and avoids associations between the test face and the training faces from the incorrect individuals. However, it is *NOT* definite that the training faces from the correct individuals can be all *selected*, while the ones from the incorrect individuals can be all excluded by the sparsity constraint, especially when the test image may contain noise. On the contrary, the case in Fig. 1 usually appears. Comparing to the general SR which represents each test image with training images and we call the forward sparse representation (fSR), the backward sparse representation (bSR), in an opposite direction, representing each training image with test images. These two SRs may contain some complementary information for classification. That is to say, the bSR can remedy the correct associations undiscovered by the fSR, and comparatively weaken the incorrect associations by strengthening the correct associations which are discovered by the fSR. The reason is that two opposite SRs are conditionally independent due to the sparsity of the SR and the diversity of images [4]. In [4], we validated the complementarity between the fSR and bSR, and proposed a cooperative sparse representation (CoSR) model for scene image annotation. Recently, Xu et al. [19] cast this idea into linear regression classification and ℓ_2-norm based representation methods.

In this paper, we will introduce the CoSR into robust face recognition and address two key problems. First, in the CoSR application in image annotation, all test images to be annotated can be utilized so that there are enough bases for the bSR to represent each training image with small residual. However, in face recognition, there is only one ready-made basis for the bSR, viz., the test face image itself. So how to construct and expand the

bases set, viz. dictionary, of the bSR to reduce the reconstruction residual is one problem to be addressed in this paper. Another key problem is how to deal with the corruption in test images. If the corruption or noise in any test image is large, the bSR tends to utilize newly added bases rather than the test image to represent the training image, even if the training image and the test image come from the same individual, which will result in smaller reconstruction error. This will lead to the failure of recognition of the test image. Imposing the noise similar to that in the test face on the learned bases of the bSR or using additional unknown faces with the similar noise as the expanded bases of the bSR can reduce the correlation between the additional bases and the training images, so that the test image is more probably selected to represent the training images from the correct individual. Thereby the robustness of the bSR is improved. The primary motivation of this paper is to extend the CoSR model to solve the robust face recognition problem, namely corrupted and occluded face recognition, and to improve the robustness and performance of the CoSR for robust face recognition by utilizing this robustness of the bSR to noise in face images.

The main contributions of this paper are two-fold. (1) There have been no previous work on the complementarity between the fSR and the bSR for face recognition, and no previous work to explore to compensate the general SR for face recognition. This paper introduces the CoSR model into the field of face recognition, obtaining very competitive recognition rates and showing high efficiency for robust face recognition. (2) This paper proposes two methods, namely '*learning bases*' and '*unknown faces help*', to expand the bases set of the bSR, which are very robust to noise or corruption in face images, especially when the noise or corruption is very large. (3) This paper also reveals that the sparse constraint is very important for the success of our CoSR model on face recognition.

2. Related work

For robust face recognition, the 'Pixel+RSR' algorithm which solves sparse representation in the original pixel space inevitably brings the problem of high time complexity due to large feature dimensionality and large number of bases. Traditional ℓ_1 minimization solution algorithms by linear programming (LP) [20] have shown their high computational complexity for real-world applications with large size of face images. Heuristic greedy algorithms such as orthogonal matching pursuit (OMP) [21] and least angle regression (LARS) [22], which are faster than using LP methods, can often fall into local optimum [23], and these sub-optimal solutions result in poor face recognition performance, even in the original pixel space. Newly developed algorithms such as gradient projection (GP) [24], augmented lagrange multiplier (ALM) [25], homotopy [26], and feature-sign search sign (FSS) [27] methods, which attempt to improve the solution efficiency of ℓ_1 minimization, however, cannot avoid the problem of high time complexity when working in the original pixel space. Though the computational complexities of GP and ALM methods are difficult to estimate exactly, the solutions all involve the matrix operations w.r.t. the dictionary of sparse representation which is a huge matrix when in pixel space. And the computational complexity of the homotopy algorithm is $O(d^2+dn)$ for each iteration [28], where d and n denote the feature dimensionality and the number of bases, respectively. Comparatively, our proposed CoSR algorithm works well in subspaces such as the PCA, LDA, and LPP, which will largely reduce time consumption.

In addition, it is worth mentioning that some studies [29,30] claimed that enforcing sparsity constraints does not help to improve the performance of object recognition and image classification. For face recognition, [31–33] also claimed that it is not the sparsity that improves the face recognition accuracy, and the simple regularized least square can attain equal or even higher accuracy while with much less time consumption than the ℓ_1-norm solution. These researches further give rise to the doubt about the necessity and significance of sparsity for image classification. However, in this paper, we will propose an efficient cooperative sparse representation (CoSR) algorithm for robust face recognition, where it is the ℓ_1-norm constraint that plays a vital role for the success of this model. The key factor which ensures the CoSR classifier to work well is that the two SRs in opposite directions are with difference or at least not so tightly correlated. The fSR associates the test image with a small subset of training images thanks to the sparse nature, while the bSR associates the test image with another small subset of training images. Due to the sparsity of the SRs, these two subsets may differ greatly from each other, which thereby results in relative independence between two opposite SRs. On the contrary, a non-sparse solution such as the ℓ_2-norm in [29–32] may make both approximations in two opposite directions associate the test image with a large number of or even all training images, which results in overlapping between two subsets and thereby causes greater dependence between two opposite approximations. In this paper, we will confirm this assumption by experimental comparison between ℓ_1-norm and ℓ_2-norm.

In the general SR models, the following reconstruction residual is usually used as the classification criterion [3,11]:

$$C(y) = \arg\min_i \| y - X\delta_i(\hat{a}) \|_2 \tag{1}$$

where y denotes the test signal, X the bases set, $\hat{a}$ the sparse vector solution, and $\delta_i(\hat{a})$ a new vector whose only nonzero elements are the ones in $\hat{a}$ associated with class i. However, for the bSR which represents training images with test ones, the reconstruction residual of sparse representation serves the training faces instead of the test one, so the above classification criterion becomes invalid. In addition, the criterion based on the reconstruction residual is inconvenient for the weighted fusion of the fSR and the bSR. So in this paper, the proportion of the positive components in sparse coefficients is used instead, which has also been verified even slightly better than the reconstruction residual for the single fSR [4]. Similar to the reconstruction residual criterion, the new classification criterion is based on the assumptions that for the fSR, the test image belonging to one certain class can be thoroughly represented by the training images from the same class, and that for the bSR, the test image can only be selected to use for representing the training images from the same class rather than the other classes. Obviously, a non-sparse solution such as the ℓ_2-norm violates these assumptions much more than the ℓ_1-norm solution. This is another evidence for the significance of sparsity for our CoSR algorithm on face recognition.

3. Extend cooperative sparse representation to robust face recognition

3.1. Face recognition based on forward sparse representation

Given m bases $X = [x_1, x_2, ..., x_m] \in \mathbb{R}^{d \times m}$ and a new test signal $y \in \mathbb{R}^d$, sparse representation aims to find a sparse vector $a \in \mathbb{R}^m$ so that y can be linearly represented as [34,35]:

$$y = Xa + e \tag{2}$$

where the vector $e \in \mathbb{R}^d$ is the signal noise in real-world data and with a small norm value. This approximation problem can be solved by minimizing the following objective function:

$$\min_a \| y - Xa \|^2 + \lambda \| a \|_1 \tag{3}$$

where $\| \cdot \|$ denotes the ℓ_2-norm, and $\| \cdot \|_1$ denotes the ℓ_1-norm; the first term is to minimize the reconstruction residual, while the second one is to control the sparsity of the vector a with the tradeoff parameter λ. A larger λ usually results in a sparser a solution.

We extend the formulation of the fSR for two-class image annotation in [4] to multi-class face recognition. A basic problem of face recognition can be described as follows. Given a set of training face images $X \in \mathbb{R}^{d \times m}$ with their class labels $L = [l_1, l_2, ..., l_m]$ and a test face image $y \in \mathbb{R}^d$, where $l_i \in \{1, 2, ..., c\}$, the face recognition task is to assign a class label to the test image. The training set can be divided into c subsets: $X_i = [x_{i,1}, x_{i,2}, ..., x_{i,m_i}]$, where $i = 1, 2, ..., c$, $\sum_i m_i = m$, and all subsets X_i are mutually exclusive to each other, consisting of the training samples from class i. So Eq. (2) can be written as:

$$y = \sum_{i=1}^{c} \sum_{j=1}^{m_i} \alpha_{i,j} x_{i,j} + e \tag{4}$$

where $\alpha_{i,j}$ are the elements in the sparse vector a associated with $x_{i,j}$. If the test image y belongs to one certain class k, then in an ideal situation, y can be thoroughly represented by a few training samples from X_k, while the elements associated with other classes than k should be all zeros, namely, $\alpha_{i,j} = 0$ for all $i \neq k$. However, in most situations, y is represented by both training samples from class k and those from other classes to minimize the reconstruction residual when there is noise or concept correlation in images. Even so, the *accumulation value* of the elements in a associated with $x_{k,j}$, namely $\alpha_{k,j}$ for all $j = 1, ..., m_k$, should be larger than that of the elements associated with $x_{i,j}$ for any one $i \neq k$ with all $j = 1, ..., m_i$. So we can have the following classification criterion for the fSR:

$$C(y) = \arg\max_i \frac{\sum_j \alpha_{i,j}}{\sum_i \sum_j \alpha_{i,j}} \tag{5}$$

The value of $\sum_j \alpha_{i,j} / \sum_i \sum_j \alpha_{i,j}$ can also be assigned to y as the score or degree with which it belongs to class i. Fig. 2 shows an example of the fSR for face recognition.

Note that the reconstruction residual $\| e \|^2 = \| y - Xa \|^2$, viz. the value of the first term in Eq. (3), is invariant when simultaneously multiplying X by a scalar and a by the inverse of the scalar, and the elements $\alpha_{i,j}$ in a are used as the similarities between the test sample y and the corresponding training samples. So the ℓ_2 norm of each basis $x_{i,j}$ should be equally normalized (usually as a unit vector) to ensure scoring by Eq. (5) to be fair in the same level.

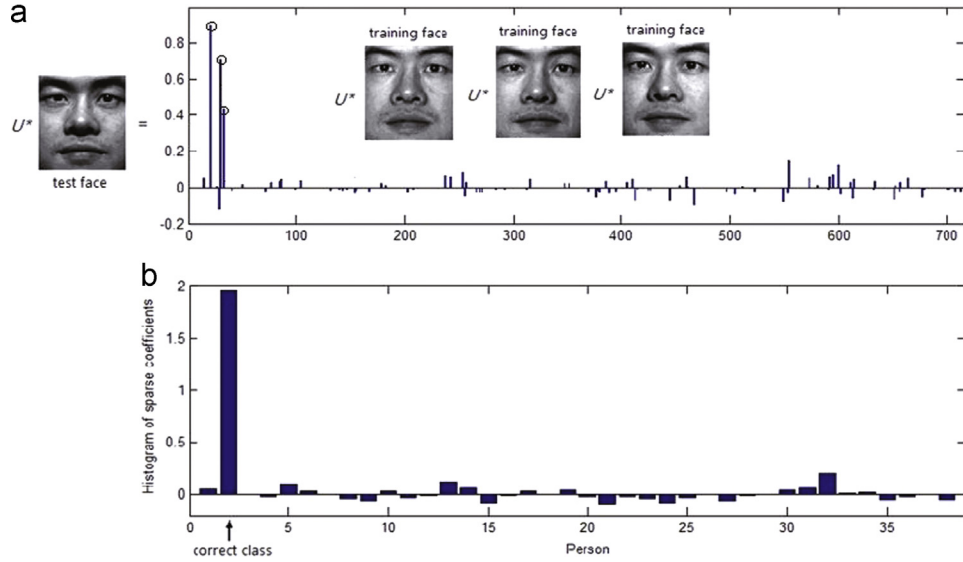## 3.2. Face recognition based on backward sparse representation

Provided a dictionary consisting of the test image y and a set of bases $D = [d_1, d_2, ..., d_n] \in \mathbb{R}^{d \times n}$, any training image $x_{i,j}$, which is the jth sample from the ith individual, can be linearly represented by sparse representation as:

$$x_{i,j} = [y, D]\begin{bmatrix} \beta_{i,j} \\ \hat{b} \end{bmatrix} + e \tag{6}$$

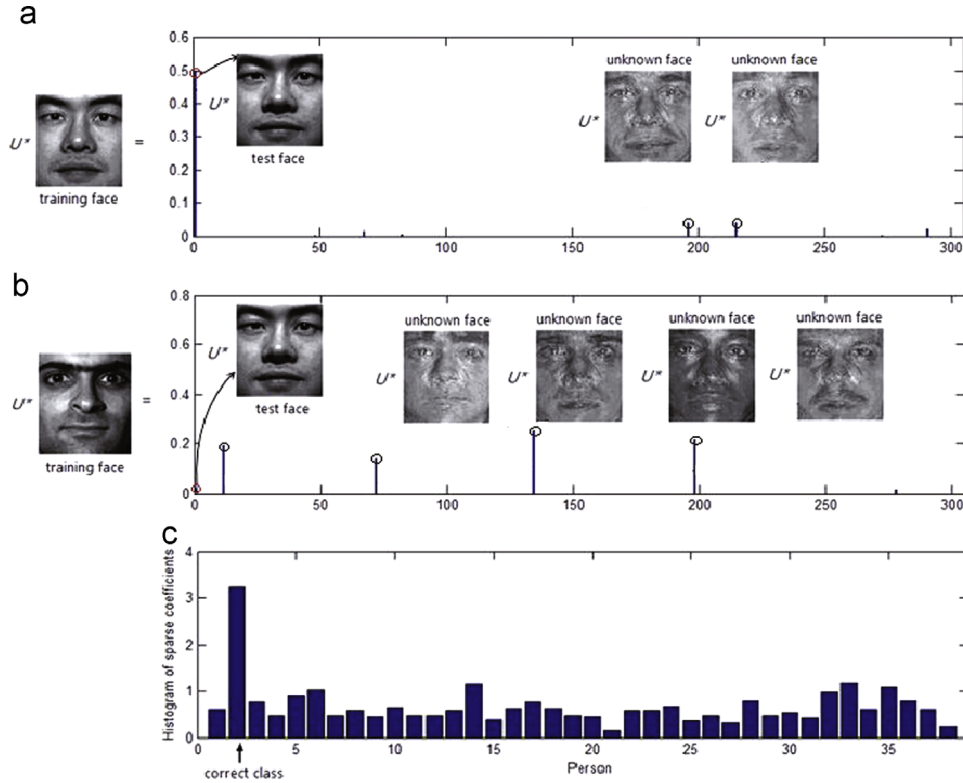where $\beta_{i,j}$ can be solved by minimizing the following objective function:

$$\min_{\begin{bmatrix} \beta_{i,j} \\ \hat{b} \end{bmatrix}} \left\| x_{i,j} - [y, D]\begin{bmatrix} \beta_{i,j} \\ \hat{b} \end{bmatrix} \right\|^2 + \lambda \left\| \begin{bmatrix} \beta_{i,j} \\ \hat{b} \end{bmatrix} \right\|_1 \tag{7}$$

Considering all training images $X = [x_{1,1}, ..., x_{1,m_1}, ..., x_{c,1}, ..., x_{c,m_c}]$,

Fig. 2. An example of the fSR for face recognition. (a) A test face image is sparsely represented by 717 training faces from various individuals, and the elements associated with the same individual as the test face correspond to the largest values. (b) The histogram of the sparse coefficients associated with the training faces for every class, by which the classification is performed. The largest histogram indicates the correct class.

Fig. 3. An example of the bSR for face recognition. (a) A training face image belonging to the same individual as the test image is sparsely represented by the dictionary consisting of the test face and 300 learned bases. (b) A training face image belonging to one individual different from that of the test face is sparsely represented by the dictionary consisting of the test face and 300 unknown faces. (c) The histogram of the sparse coefficients associated with the test image for every class, by which the classification is performed. The largest histogram indicates the correct class.

we obtain the vector

$$b = [\beta_{1,1}, \ldots, \beta_{1,m_1}, \ldots, \beta_{c,1}, \ldots, \beta_{c,m_c}] \tag{8}$$

where $\beta_{i,j}$ denotes the connection weight between the test image y and the training image $x_{i,j}$. Different from the case of the fSR, if y belongs to class k, in an ideal situation, y can only be selected to represent the training images from class k, so $\beta_{i,j} = 0$ for all $i \neq k$. However, there are exceptions due to noise or concept correlation. Even so, for all $i \neq k$, the test image y should correspond to a small value of $\beta_{i,j}$ if the bases D can

represent any training image by themselves with small reconstruction residual. Thereby, the *accumulation value* of the elements in b associated with $x_{k,j}$, namely $\beta_{k,j}$ for all $j = 1, \ldots, m_k$, should be larger than that of the elements associated with $x_{i,j}$ for any one $i \neq k$ with all $j = 1, \ldots, m_i$, and we can have the following classification criterion for the bSR:

$$C(y) = \arg \max_i \frac{\sum_j \beta_{i,j}}{\sum_i \sum_j \beta_{i,j}} \tag{9}$$

The value of $\sum_j \beta_{i,j} / \sum_i \sum_j \beta_{i,j}$ can also be assigned to y as the score or

degree with which it belongs to class *i*. Fig. 3 shows an example of the bSR for face recognition.

3.3. Fusion of forward and backward sparse representations

A simple average fusion can be adopted to combine the output score of the fSR with that of the bSR for robust face recognition. The weighted-averaging fusion of two output scores is as follows:

$$\gamma_{i,j} = (1-w)\frac{\sum_j \alpha_{i,j}}{\sum_i \sum_j \alpha_{i,j}} + w\frac{\sum_j \beta_{i,j}}{\sum_i \sum_j \beta_{i,j}} \tag{10}$$

where w is the fusion weight which can be tuned within the range [0, 1]. Note that we normalize $\alpha_{i,j}$ and $\beta_{i,j}$ to be in the range of [0,1] by dividing it by the sum of all elements in the vectors a and b, respectively, so that they both exactly express the correlations between the training images and the test image and then are mixed in the same order of magnitude.

The connection weights $\gamma_{i,j}$ then depict the relevance degrees between the training faces and the test face. So finally we have the following classification criterion for the fusion of two SRs:

$$C(y) = \arg\max_i \frac{\sum_j \gamma_{i,j}}{\sum_i \sum_j \gamma_{i,j}} \tag{11}$$

4. Additional bases for the bSR

In face recognition, there is only one basis, viz., the test face image itself, available for the bSR, and the performance of the bSR will affect the effectiveness of the combination of the fSR and the bSR. So it is necessary to construct additional bases to reduce the reconstruction residual and to improve the performance of the bSR. We propose two methods, namely 'Learning bases' and 'Unknown faces help', to enlarge the bases set of the bSR.

4.1. Learning bases for the bSR

Obviously, an enough number of *random bases* are qualified as long as they are overcomplete. However, studies have shown that given a set of training samples X, alternatively optimizing the bases D and the sparse coefficients S can reduce the objective value and improve the classification performance of sparse representation [36,37,27]. In this paper, the objective function can be defined as follows:

$$\min_{D,S} \sum_{i=1}^{m} \|x_i - [y,D]s_i\|^2 + \lambda \|s_i\|_1$$

subject to $\|d_i\|^2 \le 1, \forall i = 1,...,n$ (12)

where $D = [d_1,...,d_n]$. Though the optimization problem in (12) is not convex when simultaneously optimizing D and S, it is convex when alternatively optimizing one of them with the other fixed. Assuming $\hat{s}_0$ is a vector whose elements are the coefficients corresponding to the test sample y for all training samples X, the bases D can be learned with the fixed sparse coefficients by the following minimization function:

$$\min_D \left\| X - [y,D]\begin{bmatrix} \hat{s}_0 \\ S \end{bmatrix} \right\|^2$$

subject to $\|d_i\|^2 \le 1, \forall i = 1,...,n$ (13)

This least squares problem with quadratic constraints can be solved with gradient descent algorithm [38]. However, it is an iterative projection and we need to compute the bases repeatedly with varying sparse coefficient values to optimize it. So in this paper, we use the Lagrange dual algorithm [27] instead to optimize

the bases, which is much more efficient. Considering the Lagrangian:

$$\mathcal{L}(D, \vec{\lambda}) = \text{trace}\left(\left(X - [y,D]\begin{bmatrix} \hat{s}_0 \\ S \end{bmatrix}\right)^T \left(X - [y,D]\begin{bmatrix} \hat{s}_0 \\ S \end{bmatrix}\right)\right) + \sum_{i=1}^{k} \lambda_i \left(\|d_i\|^2 - 1\right) \tag{14}$$

where λ_i is dual variables, and $\vec{\lambda} = [\lambda_1,...,\lambda_k]$. We can obtain the following Lagrange dual by minimizing over D analytically:

$$\mathcal{D}(\vec{\lambda}) = \min_D \mathcal{L}(D, \vec{\lambda}) = \text{trace}\left((X - y\hat{s}_0)^T (X - y\hat{s}_0) - (X - y\hat{s}_0)S^T \left(SS^T + \Lambda\right)^{-1} \left((X - y\hat{s}_0)S^T\right)^T - \Lambda\right) \tag{15}$$

where $\Lambda = diag(\vec{\lambda})$. Then we compute the partial gradient and Hessian of $\mathcal{D}(\vec{\lambda})$ over λ_i as follows.

$$\frac{\partial \mathcal{D}(\vec{\lambda})}{\partial \lambda_i} = \|(X - y\hat{s}_0)S^T(SS^T + \Lambda)^{-1}I_i\|^2 - 1 \tag{16}$$

$$\frac{\partial^2 \mathcal{D}(\vec{\lambda})}{\partial \lambda_i \partial \lambda_j} = -2\left(\left(SS^T + \Lambda\right)^{-1}\left((X - y\hat{s}_0)S^T\right)^T (X - y\hat{s}_0)S^T \left(SS^T + \Lambda\right)^{-1}\right)_{i,j} \times \left(\left(SS^T + \Lambda\right)^{-1}\right)_{i,j} \tag{17}$$

where $I_i \in \mathbb{R}^{n \times 1}$ is the *i*-th unit vector. Then we optimize the Lagrange dual (15) using the conjugate gradient method. After maximizing $\mathcal{D}(\vec{\lambda})$, we obtain the optimal dictionary D as follows:

$$D^T = (SS^T + \Lambda)^{-1}SX^T - (SS^T + \Lambda)^{-1}S\hat{s}_0^T y^T \tag{18}$$

This dual algorithm is more efficient because it deals with much fewer optimization variables than the original. Moreover, in order to save the test time consumption, we only need to compute $(SS^T + \Lambda)^{-1}SX^T$ and $(SS^T + \Lambda)^{-1}S$ once for all test samples.

In addition, when dealing with robust face recognition problem where the noise in test image is large, the bSR may tend to utilize the newly added bases rather than the test image itself to represent any training image, if the expanded bases are learned well based on the reconstruction residual minimization strategy, whether the training image and the test image belong to the same individual or not. The reason lies in the tendency to small reconstruction residual of sparse representation. This phenomenon will result in the failure of the bSR classification. Adding similar noise $\hat{n}$ into the learned bases may be one solution to this problem. The similar noise can be approximated as follows:
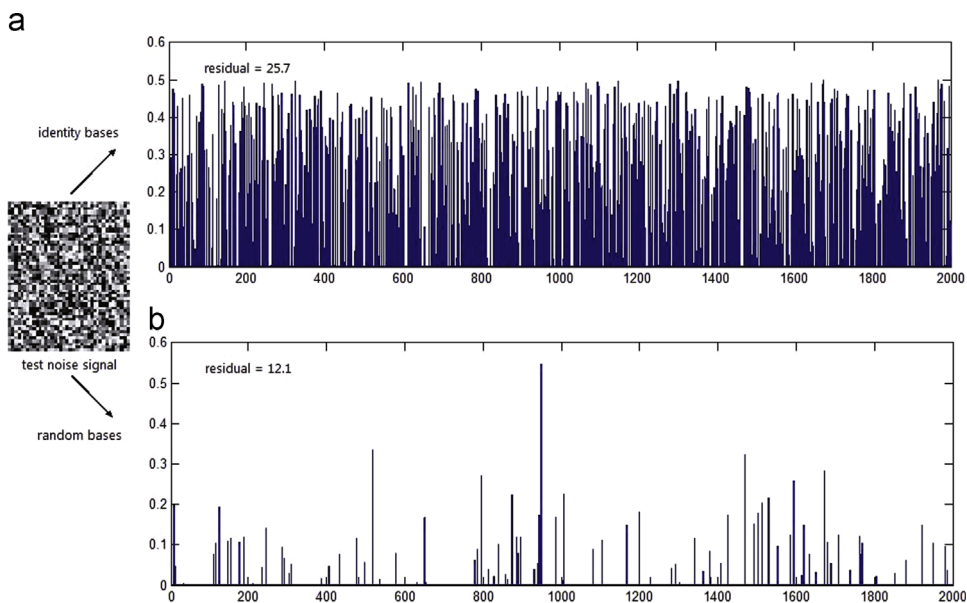
$$\hat{n} = y - \bar{x} \tag{19}$$

where $\hat{n}$ is the approximation of the noise, $\bar{x}$ is the average of all training images and $\bar{x} = \frac{1}{m}\sum_{i=1}^{m}x_i$. Then the additional bases are assigned as:

$$D \leftarrow D + \xi\hat{n}\mathbf{1} \tag{20}$$

where $\mathbf{1} \in \mathbb{R}^{1 \times n}$ is an all one vector, and ξ is a tradeoff parameter whose value is between 0 and 1.

4.2. Unknown faces help the bSR

Besides learned bases, there is another easier way for real face recognition system to expand the bases of the bSR. Generally, in real applications, additional face images without class labels can be easily collected from volunteers. We call these face images *unknown faces* in the following parts of this paper. Serving as the additional bases of the bSR, viz. *D*, *unknown faces* hold the advantage over *learned bases* that their collection surroundings are

Fig. 4. Comparison between identity bases and random bases for sparse representation on a noise signal of 50 by 40. The toy test shows that identity bases (a) result in a much denser representation with a much larger residual than random bases (b).

more similar to those of training faces including illumination, pose, and so on, even when they are from different individuals, so that the bSR can be solved more easily with small reconstruction residual. Another advantage of *unknown faces* may lie in that they naturally contain characteristics, including the noise, of the test image which may largely vary with the change of collection surroundings and devices so that the expanded bases *D*, viz. *unknown faces*, are selected and utilized by the bSR together with the test image at a fair level and thereby the bSR is robust to the noise.

5. Improvements of the CoSR for robust face recognition

5.1. Expanding the dictionary of the fSR

In this subsection, we consider improving our CoSR model by enhancing the fSR for robust face recognition. In order to deal with the robustness in face recognition, Wright et al. [3] proposed to expand the bases set with an identity matrix which can linearly reconstruct any corruption or occlusion in face images. However, since each identity vector has only one non-zero element and thereby each dimension of signal needs one independent identity vector for reconstruction, sparse representation with the identity matrix cannot efficiently reconstruct any noise signal with small residual, especially when the feature dimension is large. In an extreme case, almost all unit vectors in the identity matrix are needed to reconstruct the noise signal with small residual. Comparatively, a set of random bases *with normal distribution* may contain ones which are similar to the test noise signal, which will facilitate the sparse representation with small residual. Fig. 4 shows an example of the comparison between identity bases and random bases for sparse representation on a test noise signal of 50 by 40, which indicates that the identity bases result in a nonsparse solution in spite of the sparse constraint, while the random bases *with normal distribution* result in a much sparser representation and a much smaller residual. This phenomenon may be a good explanation to the classification performance of ℓ_1-norm which is similar to that of the ℓ_2-norm as reported in [32] or even worse in [31].

Thereby, we have the following formulation of the linear reconstruction for the test image *y*:

$$y = Xa + R\hat{a} + e \tag{21}$$

where *R* denotes a set of random bases with normal distribution,[1] $\hat{a}$ is the sparse vector for constructing the occlusion in face image. The approximation problem can then be solved by minimizing the following objective function:

$$\min_{a} \left\| y - [X\ R]\begin{bmatrix} a \\ \hat{a} \end{bmatrix} \right\|^2 + \lambda \left\| \begin{bmatrix} a \\ \hat{a} \end{bmatrix} \right\|_1 \tag{22}$$

5.2. Block partitioning

In many real applications, the occlusion on image pixels is unknown but the occluded pixels are known to be connected. Some work [3,39] has shown that partitioning the image into blocks and processing each block independently can improve face recognition since the redundant information from occluded blocks can be discarded by aggregating all blocks with voting [40]. In this paper, however, the voting strategy is not suitable for combining the soft outputs from the blocks as shown in Eq. (10). Therefore, for the convenience of the fusion of the fSR and bSR, we utilize the average fusion method instead of voting to combine all results of blocks, and modify Eq. (10) as follows:

$$\gamma_{i,j} = \frac{1-w}{K}\sum_{k=1}^{K}\left(\frac{\sum_{j}\alpha_{i,j}^k}{\sum_{i,j}\alpha_{i,j}^k}\right) + \frac{w}{K}\sum_{k=1}^{K}\left(\frac{\sum_{j}\beta_{i,j}^k}{\sum_{i,j}\beta_{i,j}^k}\right) \tag{23}$$

where *K* is the block number.

6. Experimental results and discussions

In this section, several experiments on two face recognition databases and some discussions are presented to demonstrate the effectiveness of our proposed CoSR model and to validate our claims in the previous sections. We evaluated the performance of

---

[1] Actually, uniform distribution also works, and produces the similar performance to that of normal distribution according to our experiments.

Table 1
Comparison of the recognition rates of the bSR between the clean learned dictionary and the noisy learned dictionary on Extended Yale B face database.

Corruption percentage	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%
Clean Lrd. Dict.	**0.883**	**0.905**	**0.912**	**0.861**	**0.759**	0.541	0.344	0.201	0.097	0.046
Noisy Lrd. Dict.	0.545	0.543	0.552	0.552	0.550	**0.556**	**0.570**	**0.594**	**0.671**	**0.450**

our CoSR algorithm with the bSR dictionary consisting of *learned bases* or *unknown faces*. We compared the CoSR with the 'Pixel+RSR' algorithm proposed in [3], the 'PCA+RSR' algorithm which performs the RSR in the PCA subspace, and other existing algorithms such as 'PCA+NN', 'PCA+NS', and 'PCA+Linear SVM' [41,42] which are also used for comparison in [3]. Moreover, we also explored the important role of the sparsity in the CoSR model by comparing it with the fusion of two non-sparse solutions with the ℓ_2-norm constraint in forward and backward directions using the same dictionaries as those in the CoSR algorithm, which are denoted by the 'PCA+f-ℓ_2' and 'PCA+b-ℓ_2', respectively. We also compared our CoSR with the 'Pixel+ℓ_2' algorithm proposed in [31], and for comparison fairness, we expanded the dictionary with the same number of noise bases as the fSR of our CoSR, viz. 100 box images (8 large ones and 92 small ones).

We evaluated our algorithm on the Extended Yale B dataset for robust face recognition with random pixel corruption, and on the AR database for disguise or block corruption. The Extended Yale B dataset totally contains 2414 face images from 38 individuals under various controlled lighting conditions [43], from which the cropped face images with the size of 192 by 168 were captured [44]. The AR face database contains over 4000 face images from 126 individuals, and for each individual, 26 face images were collected in two separate sessions with many facial variations including illumination change, expressions and facial disguise [45]. Following [3], we chose a subset of the dataset consisting of 1399 face images from 50 male individuals and 50 female ones (14 from each individual, except for a corrupted image w-027-14.bmp).

6.1. Learning bases of bSR

This subsection will show that constructing the bases of the bSR leads to an improvement of the CoSR. For both face databases, the dictionary size of the bSR is set to be 301, where 300 ones are learned bases and the rest one is the test image. The fusion weight w is set to be 0.8, and the parameter ξ in Eq. (20) is set to be 0.1, tuned by 5-fold *cross-validation* on the training set. The sparse representations are all solved with the FSS and the DALM toolboxes [46] implemented by the MATLAB 7.1.0.183 (R14) Service Pack 3 running on a computer with 16 Quad-Core AMD Opteron Processors 8382, 64 GB memory, and the OS of Debian GNU/Linux 5.0. The sparse tradeoff parameter λ is set to be 0.05. The PCA projects face images into a 500-dimension subspace.

6.1.1. Dictionary learning for bSR

We evaluate the effectiveness of the learned bases for the bSR by performing a comparative experiment between random bases and learned bases with the simple Lagrange dual algorithm on the Extended YaleB dataset. For each one of 38 individuals, half of the images, viz. about 32 images per subject, are randomly selected for training, and the remaining for test. As a result, the bSR with random dictionary achieves a recognition rate of 0.193, while the bSR with learned dictionary by our method achieves 0.938. This result implies that learning bases greatly improves the bSR.

6.1.2. Noisy bases for robust face recognition

As we discussed before, for robust face recognition, adding noise to the learned bases may avoid the failure of the bSR due to corruption in the test image. To confirm this claim, we compare the bSR recognition rates between the learned dictionaries without and with noise on the Extended Yale B face database. Following the literature [3], we choose Subset 1 and Subset 2 (717 images, normal-to-moderate lighting conditions) for training and Subset 3 (453, more extreme lighting conditions) for test. Then we replace the values of a percentage (0–90%) of pixels randomly selected from each test image with uniform random values. The values and the locations of the selected pixels are unknown to the algorithms. The dictionary is learned for five iterations by the alternative optimization method described in Section 4.1. Then the noise $\hat{n}$, which is added to the learned dictionary as in Eq. (20), is approximately computed by Eq. (19). Finally, the comparative results are shown in Table 1, from which we can see that the recognition rates of the bSR with the clean learned dictionary vary between a very large range: from 0.883 to 0.046 with the change of the corruption percentage, while the recognition rates of the bSR with the noisy learned dictionary vary between a much smaller range. The bSR with the noisy learned dictionary is much more robust to the corruption, and performs better than that with the clean learned dictionary when the corruption percentage becomes large. Also we can find that the recognition rates of the bSR with the noisy learned dictionary are lower than those with the clean learned dictionary when the corruption percentage is relatively small, but the powerful recognition capability of the fSR can compensate it. We will confirm this in the following experiments.

6.1.3. Experimental results of our CoSR on two face datasets

We evaluate our CoSR algorithm, which adopts the bSR with noisy learned dictionary on two face datasets: Extended Yale B and AR face.

On the Extended Yale B dataset, we evaluate the performance of our CoSR algorithm for the corrupted face images with random noise of a uniform distribution. First, we learn a dictionary of the bSR using the method in Section 4.1; then, we compute the additional bases of the bSR as in Eq. (20); finally, the fSR and bSR are performed respectively and combined together to produce the final classification decision. Table 2 shows the experimental results of our algorithm, comparing with other algorithms.

On the AR face dataset, we evaluate the performance of our CoSR algorithm for the corrupted face images with unknown noise. For AR face database, following [3], we use 799 images (about 8 from each) without disguise for training, and the remaining 600 ones with disguise (wearing sunglasses or scarves) for test. First, we learn the dictionary of the bSR using the method in Section 4.1; then, we compute the additional bases of the bSR as in Eq. (20), in which the noise is approximately computed by Eq. (19); finally, the fSR and bSR are performed respectively and combined together to produce the final classification decision. Table 4 shows the recognition rates of our algorithm, comparing with other algorithms.

From the above results, we have the following observations.

(1) Our CoSR algorithm obtains relatively good recognition performance. The fSR and the bSR are complementary to each other, and for most cases, the fusion of them can improve the recognition rates. Especially for the case of large corruption

Table 2
Recognition rates of PCA+CoSR (bSR with noisy learned dictionary) and 'pixel+RSR' algorithms on Extended Yale B face database.

Corruption percentage	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%
Pixel+RSR (FSS)	–	–	–	–	–	–	–	–	–	–
Pixel+RSR (DALM)	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**0.996**	**0.892**	0.391	0.071
Pixel+ℓ_2 [31]	**1.000**	**1.000**	**1.000**	**0.998**	**0.996**	0.923	0.753	0.483	0.219	0.091
PCA+RSR (FSS)	**1.000**	0.998	0.998	0.998	0.982	0.899	0.613	0.218	0.086	0.038
PCA+RSR (DALM)	**1.000**	0.998	0.998	0.998	0.989	0.894	0.609	0.215	0.083	0.041
PCA+f-ℓ_2	*1.000*	*1.000*	*1.000*	*0.993*	*0.980*	*0.907*	*0.739*	*0.501*	*0.232*	*0.095*
PCA+b-ℓ_2	*0.949*	*0.947*	*0.940*	*0.921*	*0.938*	*0.927*	*0.896*	*0.832*	*0.691*	*0.316*
Fusion	0.989	0.982	0.978	0.971	0.958	0.929	0.899	0.766	0.589	0.190
PCA+fSR (FSS)	*1.000*	*1.000*	*1.000*	*0.993*	*0.980*	*0.901*	*0.724*	*0.457*	*0.210*	*0.080*
PCA+bSR (FSS)	*0.536*	*0.534*	*0.541*	*0.534*	*0.528*	*0.536*	*0.545*	*0.556*	*0.680*	*0.461*
Fusion (PCA+CoSR)	**1.000**	**1.000**	**1.000**	**1.000**	0.991	0.960	0.903	0.753	**0.687**	**0.453**
PCA+fSR (DALM)	*1.000*	*1.000*	*1.000*	*0.993*	*0.980*	*0.901*	*0.724*	*0.457*	*0.212*	*0.080*
PCA+bSR (DALM)	*0.543*	*0.534*	*0.545*	*0.554*	*0.559*	*0.552*	*0.578*	*0.614*	*0.713*	*0.536*
Fusion (PCA+CoSR)	**1.000**	**1.000**	**1.000**	**1.000**	**0.992**	0.960	0.910	0.768	**0.704**	**0.528**

Table 3
Time consumptions (s) of our CoSR model (BSR with learned dictionary + noise) and the 'pixel+RSR' algorithm for recognizing each test face.

Algorithms	Extended Yale B	AR Database
Pixel+RSR (FSS)	> 2000	1988.3
Pixel+RSR (DALM)	1108.8	582.2
Pixel+ℓ_2 [31]	**1.8**	**0.3**
PCA+RSR (FSS)	26.8	97.5
PCA+RSR (DALM)	3.1	4.5
PCA+CoSR (FSS)	**1.5**	**3.8**
PCA+CoSR (DALM)	243.6	550.6

Table 4
The recognition rates of our CoSR model (bSR with noisy learned dictionary) and other algorithms on the AR face dataset.

Algorithms	Sunglasses	Scarves	Average
Pixel+RSR (FSS)	0.827	0.563	0.695
Pixel+RSR (DALM)	0.827	0.560	0.694
Pixel+ℓ_2 [31]	0.677	0.810	0.744
PCA+RSR (FSS)	0.653	0.320	0.487
PCA+RSR (DALM)	0.647	0.320	0.483
PCA+f-ℓ_2	*0.543*	*0.773*	*0.658*
PCA+b-ℓ_2	*0.157*	*0.200*	*0.179*
Fusion	0.533	0.773	0.653
PCA+fSR (FSS)	*0.620*	*0.820*	*0.720*
PCA+bSR (FSS)	*0.533*	*0.400*	*0.467*
Fusion (PCA+CoSR, FSS)	**0.743**	**0.840**	**0.792**
PCA+fSR (DALM)	*0.620*	*0.820*	*0.720*
PCA+bSR (DALM)	*0.483*	*0.420*	*0.452*
Fusion (PCA+CoSR, DALM)	**0.740**	**0.833**	**0.787**

percentage (the Yale B dataset with corruption percentage larger than 80% and the AR dataset), our CoSR model obtains the best results.

(2) From Table 2, we can see that the performance of the bSR is relatively stable with the varying percentage of noise, and it performs better than the fSR when the corruption percentage is relatively large. This indicates that the bSR contributes the most to the robustness of our CoSR model when the corruption percentage is large.

(3) For the Extended Yale B face dataset, the single 'PCA+f-l2' obtains better results than the single 'PCA+fSR', and the single 'PCA+b-l2' obtains better results than the single 'PCA+bSR'. And for the AR face dataset, the situation is reversed. But for both databases, the fusion of non-sparse solutions in the forward and backward directions, viz. 'PCA+f-l2' and 'PCA+b-l2', cannot improve the recognition rates, and underperforms the fusion of 'PCA+fSR' and 'PCA+bSR', which indicates that the sparsity plays an important role in our CoSR algorithm. Though the 'PCA+b-ℓ_2' shows more robustness than the 'PCA+bSR' when the corruption percentage is relatively small, its complementarity to the 'PCA+f-ℓ_2' is weaker than that of the 'PCA+bSR' to the 'PCA+fSR', according to the comparison of the fusion performance. For the case of large corruption percentage (the Yale B dataset with corruption percentage larger than 80% and the AR dataset), the 'PCA+b-ℓ_2' shows less robustness than the 'PCA+bSR', which results in lower fusion performance. Moreover, our CoSR algorithm outperforms the 'Pixel+ℓ_2' algorithm in [31] on recognition rates.

6.1.4. Time complexity

In order to validate the efficiency of our CoSR algorithm, we also compared our CoSR algorithm on time consumption with the classical 'Pixel+RSR' algorithm, the 'PCA+RSR' algorithm, and the 'Pixel+ℓ_2' algorithm, as shown in Table 3. From the table we can see that our CoSR algorithm consumes much less recognition time than the classical 'Pixel+RSR' algorithm. The reason lies in that

our CoSR algorithm is performed in a subspace with a much lower dimensionality, and our model adopts a dictionary with much fewer bases. The time consumption comparison between our CoSR model and the 'PCA+RSR' algorithm is also given, which shows that our CoSR algorithm consumes less time than the 'PCA+RSR' algorithm when the FSS solution is adopted, but in reverse when the DALM solution is adopted. The 'Pixel+ℓ_2' algorithm takes the least time on the AR face dataset. Overall, however, our CoSR algorithm with the FSS solution consumes relatively little recognition time (several seconds per test face), and is therefore suitable for real face recognition systems.

For the FSS algorithm, the time complexity to solve one single sparse optimization problem is $O(n(d^2+d))$. And since the bSR is solved by performing sparse optimization problem n times with n input vectors (training images), respectively, the time complexity of our CoSR algorithm is $O(n^2(d^2+d))$. Even so, since our CoSR algorithm, which solves the robustness problem by the bSR, is not necessary to be performed in the original pixel space, the vector dimensionality d and the dictionary size n are both largely reduced. Thereby, the time consumption of our algorithm can be much less than that of the traditional 'Pixel+RSR' algorithm, especially for high resolution images. The above analysis is also applicable to other sparse solution algorithms such as the DALM, GP, and homotopy.

6.2. Unknown faces help the bSR

This subsection shows the experimental evaluation of our CoSR model adopting unknown faces as the additional bases of the bSR. In these experiments, the PCA projects face images into a 300-dimension

Table 5
The recognition rates of our CoSR model (unknown faces help) on the YaleB face database with random pixel corruption.

Num. of classes unknown/known		3/35	6/32	9/29	12/26
10%	fSR	*0.999 ± 0.001*	*1.000 ± 0*	*1.000 ± 0*	*0.999 ± 0.001*
	bSR	*0.883 ± 0.006*	*0.965 ± 0.009*	*0.971 ± 0.028*	*0.988 ± 0.007*
	Fusion	0.995 ± 0.003	0.998 ± 0.004	0.997 ± 0.004	**0.999** ± 0.001
30%	fSR	*0.995 ± 0.001*	*0.990 ± 0.001*	*0.991 ± 0.006*	*0.992 ± 0.003*
	bSR	*0.876 ± 0.012*	*0.968 ± 0.007*	*0.969 ± 0.030*	*0.986 ± 0.003*
	Fusion	0.975 ± 0.005	**0.992** ± 0.003	**0.993** ± 0.010	**0.999** ± 0.001
50%	fSR	*0.887 ± 0.017*	*0.855 ± 0.008*	*0.866 ± 0.024*	*0.854 ± 0.029*
	bSR	*0.870 ± 0.004*	*0.947 ± 0.014*	*0.957 ± 0.023*	*0.983 ± 0.005*
	Fusion	**0.939** ± 0.015	**0.976** ± 0.006	**0.977** ± 0.017	**0.996** ± 0.001
70%	fSR	*0.411 ± 0.032*	*0.350 ± 0.026*	*0.385 ± 0.098*	*0.323 ± 0.120*
	bSR	*0.833 ± 0.022*	*0.910 ± 0.007*	*0.935 ± 0.026*	*0.946 ± 0.012*
	Fusion	**0.851** ± 0.016	**0.905** ± 0.020	**0.933** ± 0.030	**0.932** ± 0.005
90%	fSR	*0.085 ± 0.029*	*0.096 ± 0.022*	*0.076 ± 0.029*	*0.073 ± 0.049*
	bSR	*0.563 ± 0.022*	*0.549 ± 0.018*	*0.550 ± 0.010*	*0.550 ± 0.022*
	Fusion	**0.509** ± 0.023	**0.495** ± 0.035	**0.424** ± 0.030	**0.329** ± 0.055

Table 6
The recognition rates of our CoSR model (unknown faces help) on the AR face dataset.

Number of classes unknown/known		5/95	10/90	15/85	20/80	25/75
Sunglasses	fSR	*0.606 ± 0.017*	*0.602 ± 0.028*	*0.606 ± 0.025*	*0.618 ± 0.027*	*0.634 ± 0.031*
	bSR	*0.813 ± 0.058*	*0.867 ± 0.042*	*0.901 ± 0.031*	*0.917 ± 0.022*	*0.930 ± 0.021*
	Fusion	**0.895** ± 0.032	**0.914** ± 0.026	**0.927** ± 0.013	**0.938** ± 0.016	**0.942** ± 0.019
Scarves	fSR	*0.762 ± 0.022*	*0.767 ± 0.022*	*0.771 ± 0.017*	*0.771 ± 0.029*	*0.773 ± 0.026*
	bSR	*0.708 ± 0.047*	*0.803 ± 0.042*	*0.845 ± 0.032*	*0.875 ± 0.023*	*0.893 ± 0.031*
	Fusion	**0.836** ± 0.024	**0.879** ± 0.023	**0.900** ± 0.019	**0.915** ± 0.023	**0.928** ± 0.021

subspace, which is smaller than that in the previous experiments since the number of training images is reduced. The sparse representations are solved with the FSS algorithm which has been proved faster in the previous subsection. The fusion weight w is set to be 0.2, tuned by 5-fold *cross-validation* on the training set.

For Extended Yale B database, we also choose Subsets 1, 2 and 3 for evaluation. Firstly, we randomly select a small number of individuals (3, 6, 9, 12) from 38, and use the face images from these individuals in three subsets as 'unknown faces'. Second, we choose the remaining of Subsets 1 and 2 for training and the remaining of Subset 3 for test. Finally, we replace the values of a percentage (10%, 30%, 50%, 70%, 90%) of pixels randomly selected from each test image with uniform random values, and to obtain the unknown faces with the collection surroundings and devices similar to the test faces, we also randomly corrupt a random percentage of pixels in each unknown face with uniform random values. Table 5 shows the recognition rates of our CoSR algorithm, which are the average of ten runs on different random dataset splits for unknown faces.

For AR face database, we choose a subset consisting of 1399 face images from 50 male individuals and 50 female ones for evaluation. Firstly, we randomly select a small number of individuals (5, 10, 15, 20, 25) from 100, and use the face images (with and without disguise) from these individuals in the subset as the *unknown faces*. Secondly, we choose the face images without disguise from the remaining for training, and those with disguise for test. Finally, we obtain the recognition rates of our CoSR algorithm shown in Table 6, which are the average of ten runs on different random dataset splits described as above.

From the above results, we have the following observations.

(1) The fSR and the bSR are complementary to each other. The performance of the bSR is relatively stable with the varying

noise percentage, which indicates that it contributes the most to the robustness of our CoSR model. The results also indicate that our CoSR model works well and greatly outperforms over the general sparse representation which we call the 'fSR' when the percentage of corruption is large.

(2) By a rough[2] comparison between the CoSR adopting the learned bases and that adopting the least unknown faces, we can see that, for the cases of intense noise or large block disguise, the latter performs better than the former. This observation confirms our speculation in Subsection 4.2 that *unknown faces* are more similar to input faces than *learned bases* and naturally contain similar noise to that in test images. Another factor is that additional faces provide more complementary information for the fusion.
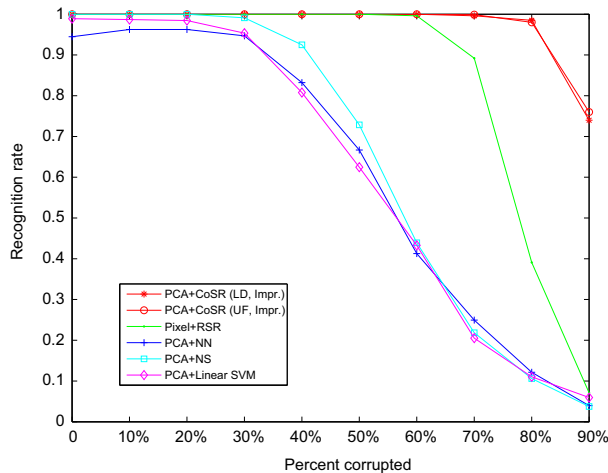
6.3. Improvements of our CoSR algorithm

On the YaleB face dataset, we evaluate our CoSR algorithm improved by expanding the dictionary of the fSR with random vectors as described in Section 5.1, for both strategies of expanding the dictionary of the bSR with *learned bases* and *unknown faces*. For the *unknown faces* strategy, to compare with other state-of-the-art algorithms as most fair as possible (see footnote 1), we adopt a very small subset, viz. 3 of 38 individuals to expand the bases set of the bSR. We use a dictionary consisting of all training images X together with an expanded noise bases set R containing 100 vectors with normally distributed random numbers for the fSR. Table 7 shows the recognition rates of our improved CoSR

---

[2] There can be no strictly fair comparison since additional unknown faces are used in this model.

Table 7
Recognition rates of our improved CoSR algorithm on Extended Yale B face database.

Corruption percentage	0–60%	70%	80%	90%
PCA+CoSR (LD, Impr.)	1.000	0.996	0.985	0.740
PCA+CoSR (UF, Impr.)	1.000 ± 0.000	0.999 ± 0.001	0.980 ± 0.008	0.760 ± 0.018

Fig. 5. The recognition rates of various algorithms on the YaleB face dataset.

Table 8
The recognition rates of various algorithms with image block partition on the AR face dataset.

Algorithms	Sunglasses	Scarves	Average
PCA+NN	0.693	0.123	0.408
PCA+NS	0.697	0.147	0.422
PCA+SVM	0.523	0.277	0.400
Pixel+SSR [47]	0.930	0.927	0.929
Pixel+RSR (Impr.) [3]	0.975	0.935	0.955
PCA+CoSR (LD, Impr.)	**0.977**	**0.947**	**0.962**
PCA+CoSR (UF, Impr.)	**0.983** ± 0.004	**0.968** ± 0.007	**0.976** ± 0.005

algorithm, and Fig. 5 plots the comparison between our algorithm and other state-of-the-art algorithms.

On the AR face dataset, we evaluate our CoSR algorithm improved by partitioning face images into eight (4×2, namely $K=8$) blocks of 41×60 pixels as described in Section 5.2, for both strategies of expanding the dictionary of the bSR with *learned bases* and *unknown faces*. For the CoSR with *unknown faces* expanding the bases set of the bSR, to compare with other state-of-the-art algorithms as most fair as possible (see footnote 2), we adopt a very small subset, viz. 5 of 100 individuals to expand the bases set of the bSR. Table 8 shows the recognition rates of our algorithm, comparing with other state-of-the-art algorithms.

From the experimental results, we can see that our CoSR algorithm outperforms the other state-of-the-art algorithms, and obtains the highest recognition rates on these two face datasets in all cases we have tried so far.

In addition, the time consumption of our CoSR algorithm to recognize a single test face is shown in Table 9, which indicates that our algorithm is much more efficient and can be more easily taken by real face recognition systems, comparing with the classical 'Pixel+RSR' algorithm.

Table 9
Time consumptions (seconds) of our improved CoSR algorithm to recognize a single test face.

Algorithms	Extended Yale B	AR database
PCA+CoSR (UF, Impr.)	2.6	5.8
PCA+CoSR (LD, Impr.)	1.7	5.2

7. Conclusion

This work has extended the CoSR algorithm to solve the robust face recognition problem by expanding the bases set of the bSR with *learned bases* or *unknown faces*. Compared with the other state-of-the-art algorithms, our CoSR algorithm obtains inspiring and competitive recognition rates, especially when the noise percentage in test faces is large. Moreover, since our CoSR algorithm is performed in a subspace with a small dimensionality, it has an overwhelming advantage on time consumption over the traditional 'Pixel+RSR' algorithm which works in image pixel space. Further, our study has also shown that the bSR contributes the most to the robustness of our CoSR model for robust face recognition, and *unknown faces* do better than *learned bases* for the expanded bSR dictionary. When additional faces are available in real applications, preference will be given to the former. But when it is uneasy to obtain additional faces, the latter can be a proper alternative method.

In addition, this paper also reveals that the sparsity plays a very important role in our CoSR algorithm for face recognition.

Conflict of interest

None declared.

Acknowledgments

References

[1] M. Yang, L. Zhang, Gabor feature based sparse representation for face recognition with gabor occlusion dictionary, In: Proceedings of the 11th European Conference on Computer Vision: Part VI, 2010, pp. 448–461.

[2] W. Deng, J. Hu, J. Guo, Extended SRC:: undersampled face recognition via intraclass variant dictionary, IEEE Trans. Pattern Anal. Mach. Intell. 34 (9) (2012) 1864–1870.

[3] J. Wright, A. 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.

[4] Z.-Q. Zhao, H. Glotin, Z. Xie, J. Gao, X. Wu, Cooperative sparse representation in two opposite directions for semi-supervised image annotation, IEEE Trans. Image Process. 21 (9) (2012) 4218–4231.

[5] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, Y. Ma, Towards a practical face recognition system: robust alignment and illumination via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 34 (2) (2012) 372–386.

[6] K. Blekas, A. Likas, Sparse regression mixture modeling with the multi-kernel relevance vector machine, Knowl. Inf. Syst. 39 (2) (2014) 241–264.

[7] Q. Gao, Y. Huang, H. Zhang, X. Hong, K. Li, Y. Wang, Discriminative sparsity preserving projections for image recognition, Pattern Recognit. 48 (8) (2015) 2543–2553.

[8] R. Duda, P. Hart, D. Stork, Pattern Classification, second ed. John Wiley & Sons, New York, 2001.

[9] N. Tomasev, D. Mladenic, Hubness-aware shared neighbor distances for high-dimensional k-nearest neighbor classification, Knowl. Inf. Syst. 39 (1) (2014) 89–122.

[10] J. Ho, M. Yang, J. Lim, K. Lee, D. Kriegman, Clustering appearances of objects under varying illumination conditions, Comput. Vis. Pattern Recognit. (2003) 11–18.

[11] J. Yang, D. Chu, L. Zhang, Y. Xu, J. Yang, Sparse representation classifier steered discriminative projection with applications to face recognition, IEEE Trans. Neural Netw. Learn. Syst. 24 (7) (2013) 1023–1035.

[12] D. Wang, S. Hoi, Y. He, J. Zhu, Mining weakly labeled web facial images for search-based face annotation, IEEE Trans. Knowl. Data Eng. 26 (1) (2014) 166–179.

[13] G. Hua, M.-H. Yang, E. Learned-Miller, Y. Ma, M. Turk, D.J. Kriegman, T.S. Huang, Introduction to the special section on real-world face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 33 (10) (2011) 1921–1924.

[14] J. Wang, Y. Xu, J. You, Sparse residue for occluded face image reconstruction and classification, In: 2012 21st International Conference on Pattern Recognition (ICPR), 2012, pp. 1707–1710.

[15] M. Turk, A. Pentland, Face recognition using eigenfaces, In: CVPR, 1991, pp. 586–591.

[16] Z.Q. Zhao, D.S. Huang, B.Y. Sun, Human face recognition based on multi-features using neural networks committee, Pattern Recognit. Lett. 25 (12) (2004) 1351–1358.

[17] P.N. Belhumeur, J.a.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.

[18] X. He, S. Yan, S. Hu, P. Niyogi, H. Jiang Zhang, Face recognition using Laplacian faces, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2005) 328–340.

[19] Y. Xu, X. Li, J. Yang, Z. Lai, D. Zhang, Integrating conventional and inverse representation for face recognition, IEEE Trans. Cybern. 44 (10) (2014) 1738–1746.

[20] M. Kojima, N. Megiddo, S. Mizuno, Theoretical convergence of large-step primal–dual interior point algorithms for linear programming, Math. Program. 59 (1993) 1–21.

[21] G. Davis, S. Mallat, M. Avellaneda, Adaptive greedy approximations, Constr. Approx. 13 (1) (2004) 57–98.

[22] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, Ann. Stat. 32 (2004) 407–499.

[23] M.D. Plumbley, Recovery of sparse representations by polytope faces pursuit, In: Proceedings of the 6th International Conference on Independent Component Analysis and Blind Source Separation (ICA 2006), 2006, pp. 206–213.

[24] M.A.T. Figueiredo, R. Nowak, S. Wright, Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems, IEEE J. Select. Top. Signal Process. 1 (4) (2007) 586–597.

[25] J. Yang, Y. Zhang, Alternating direction algorithms for ℓ_1−problems in compressive sensing, SIAM J. Sci. Comput. 33 (1) (2011) 250–278.

[26] D.L. Donoho, Y. Tsaig, Fast solution of ℓ_1−norm minimization problems when the solution may be sparse, IEEE Trans. Inf. Theory 54 (11) (2008) 4789–4812.

[27] H. Lee, A. Battle, R. Raina, A.Y. Ng, Efficient sparse coding algorithms, In: Advances in Neural Information Processing Systems, 2007, pp. 801–808.

[28] A.Y. Yang, S.S. Sastry, A. Ganesh, Y. Ma, Fast ℓ_1−minimization algorithms and an application in robust face recognition: a review, In: ICIP, IEEE, Hong Kong, 2010, pp. 1849–1852.

[29] R. Rigamonti, M.A. Brown, V. Lepetit, Are sparse representations really relevant for image classification? In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, 2011.

[30] C.-Y. Lu, H. Min, Z.-Q. Zhao, L. Zhu, D.-S. Huang, S. Yan, Robust and efficient subspace segmentation via least squares regression, In: Proceedings of the 12th European Conference on Computer Vision – Volume Part VII, 2012, pp. 347–360.

[31] Q. Shi, A. Eriksson, A. van den Hengel, C. Shen, Is face recognition really a compressive sensing problem? In: CVPR, 2011, pp. 553–560.

[32] L. Zhang, M. Yang, X. Feng, Sparse representation or collaborative representation: which helps face recognition? In: Proceedings of the 2011 International Conference on Computer Vision, 2011, pp. 471–478.

[33] W. Yang, Z. Wang, C. Sun, A collaborative representation based projections method for feature extraction, Pattern Recognit. 48 (1) (2015) 20–27.

[34] P. Zhao, B. Yu, D. Madigan, On model selection consistency of lasso, J. Mach. Learn. Res. 7 (2006) (2006).

[35] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B 58 (1994) 267–288.

[36] S.S. Chen, D.L. Donoho, Michael A. Saunders, Atomic decomposition by basis pursuit, SIAM J. Sci. Comput. 20 (1998) 33–61.

[37] S. Gao, I.W.-H. Tsang, L.-T. Chia, P. Zhao, Local features are not lonely - laplacian sparse coding for image classification, In: The 23rd IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 3555–3561.

[38] Y.A. Censor, S.A. Zenios, Parallel Optimization: Theory, Algorithms and Applications, Oxford University Press, New York, 1997.

[39] S. Fidler, D. Skocaj, A. Leonardis, Combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling, IEEE Trans. Pattern Anal. Mach. Intell. 28 (3) (2006) 337–350.

[40] L. Kuncheva, J. Rodríguez, A weighted voting framework for classifiers ensembles, Knowl. Inf. Syst. 38 (2) (2014) 259–275.

[41] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (3) (1995) 273–297.

[42] Y.-C. Wu, A sparse l_2-regularized support vector machines for efficient natural language learning, Knowl. Inf. Syst. 39 (2) (2014) 305–328.

[43] A.S. Georghiades, P.N. Belhumeur, D.J. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6) (2001) 643–660.

[44] K.-C. Lee, J. Ho, D.J. Kriegman, Acquiring linear subspaces for face recognition under variable lighting, IEEE Trans. Pattern Anal. Mach. Intell. 27 (5) (2005) 684–698.

[45] A. Martínez, R. Benavente, The AR face database, June 1998. ⟨http://www.cat.uab.cat/Public/Publications/1998/MaB1998⟩.

[46] ⟨http://www.eecs.berkeley.edu/~yang/software/l1benchmark⟩, accessed: 2013-10-09.

[47] W. Ou, X. You, D. Tao, P. Zhang, Y. Tang, Z. Zhu, Robust face recognition via occlusion dictionary learning, Pattern Recognit. 47 (4) (2014) 1559–1572.

Zhong-Qiu Zhao is a Professor at Hefei University of Technology, China. He obtained the Master's degree in Pattern Recognition & Intelligent System at Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei, China, in 2004, and the Ph.D. degree in Pattern Recognition & Intelligent System at University of Science and Technology, China, in 2007. From April 2008 to November 2009, he held a postdoctoral position in image processing in CNRS UMR6168 Lab Sciences de l'Information et des Systmes, France. From January 2013 to December 2014, he held a research fellow position in image processing at the Department of Computer Science of Hongkong Baptist University, Hongkong, China. Now he works in Laboratory of Data Mining and Intelligent Computing, Hefei University of Technology, China. His research is about pattern recognition, image processing, and computer vision.

Yiu-ming Cheung is a Professor at Department of Computer Science in Hong Kong Baptist University. He received the Ph.D. degree from the Department of Computer Science and Engineering, The Chinese University of Hong Kong in 2000, and then joined the Department of Computer Science at Hong Kong Baptist University in 2001. His current research interests include machine learning, pattern recognition, watermarking, visual computing, and optimization. He is the founding Chairman of IEEE (Hong Kong) Computational Intelligence Chapter. Currently, he is also the Associate Editor of IEEE Transactions on Neural Networks and Learning Systems, Knowledge and Information Systems, as well as the guest co-editor and editorial board member of several international journals.

Haibo Hu received the Ph.D. degree in computer science from the Hong Kong University of Science and Technology in 2005. He is an assistant professor in the Department of Electronic and Information Engineering at Hong Kong Polytechnic University. Prior to this, he held several research and teaching posts at HKUST and HKBU. His research interests include mobile and wireless data management, location-based services, and privacy-aware computing. He has published more than 50 research papers in international conferences, journals, and book chapters. He is also the recipient of many awards, including ACM-HK Best PhD Paper Award and Microsoft Imagine Cup. He is a member of the IEEE.

Xindong Wu is a Professor of Computer Science at the University of Vermont (USA), a Yangtze River Scholar in the School of Computer Science and Information Engineering at the Hefei University of Technology (China), and a Fellow of the IEEE and the AAAS.
Dr. Wu received his Bachelor's and Master's degrees in Computer Science from the Hefei University of Technology, China, and his Ph.D. degree in Artificial Intelligence from the University of Edinburgh, Britain. His research interests include data mining, knowledge-based systems, and Web information exploration. He is the Steering Committee Chair of the IEEE International Conference on Data Mining (ICDM), the Editor-in-Chief of Knowledge and Information Systems (KAIS, by Springer), and a Series Editor of the Springer Book Series on Advanced Information and Knowledge Processing (AI&KP). He was the Editor-in-Chief of the IEEE Transactions on Knowledge and Data Engineering (TKDE, by the IEEE Computer Society) between 2005 and 2008. He served as Program Committee Chair/Co-Chair for ICDM '03 (the 2003 IEEE International Conference on Data Mining), KDD-07 (the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining), and CIKM 2010 (the 19th ACM Conference on Information and Knowledge Management).