PLOS ONE

Low-Rank and Eigenface Based Sparse Representation for Face Recognition

Yi-Fu Hou[1], Zhan-Li Sun[1], Yan-Wen Chong[2], Chun-Hou Zheng[1]*

1 College of Electrical Engineering and Automation, Anhui University, Hefei, China, 2 State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China

Abstract

In this paper, based on low-rank representation and eigenface extraction, we present an improvement to the well known Sparse Representation based Classification (SRC). Firstly, the low-rank images of the face images of each individual in training subset are extracted by the Robust Principal Component Analysis (Robust PCA) to alleviate the influence of noises (e.g., illumination difference and occlusions). Secondly, Singular Value Decomposition (SVD) is applied to extract the eigenfaces from these low-rank and approximate images. Finally, we utilize these eigenfaces to construct a compact and discriminative dictionary for sparse representation. We evaluate our method on five popular databases. Experimental results demonstrate the effectiveness and robustness of our method.

Introduction

Sparse representation algorithm has been successfully applied in image restoration [1] and compressed sensing [2] in the past several years. Recently, it has also led to promising results in image classification such as face recognition [3–5] and texture recognition [6]. For face recognition, given an over-complete dictionary, a testing face image can be linearly represented as a sparse coefficient vector. With the recent progress of l_0-norm and l_1-norm minimization techniques, classification based on sparse representation has became a hot topic owing to the fact that a high-dimensional image vector can be well mapped to a low-dimension manifold.

The sparse representation model can be succinctly denoted as the following optimization equation:

$$\min_{x} \ \|x\|_1 \quad s.t. \quad y = Dx \tag{1}$$

Where D is an over-complete dictionary, y is a testing sample, e.g., a face image need to be identified, x is the sparse coefficient representation vector over dictionary D.

In the early stage, some literatures [7–9] directly grouped the off-the-shelf bases together as the dictionary atoms. However, with the further research in sparse representation, learning an excellent dictionary has been proved to be an effective way to improve the capacity of signal reconstruction noticeably. To improve the performance of sparse representation, many algorithms have been proposed to optimize dictionary and acquired corking performance. E.g., Aharon et al. [10] generalized the k-means clustering process and proposed K-SVD algorithm, which iteratively updated the sparse representation coefficients based on the current

dictionary and then optimized the dictionary atoms to better fit the data. Mairal et al. [11] presented an energy formulation with both sparse re-construction and class discriminative components. An on-line dictionary learning algorithm based on stochastic approximation was also developed to handle large datasets with millions of training samples [12]. In addition, Yang et al. [13] applied the fisher discrimination dictionary learning (FDDL) scheme to learn a discriminative dictionary for sparse representation.

These algorithms work well for the clean images or those images corrupted by slight noises. However, if both training images and testing images are corrupted by heavy noises, we should optimize the dictionary to better reconstruct the images including the noises. Although an identity matrix [4] can be introduced into dictionary to encode the corrupted pixels, it affects the sparsity of representation coefficients. On the other hand, supposing that the corrupted pixels are separated or suppressed so that images become clean as much as possible, and then use the clean data to learn a dictionary, the results of classification or reconstruction should be improved.

Recently, some algorithms have been proposed to separate the low rank and approximate data from original corrupted data. E.g., Candes et al. [14] alleviated the influence of noises and occlusions by Robust Principal Component Analysis (Robust PCA) algorithm. Liu et al. [15] generalized the Robust PCA and proposed Low-Rank Representation(LRR)to deal with the original data and utilized the low-rank structures to segment subspaces or classify face images. Long et al. [16] improved sparse representation for face recognition based on the discriminative low-rank dictionary learning method. Zhou et al. [17] applied Go Decomposition (GoDec) algorithm to extract the low-rank term quickly and obtained commendable performance in image and video processing.
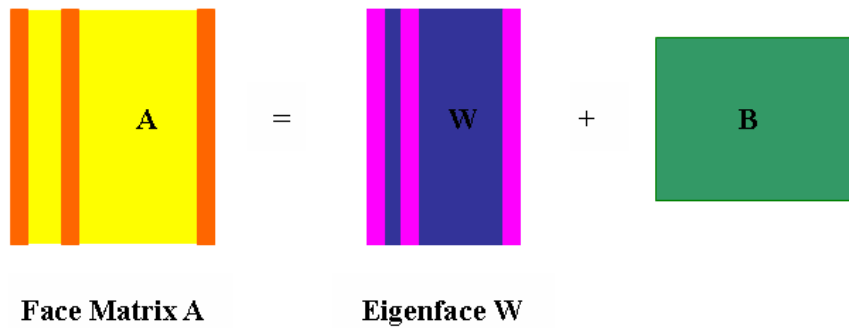
Figure 1. The relation between face matrix and eigenface matrix. Each column of matrix W represents a eigenface, and each sample (column) in face matrix A can be represented by eigenfaces in eigenface matrix W with eigenface expression pattern (column in matrix B) of the corresponding sample.

Thereby, we, in this paper, alleviate the influence of noises exiting in training samples according to the research about low rank representation.

For dictionary learning, another attractive aspect is the compactness of the constructed dictionary [18–20]. Zheng et al. [20] constructed a compact dictionary with meta-samples and regarded samples (genes expression data) as a linear mixture of independent basis snapshots (i.e., meta-samples). The results indicated that the compact dictionary achieved better recognition accuracy and decreased time consuming greatly. Actually, every face image could be represented by a large number of atoms in face space. We could extract these atoms (i.e., so-called eigenfaces) capturing the intrinsic structural information of each face space and use them to construct a compact dictionary for sparse representation.

Based on the aforementioned analysis, in this paper we incorporate low-rank and eigenfaces extraction into sparse representation for dictionary learning. Firstly, the low-rank images of training samples are extracted by Robust PCA to make the training images as clean as possible. Then, SVD is applied to extract eigenfaces from these low-rank and approximate images. Finally, these eigenfaces are combined to construct a compact dictionary for sparse representation based classification. The proposed algorithm can not only separate the influence of sparse noises but also enhance the compactness of dictionary. Compared with other dictionary learning algorithms, the merits of the proposed approach are listed as follows: First, we apply low rank transformation to compress the noises existing in training images so that the extracted features are more discriminative for classification. Second, we extract eigenfaces from the low-rank

face images by mathematical method to make the dictionary more compact. Third, comparing with the classical dictionary learning algorithms, our method is robust to some large but sparse errors such as block occlusions.

The remainder of the paper is organized as follows: Some related works about low-rank representation are introduced in Section 2. Section 3 presents our algorithm in detail. Section 4 shows the experimental results. And some theoretic analysis are also listed in this Section. Section 5 concludes the paper and outlines the future work.

Low-Rank Representation Algorithm

Robust PCA for Low-Rank Matrices Recovery

How to exploit low-rank structure from high-dimensional data is taking on increasing attention in image, audio and video processing. Meanwhile, it is also a challenging task to exactly recover the low-rank structure from the high-dimensional and corrupted data. The application of classical Principal Component Analysis (PCA) [21] suffers from a prodigious limitation owing to its brittleness with respect to serious corrupted data. Although some approaches such as multivariate trimming [22], alternating minimization [23] and random sampling techniques [24] can improve the robustness of PCA, none of them obtains a polynomial-time algorithm with strong performance guarantee. Motivated by the recent research on robust solutions of over-determined linear systems and low-rank solutions of under-determined linear systems, Candes et al. [14] proposed the Robust PCA algorithm which decomposed a corrupted matrix into a low-rank matrix and a sparse errors matrix.
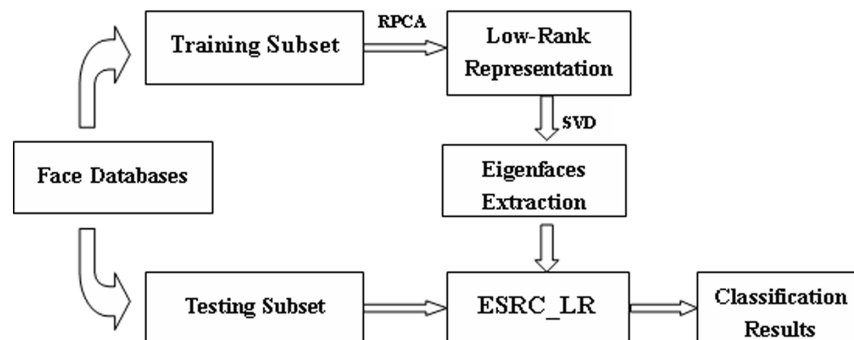
Figure 2. Overview of our method.

Table 1. The information of utilized databases.

Datasets	number of classes	samples per classes	number of features	train samples/test samples(per class)
the Extended Yale B dataset	38	64	504 (downsample ratio:1/8)	40/24
the AR dataset (part)	100	26	1600(40*40)	7/7(without occlusion, each session)
the ORL dataset	40	10	750(30*25)	8/2
the CMU-PIE dataset	68	49	4096(68*68)	25/24(almost equal)
The UMIST dataset	15	almost 28	600(30*20)	almost equal

Assuming that a cleaning data matrix L is corrupted by noises (i.e., error term S) and becomes the corrupted data matrix D, i.e.,

$$D = L + S \qquad (2)$$

where L and S are unknown but L is a low rank matrix and S is sparse. The process of recovering L from D can be reformulated as the following optimization problem:

$$\min_{L,S} \ rank(L) + \lambda\|S\|_0 \qquad s.t. \qquad D = L + S \qquad (3)$$

where λ indicates the percentage of sparse errors. Unfortunately, it is a highly non-convex optimization problem so that no efficient solution can be obtained. We relax formula (3) (i.e., replacing the l_0-norm and the rank term with the l_1-norm and the nuclear norm, respectively) and yield the following convex function:

$$\min_{L,\ S} \ \|L\|_* + \lambda\|S\|_1 \qquad s.t. \qquad D = L + S \qquad (4)$$

the minimized solutions (denoted as L^* and S^*) are named as the low-rank part and sparse errors part of the corrupted matrix D.

Augmented Lagrange Multiplier (ALM) Method for Robust PCA

We adopt in this paper the Augmented Lagrange Multiplier (ALM) [25–26] to solve problem (4). Generally, the ALM algorithm is used to solve the following optimization function:

$$\min_{X} \ G(X) \qquad s.t. \qquad F(X) = 0 \qquad (5)$$

where G is a convex function, F is a linear function. Equation (5) is solved by defining augmented Lagrange function:

$$L(X,Y,u) = G(X) + tr[Y^t F(X)] + \frac{u}{2}\|F(X)\|_F^2 \qquad (6)$$

where u is a positive scalar, Y is a Lagrange multiplier vector. Correspondingly, the augmented Lagrange function of Equation (4) is:

$$P(L,S,Y,u) =$$
$$\|L\|_* + \lambda\|S\|_1 + tr[Y^t(D-L-S)] + \frac{u}{2}\|D-L-S\|_F^2 \qquad (7)$$
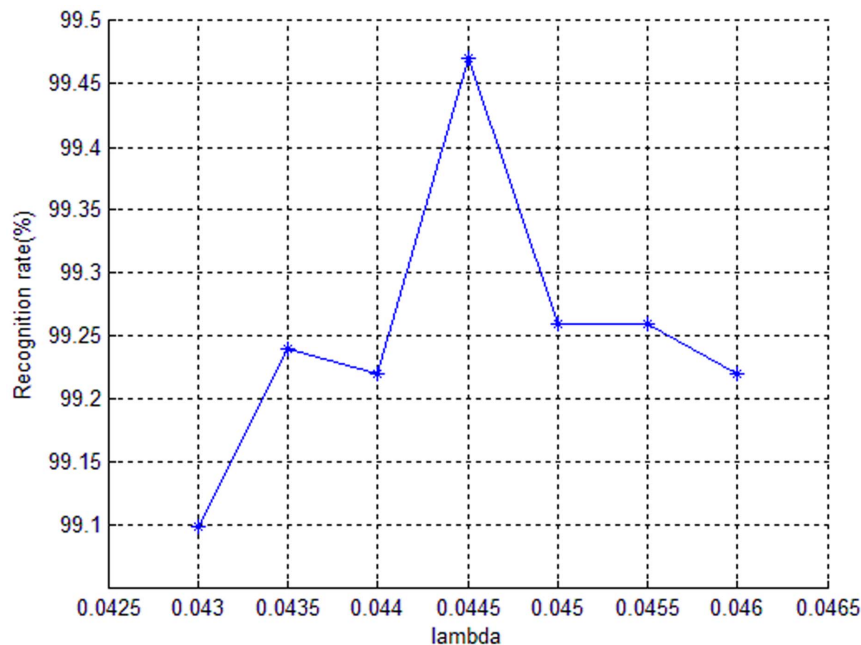
Figure 3. The relation between recognition rates and the corresponding parameter λ.

Figure 4. Accuracy rate of different number of eigenfaces. The number of eigenfaces 1, 2, 3, 4, 6, 8, 10, 12, 14 correspond to 25.95%, 74.59%, 90.30%, 96.46%, 98.31%, 98.89%, 99.47%, 99.30%, 99.47%.
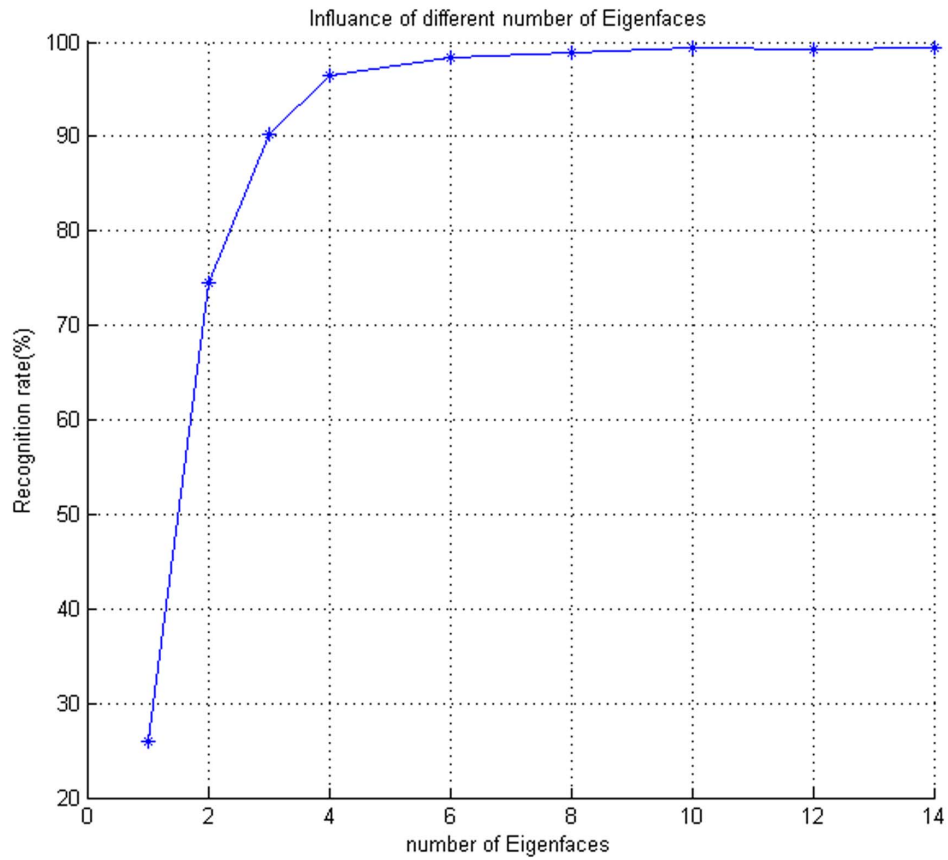doi:10.1371/journal.pone.0110318.g004

here, λ is a parameter and generally computed through the following formula:

$$\lambda = \frac{1}{\sqrt{\max(m,n)}} \qquad (8)$$

where m and n are the number of columns and rows of matrix D, respectively. In the experiment section, we will show how to select the optimal λ for face matrix D. The parameter u in Equation (7) can be adjusted by the following strategy:

$$u = \min(u * \rho, u_{\max}) \qquad (9)$$

where ρ satisfies the condition $\rho > 1$ (e.g., $\rho = 1.1$) to control the velocity of the increase of u. $u_{\max}$ is the upper bound of ρ (e.g., $u_{\max} = 1e7$).

The optimization for Equation (7) can be divided into two sub-problems. The first sub-problem is to compute L for a fixed S. The second sub-problem is to solve S for the fixed L computed

from the first sub-problem. We update L and S as follows:

$$L = \arg\min_{L} \ \|L\|_* + \ tr\left[Y^t(D-L-S)\right] + \frac{u}{2}\|D-L-S\|_F^2$$

$$= \arg\min_{L} \frac{1}{u}\|L\|_* + \frac{1}{2}\left\|L-\left(\frac{1}{u}Y+D-S\right)\right\|_F^2 \qquad (10)$$

$$S = \arg\min_{S} \ \lambda\|S\|_1 + \ tr[Y^t(D-L-S)] + \frac{u}{2}\|D-L-S\|_F^2$$

Table 2. The recognition rates on Extended Yale B Database.

Algorithms	ESRC_LR	FDDL	SRC	SVM (RBF)	SVM (linear)	NN
Recognition rate(%)	99.47	93.31	89.95	88.25	78.25	77.83

doi:10.1371/journal.pone.0110318.t002
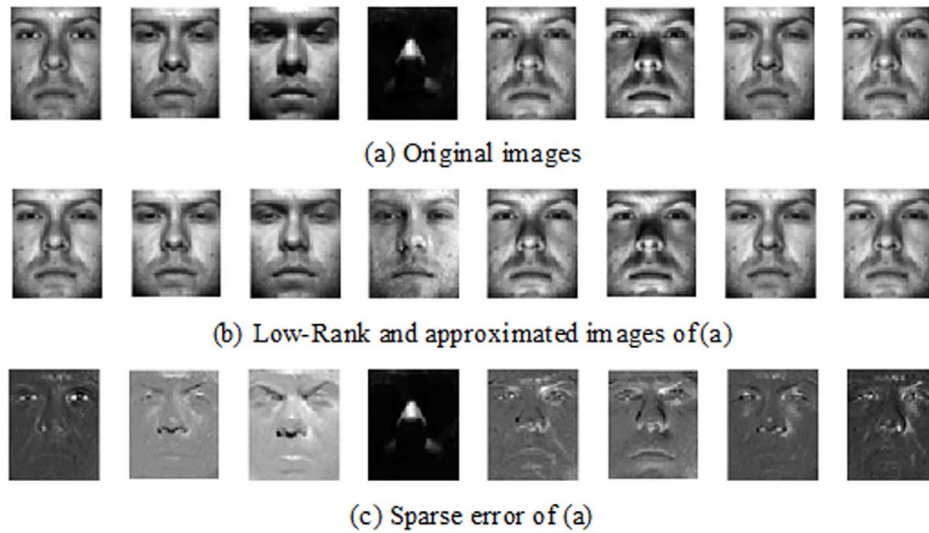
Figure 5. An example of Robust PCA algorithm on the Extended Yale B Face Database. First row is the original images with vary illumination and expression changes. Second row shows the low rank and approximate images of (a). Third row is the sparse error images of (a) which is the difference of (a) and (b).
doi:10.1371/journal.pone.0110318.g005
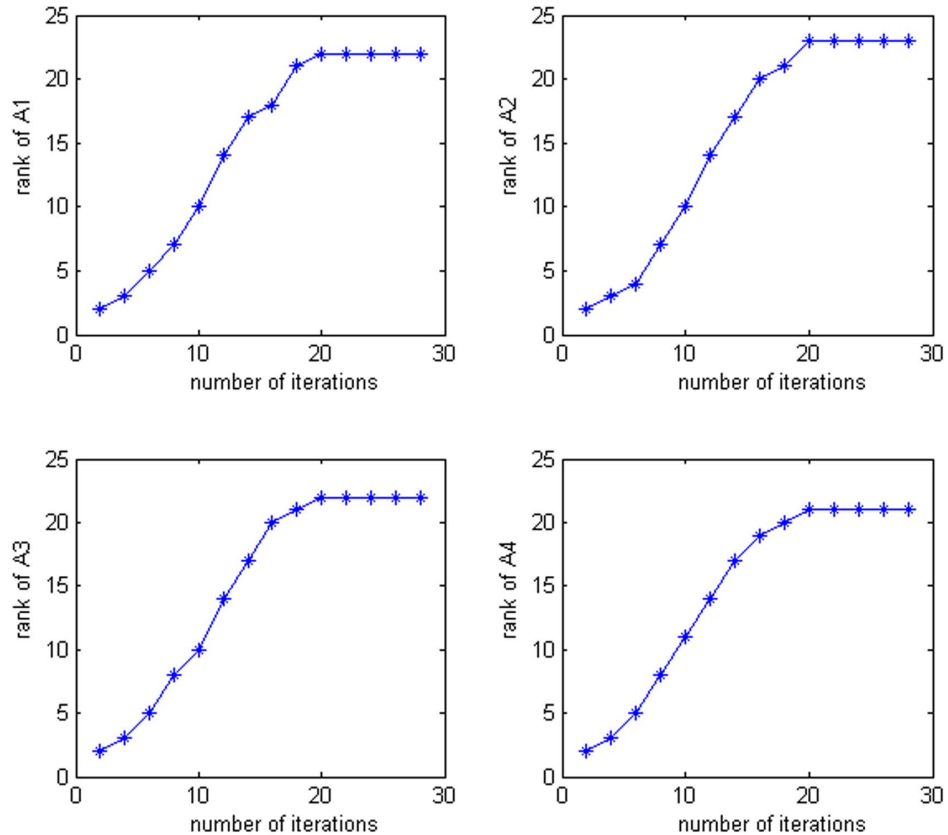
Figure 6. The relation between ranks of face matrices and number of iterations. Top-left represents the relation between the rank of first class face matrix and the number of iterations on Yale B Face Database with 40 training samples. Top-right, bottom-left and bottom-right represent that of second, third and forth class face matrices, respectively.
doi:10.1371/journal.pone.0110318.g006

Table 3. The recognition rates on AR Face Database.

Algorithms	ESRC_LR	FDDL	SRC	SVM (RBF)	SVM (linear)	NN
Recognition rate (%)	97.66	91.29	89.14	73.14	66.00	75.94

$$= \arg\min_{L} \frac{\lambda}{u}\|S\|_1 + \frac{1}{2}\left\|S - \left(\frac{1}{u}Y + D - L\right)\right\|_F^2 \qquad (11)$$

The ALM algorithm for Robust PCA is summarized as follows:
Algorithm: ALM algorithm for Robust PCA [27]
Input: Corrupted matrix $D, \lambda, S = Y_0 = 0,\; u > 0$.
Step1: while not converged do
    Compute L_{k+1} according to formula (10);
    Compute S_{k+1} according to formula (11);
    Compute $Y_{k+1} = Y_k + u\,(\,D - L_{k+1} + S_{k+1})$;
End while
Output: L, S.

The ESRC_LR Based Face Recognition Algorithm

Eigenfaces of Face Images

Generally speaking, eigenfaces of face images are defined as several special faces which capturing intrinsic structural information of face space. They are similar to a group of bases of linear space in matrix theory. For sparse representation, the number of training samples may be considerable. This probably results in redundancy of the constructed dictionary and more time expending during practical applications. Actually, by applying mathematical methods to extract eigenfaces and combining them to construct a compact dictionary for sparse representation, we still can obtain a sparse solution by solving l_1-norm minimization problem [28] for each testing sample. Mathematically, the face images dataset matrix A can be decomposed as the product of two matrices:

$$A = W \times B \qquad (12)$$

where matrix A is of size $m \times n$ with m dimensions and n samples, matrix W is of size $m \times p$ and each of the p columns is defined as an eigenface, matrix B is of size $p \times n$ and each of the n columns represents eigenface expression pattern of corresponding sample.

Figure 1 shows the relation between face matrix A and eigenface matrix W.

The methods about how to extract eigenfaces (eigengenes or metasamples) have been published in many literatures [29–31]. E.g., Alter et al. [29] extracted eigengenes by SVD to transform the "genes × samples" space to the diagonalized "eigengenes × eigenarrays" space. Zheng et al. [30] applied Independent Component Analysis (ICA) to model genes expression data regarding samples as a linear mixture of independent basis snapshots (i.e., metasamples). Brunet et al. [31] similarly utilized a small number of metasamples extracted by Non-negative Matrix Factorization (NMF) to represent testing samples. Consequently, we can extract eigenfaces from face data by means of several methods such as SVD, ICA and NMF.

Eigenface-based Sparse Representation Classification

Owing to the characteristic of eigenfaces, in this paper we extract eigenfaces from face images of each class, respectively, and then combine them instead of the original training data to design a compact dictionary for sparse representation. In practice, each sub-dataset (i.e., including only one class) A_i is factorized into two matrices:

$$A_i = W_i \times B_i \qquad (13)$$

where i represents the ith class samples. Matrix W_i is of size $m \times p_i$ and each of the p_i columns represents an eigenface.

After extracting the eigenface matrix W_i from each class, we combine them to build a collaborative dictionary W for sparse representation, i.e., $W = [W_1\ W_2\ \ldots\ldots\ W_k]$, where k represents the number of classes. Now, given a testing sample y, we compute its sparse representational solution over the collaborative dictionary, i.e., $y = W\,x$. Ideally, $x = [0, \cdots\cdots, 0, a_{i,1}, \cdots\cdots, a_{i,n_i}, 0, \cdots\cdots, 0]$ is a sparse vector whose entries are zeros except for these associating with the ith class. We identify the class label to which a testing sample belongs by analyzing its nonzero entries of coefficient representation vector [4,28].

Obviously, for a $m \times n_e$ matrix W (n_e represents the total number of eigenfaces), if $m > n_e$, $y = Wx$ is over-determined and its correct solution is unique. But in the face recognition system,

Table 4. The comparison of kernel optimization based algorithms with our algorithm.

	KSDAH	KDAH	KNDAH
Nearest Mean	88.1	87.5	71.3
Nearest Neigbor	96.7	88.3	69.2
Linear SVM	88.1	87.5	79.4
Method in[37]	96.6	90.6	70.9
ESRC_LR	89.27		

KSDA, KDA and KNDA present kernel discriminant analysis, kernel subclass discriminant analysis and kernel non-parametric discriminant analysis, respectively. Subscript 'H' presents homoscedastic-based optimization.

Table 5. The recognition rates on ORL Face Database.

Algorithms	ESRC_LR	FDDL	SRC	SVM (RBF)	SVM (linear)	NN
Recognition rate (%)	96.88	95.31	95.5	85.00	81.87	72.25

$y = Wx$ typically represents under-determined system, and its solution is not unique. To achieve a sparse solution, we adopt l_1-norm minimization:

$$x_1 = \arg\min_x \|x\|_1 \quad s.t. \quad y = Wx \qquad (14)$$

The standard linear programming methods is applied to solve (14) in polynomial time. Moreover, to certain degree of noises, we adopt the following objective function to solve x, i.e.,

$$J(x,\sigma) = \min_x \left\{ \|y - Wx\|_2 + \sigma\|x\|_1 \right\} \qquad (15)$$

where the positive parameter σ is a scalar regulator.

In practice, model errors and certain degree of noises inevitably cause some nonzero entries associating with different classes, which may bring about misclassification. To design a more robust classifier, we compute reconstruct residual as the basis of judgment [4], i.e.,

$$\min_i r_i(y) = \|y - W\theta_i(x)\|_2 \qquad (16)$$

where $\theta_i(x)$ is a vector whose nonzero entries are these from class i in x.

The eigenface-based sparse representation classification algorithm is described as follows:

Input: Matrix of training samples $A = [A_1 \ A_2 \ \cdots\cdots \ A_k]$($k$classes); a testing sample$y$.

    Step1: Extract the eigenfaces matrix $W = [W_1 \ W_2 \ \cdots\cdots \ W_k]$ by SVD;

    Step2: Solve the optimization problem defined in (15);

    Step3: Compute the residuals through formula (16);

Output: *identify* $(y) = \arg\min_i r_i(y)$.

Overview of Our Method

Figure 2 shows the overview of our method. For a face database, we firstly divide it into two parts, i.e., the training subset and the testing subset. Then, we extract low rank images for each class images in training subset by the Robust PCA, which will make the training images as clean as possible. Next, SVD is applied to extract the eigenfaces from the clean images. These eigenfaces are used as a compact dictionary for sparse representation based classification. We named our strategy as Low-Rank and Eigenface based Sparse Representation Classification (i.e., ESRC_LR). Finally, we apply the proposed ESRC_LR to classify all of the testing samples.

Since the face images are captured in real world and more or less corrupted by some negative factors such as varying illumination, shadowing, uniform noises and even random block occlusions, we had better adopt a preprocessing operation to acquire clean data as much as possible. Aiming at alleviating or eliminating the influence of negative factors, we extract the low-rank images by Robust PCA to get relatively more clean face images. In addition, we apply SVD to extract eigenfaces from the corresponding low-rank images. These eigenfaces could all-rightly describe intrinsic structural information of corresponding face spaces. By combining all of eigenfaces to construct a compact dictionary, we can improve the performance of standard sparse representation based classification algorithm. In practice, to alleviate the influence of noises and model error, we classify a testing sample based on how well it can be reconstructed by the representational coefficients of corresponding class.

The proposed ESRC_LR algorithm is summarized as follows:

Input: matrix of training samples $A = [A_1 \ A_2 \ \cdots\cdots \ A_k]$ ($k$classes); a testing sample$y$.

    Step1: while $i < =k$ do

$$\min_{L_i, S_i} \ \|L_i\|_* + \lambda\|S_i\|_1 \quad s.t. \quad A_i = L_i + S_i$$

end while

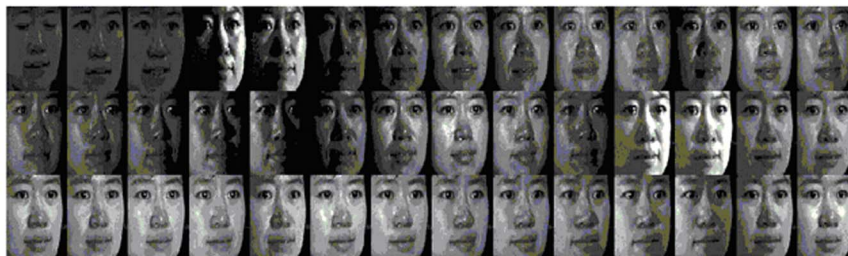$$L = [L_1 \ L_2 \ \cdots\cdots \ L_k]$$

    Step2: while $i < =k$ do

$$L_i = W_i \times B_i$$

end while

$$W = [W_1 \ W_2 \ \cdots\cdots \ W_k]$$

Figure 7. Some examples from the CMU-PIE Face Database with variation in illumination, expression.

Table 6. The recognition rates on CMU-PIE Face Database.

Algorithms	ESRC_LR	FDDL	SRC	SVM (RBF)	SVM (linear)	NN
Recognition rate (%)	97.97	97.01	97.89	97.72	94.37	90.28

Step3: Solve the optimization problem $J(x,\sigma) = \min\{\|y - Wx\|_2 + \sigma\|x\|_1\}$;

Step4: Compute the residuals $\min r_i(y) = \|y - W\theta_i(x)\|_2$;

Output: *identify* $(y) = \arg\min_i r_i(y)$;

Experimental Results

Firstly, in this section, we conduct experiments to illustrate how to make an optimal choice for parameter λ in Equation (4) and the influence of different number of eigenfaces to recognition rate on the extended Yale B Database [32]. Secondly, we apply the proposed ESRC_LR algorithm to face recognition on four publicly available face datasets, i.e., the extended Yale B Database, the AR Database [33], the ORL Database [34] and CMU-PIE Face Database [35]. Thirdly, we evaluate the robustness of ESRC_LR algorithm to different percentage of uniform noises and block occlusions. The summary of utilized databases are listed in Table 1:

To show the effective of our method, we also used other methods to classify the datasets for comparison. They are Fisher Discrimination Dictionary Learning algorithm (FDDL) [13], Sparse Representation based Classification (SRC) [4], Support Vector Machine (SVM) and Nearest Neighbor (NN). In addition, during all of the following experiments, we set the number of

iterations to 25 for learning dictionary in FDDL. Parameter σ in formula (15) is fixed to 0.15 for SRC according to [20]. SVM with different kernel function (i.e., linear kernel and RBF kernel) is applied and Nearest Neighbor with K = 1.

The Choice of Parameters

Tuning Parameter λ for Face Matrix A. In Equation (4), λ presents the percentage of sparse errors. The low-rank representation results are extremely sensitive to parameter λ. Although we could compute it using Equation (8), it is not optimal. In practice, we initialize λ by Equation (8) and then increase or decrease λ step by step at interval of 0.005 to choose the best one. We used the experiment on extended Yale B Database to adjust parameter λ. The former 40 images were selected as training images and the remaining as testing images, and all of the images were down-sampled with ratio 1/8.

Figure 3 shows the relation between recognition rates and the corresponding parameter λ. From this figure we can confirm that 0.0445 is the best choice for λ, since the recognition rate is the highest one with this value. Through this way we can also obtain the best λ for other different databases.

Influence of Different Number of Eigenfaces. As mentioned above, eigenfaces can capture intrinsic structural information of face space, so different number of eigenfaces performs
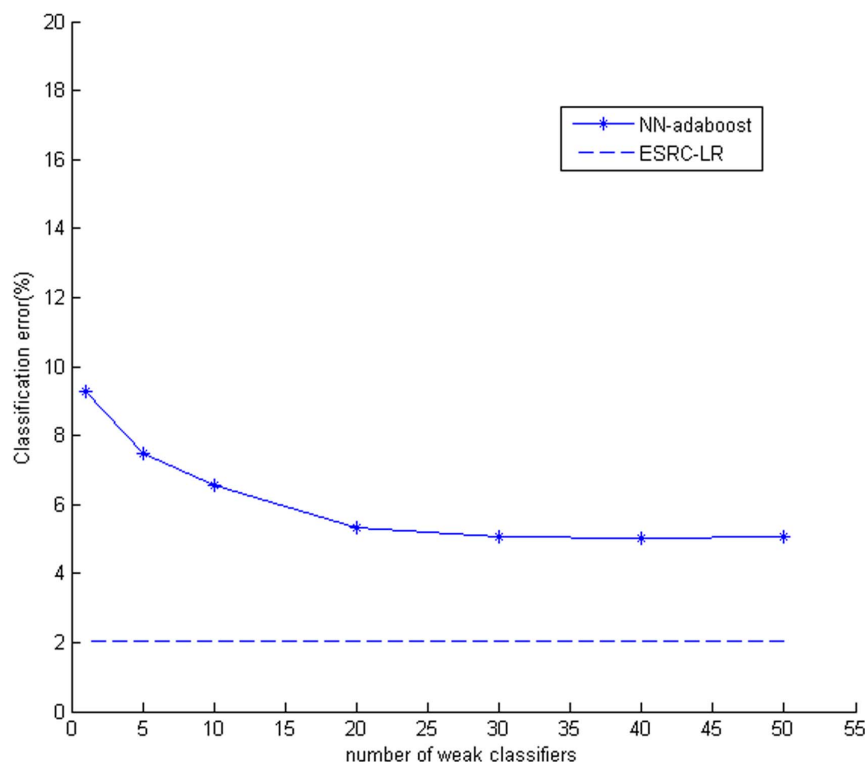
Figure 8. Classification errors under different number of weak classifier.
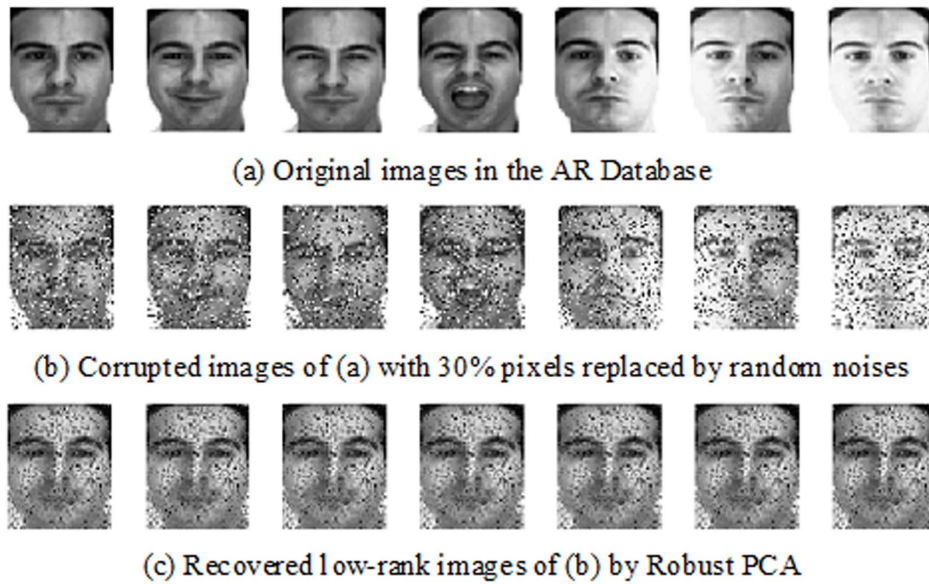
Figure 9. An example from the AR Database with 30% pixels corruptions. The top row is the original images. The middle row shows corrupted images of (a) with 30% pixels replaced by random noises. The noise values are random selected from [0, 255] and the locations are unknown. The below row is the recovered low-rank images of (b) by Robust PCA algorithm.
doi:10.1371/journal.pone.0110318.g009

different performance. Need not go into detail that overmuch eigenfaces result in much time expending in applications, and too few may be insufficient to express integrity of face space, classification accuracy is not guaranteed as a consequent. Experimental results about the performance of different number

of eigenfaces on extended Yale B Database are listed in Figure 4. We set $\lambda = 0.0445$ to extract the low-rank images of training samples by Robust PCA algorithm.

Figure 4 shows that the recognition rate goes up promptly with the increasing of the number of eigenfaces, and it is low when the
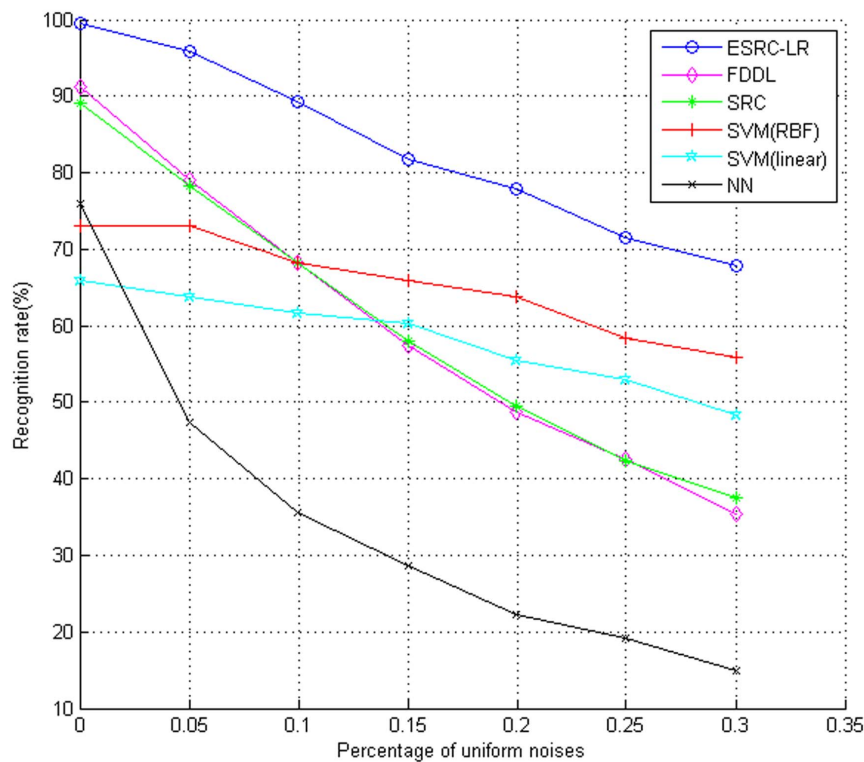
Figure 10. Recognition accuracy on the AR Database with different percentage of pixels corruptions.
doi:10.1371/journal.pone.0110318.g010
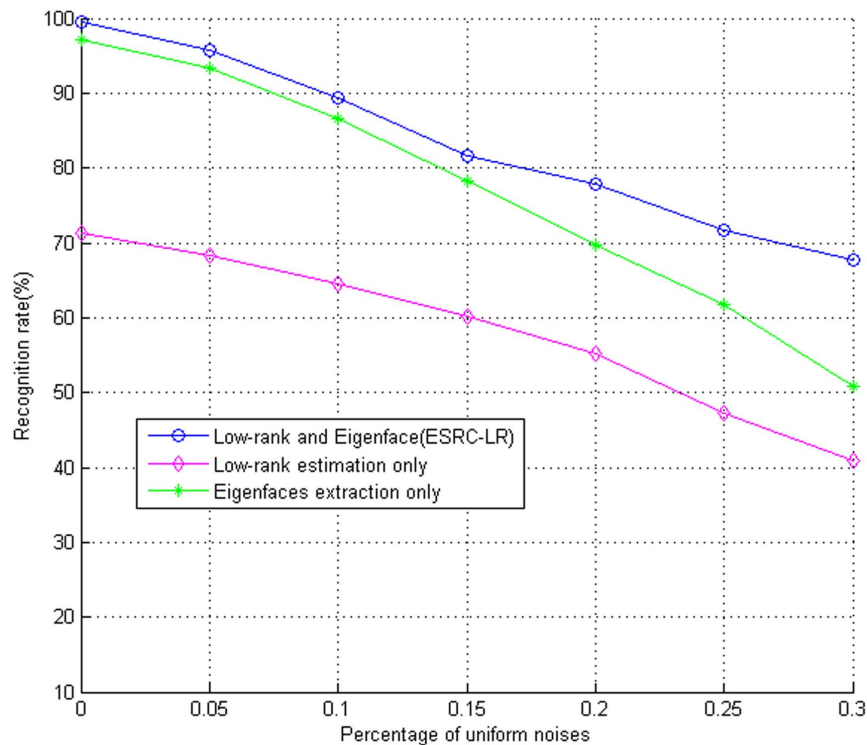
Figure 11. The separate effect of these two sub-processes of ESRC_LR algorithm on the AR Database with increasing corrupted pixels.
doi:10.1371/journal.pone.0110318.g011

number of eigenfaces is less than three. The reason should be that minority eigenfaces cannot integrally capture inherent structural information of face space. Therefore, some testing samples can not be well reconstructed by these eigenfaces. After greater than 4, with the further increase of the number of eigenfaces, recognition rate rises slowly and reaches a relatively steady point, which illustrates the selected eigenfaces are enough to span face space. Certainly, to different face databases, the suited number of eigenfaces still needs to conduct experiments to determine.

Face Recognition without Occlusions

In this subsection, we apply the proposed ESRC_LR algorithm to classify face images on four popular databases (i.e., the Extended Yale B Database [32], the AR Database [33], the ORL Database [34] and CMU-PIE Face Database [35]), and compare it with FDDL[13], SRC[4], SVM with linear kernel function and RBF kernel function and NN with K = 1.

Evaluation on the Extended Yale B Database. The Extended Yale B Database consists of 2414 frontal-face images of 38 individuals. The images were captured under various laboratory-controlled lighting conditions. We selected the former 40 images as training images and the reminder as testing images

Figure 12. An example of images from UMIST Face Database with random block occlusions ($\eta = 0.3$).
doi:10.1371/journal.pone.0110318.g012
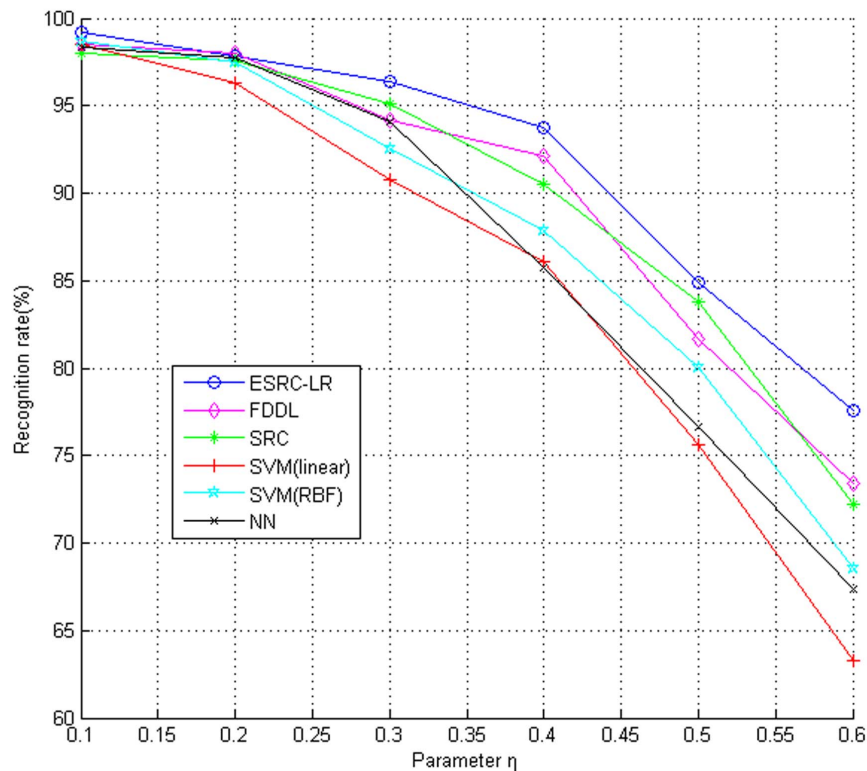
Figure 13. Recognition rates on UMIST Face Database with different level of block occlusions.
doi:10.1371/journal.pone.0110318.g013

and all of the images were down-sampled with ratio 1/8. In the experiment we chose the popular methods such as SRC, SVM, NN and the recently published FDDL for comparison. The recognition rates are listed in Table 2.

Figure 5 shows an example of the extracted low-rank and approximate images on Extended Yale B Database. It is easy to find that the low rank images are similar to original images but illumination difference and expression changing have been alleviated. Figure 6 illustrates the relation between the ranks of face images of four classes and the number of iterations. We can see from figure 6 that the low rank algorithm is convergent. The rank of each class is stabilized at about 22 finally, which is much smaller than the rank of original corrupted face images (i.e., about 40). The reduction of rank leads to the alleviation of illumination difference and expression changing. In the experiment we selected 14 eigenfaces from each class according to subsection 4.1, and set $\lambda = 0.0445$ to extract the low rank images. Comparing with other classical methods shown in Table 2, the proposed ESRC_LR algorithm exhibits better performance than FDDL, SRC, SVM(RBF), SVM(linear) and NN(K = 1).

Evaluation on the AR Database. The AR face database consists of over 4000 frontal images with 126 individuals. For each individual, 26 pictures were composed of two sessions. In each session, 7 pictures were about varied expression, 3 pictures were covered by mask and the rest 3 pictures wore sunglasses. In our experiment we chose 50 males and 50 females and all of images were cropped to 40×40 pixels and converted to grayscale. For each individual, the 7 images with illumination and expression changes from session 1 were used for training, and the other 7 images with the same changes from session 2 were chosen for testing. We fixed $\lambda = 0.0183$ for extracting the low rank representation of face matrix, and selected 4 eigenfaces for each

class to construct a compact dictionary. The comparison between several competing methods is shown in Table 3. It can be seen that the proposed algorithm (recognition rate 97.66%) improves about 6.37% over the second best performance (FDDL: 91.29%), and 8.53% over the third best performance (SRC: 89.14%). The experimental results show that the low-rank processing and eigenfaces extraction can do improve the performance of dictionary, thereby, coding the query image over the learnt dictionary is more reasonable. From this experiment, we can also draw a conclusion that our method is quite effective to deal with illumination difference and expression changing.

Recent years, the kernel optimization based algorithms are popular and competitive. It is necessary to compare our algorithm with some kernel optimization based algorithms mentioned in [36] on the AR database. For simplification, we also resized the images into 29×21 pixels. The images in session 1 were used for training and those in session 2 for testing. One could refer to [36] for details. The results are shown in Table 4 (some results originate from [36] directly). As we all known, kernel discriminant analysis (KDA) maps the original data with unequal covariances into a homoscedastic space. However, to a multi-model structure in each class, KDA may be limited to smooth solutions which would result in large classification error. In our method, we try to seek a better representation over a learnt dictionary for each testing sample (i.e., learning a dictionary with low-rank processing and eigenface extraction). By comparison, our method achieves slightly better results than KDA. However, to multi-model class representations, kernel based subclass DA(KSDA) can find some smoother functions than KDA so as to carry much smaller classification error (in some case, surpass our algorithm a lot).

Evaluation on ORL Face Database. The ORL database contains 400 images in total, each of 40 individuals has ten

different images shot under different times, varying the lighting, facial expressions and facial details. The background of the images is uniform while the subjects are in frontal, upright positive. For each individual, we randomly selected 8 images as training samples and the rest as testing samples and repeated the experiment 10 times. All of images were normalized to 30×25 pixels. We fixed $\lambda = 0.0408$ for extracting the low rank representation of face matrix, and selected 6 eigenfaces for each class to construct a compact dictionary. The average recognition rates are shown in Table 5. According to this Table we can easily see that the recognition rate of ERSC_LR reaches 96.88% which outperforms FDDL (95.31%) and SRC (95.50%) and far better than SVM(RBF) (85.00%), SVM(linear) (81.87%) and NN(K = 1) (72.75%). The ESRC_LR algorithm has at least 1.3% improvement over the second best performance, which also illustrates that our method is more powerful than other methods.

Evaluation on CMU-PIE Face Database. The CMU-PIE database contains about 41,368 images of 68 distinct individuals. The images of each subject are captured under different poses, variable illumination conditions, and with different expressions. In this paper, one of near frontal pose subsets, namely C05, is chosen for experiments. There are 3332 images in total with size of 64×64 pixels which already detect the face and eyes regions from the original images. Figure 7 shows an example of images in the subset. For each person, we randomly selected about half number of images as training samples and the rest as testing samples and repeated each experiment 10 times. We fixed $\lambda = 0.0206$ for extracting the low rank representation of face matrix, and selected 20 eigenfaces for each class to construct a compact dictionary. The average recognition results of different approaches are shown in Table 6. From which, we can see that the recognition rate of ERSC_LR algorithm (97.97%) is better than SVM (linear) (94.37%) and NN (90.28%) but similar to FDDL (97.01%), SVM (RBF) (97.72%) on CMU-PIE Face Database. Actually, SVM with RBF kernel function nonlinearly maps the original face data into a high-dimensional space to make linearly impartible problem divisible, so the performance of RBF kernel function is superior to linear kernel function according to our experimental results. Notice that the mapped data in high-dimensional space may still be linearly impartible due to the curse of dimensionality. However, in our algorithm, each testing sample is represented by eigenfaces extracted from low-rank images as a coefficient vector so that it is easier for classification than that extracted from the corrupted data. Although FDDL uses of the fisher discrimination criterion to learn a structured dictionary, the large-scale sparse errors in training samples limit its performance. In fact, our method and FDDL are both improved methods based on standard SRC. Table 6 shows that our method obtains slightly better recognition accuracy than SVM with RBF kernel function and FDDL.

Recently, the cascade classifier, another kind of face recognition system, is popular. Shan et al. [38] proposed the AdaBoost Gabor Fisher Classifier for robust face recognition, in which a chain Adaboost learning method based on Booststrap re-sampling was proposed and applied to face recognition. To compare ESRC_LR algorithm with the cascade classifier based face recognition system, we made another experiment on the CMU_PIE face dataset. The popular nearest neighbor based adaptive boosting (NN_adaboost) classifier was applied and we also set the same experimental condition as above. Different number of weak classifiers was cascaded and the corresponding classification results were shown in Figure 8.

From Figure 8 we can see that classification error decreases gradually with the increase of number of weak classifiers. In the experiment, the classification error of single NN classifier is 9.72% according to the above experiment. After cascading 50 NN classifiers, the classification error drops to about 5%, Owing to the augment of weight for misclassified samples to make them classify easily in each training process of weak classifier, the final strong classifier (cascade of weak classifiers) achieves much better performance. However, experimental result also shows our ESRC_LR algorithm (classification error: 2.03%) outperforms the NN_adaboost algorithm in spite of increasing the number of weak classifiers (classification error is no longer change apparently for more than 30 weak classifiers).

Face Recognition with Occlusions

One of the most fascinating features of sparse representation based algorithms is the robustness to occlusions and sparse noises. Since they universally exist in both training and testing images, in this subsection we investigate the robustness of our algorithm to random distributed noises and random block occlusions. Firstly, we evaluate the robustness to random pixel corruptions on the AR Database. Secondly, the UMIST Face Database [39] is adopted to test the robustness to random block occlusions. We still choose previous methods (FDDL, SRC, SVM (linear and RBF), NN(K = 1)) as the baselines.

Robustness to Random Noises. Firstly, we investigate the robustness of our algorithm to random noises on the AR Database. For each individual, the seven clean images from Session 1 were selected for training, and the seven images from Session 2 were chosen for testing. All of the images were resized into 40×40 pixels and a certain percentage of pixels were replaced by uniformly distributed random values within [0, 255]. The corrupted pixels were randomly chosen in both training images and testing images and the locations were blind to algorithms. Figure 9 shows an example of face images corrupted by random pixel noises and the corresponding low-rank images extracted by Robust PCA algorithm. The recognition rates under different level of noises are given in figure 10. Obviously, all classification performance decreases with the increasing random noises. Moreover, the advantage of ESRC_LR algorithm over the other popular methods is clear. With higher random noises, the recognition rate of ESRC_LR algorithm decreases slowly, and outperforms FDDL and SRC algorithm by 15% improvement on average.

As we can see from the above experiment that the combined performance of low-rank estimation and eignefaces extraction outperforms the other popular algorithms evidently such as FDDL, SRC, SVM(linear and RBF) and NN(K = 1). To illustrate the separate effect of the two sub-processes, i.e., low-rank matrix estimation only and eigenfaces extraction only, we also conduct an experiment to research independent effect of the two sub-processes on the AR database with increasing corrupted pixels. The scenarios are: (i) low rank estimation followed by eigenfaces extraction, (ii) low-rank matrix estimation only, and (iii) eigenfaces extraction only. Figure 10 presents the experimental results of three different scenarios.

The comparison results between ESRC_LR and its two sub-processes (i.e., low-rank matrix estimation only and using eigenfaces only) are shown in figure 11. From this figure we can see that the proposed ESRC_LR algorithm outperforms its two sub-processes. With lower random noises, the ESRC_LR algorithm performs better than independent eigenfaces by about 2% improvement but far better than low-rank estimation. With the increase of random noises, the ESRC_LR algorithm is superior to the two sub-processes more obviously. The experiment also illustrates that the two sub-processes have a positive effect on the combined ESRC_LR algorithm.

Robustness to Random Block Occlusions. In this subsection we test the robustness of our method to random block occlusions on UMIST face database. The UMIST Face Database consists of 564 images of 20 individuals and each individual is shown in a range of poses from profile to frontal views. In the experiment, 15 individuals' images were used and all of images were resized into 30×20 pixels. We replaced a randomly located block of each image with an unrelated random image. The values in each block were randomly chosen from [0, 255] and the position of block in each image was randomly selected. A parameter η was set to control the size of block which met the following relations:

$$block\_h = floor(sqrt(im\_h * im\_w) * \eta) \qquad (17)$$

$$block\_w = floor(sqrt(im\_h * im\_w) * \eta) \qquad (18)$$

where $block\_h$ and $block\_w$ were the height and width of block, im_h and im_w were the height and width of images, respectively. Figure 12 shows an example of training images and testing images with random block occlusions ($\eta = 0.3$) from UMIST Face Database. We adopted 4-fold cross validation to evaluate our algorithm. The recognition rates under different level of occlusions are given in figure 13. We can see the proposed ESRC_LR algorithm performs better than the other popular algorithms such as FDDL, SRC, SVM(RBF), SVM(linear) and NN(K = 1). For small block corruptions, the advantage of ESRC_LR algorithm is not obvious comparing with the other methods. However, as the block occlusion becoming larger, our ESRC_LR algorithm decreases slowly and outperforms other algorithms more and more apparent. Which shows the strong robustness of our method to random block occlusion.

Conclusions

In this paper, we present an improvement to the well known Sparse Representation based Classification for face recognition. We firstly extract low rank images for each class in training subset to alleviate the influence of noises such as illumination difference and occlusions. Then, SVD is applied to extract eigenfaces from the clean face images. These eigenfaces are organized to construct a compact but discriminative dictionary for sparse representation based classification. We evaluate the proposed ESRC_LR algorithm under different conditions, i.e., clean images, uniform distributed noises and random block corruptions. Experimental results verify that our ESRC_LR algorithm is advantage and robust. However, how to alleviate the influence of large but sparse noises is still worth deep studying. Moreover, A new angle of view about how to extract effective feature on low-rank images to construct a better dictionary is worthy of consideration.

Author Contributions

References

1. Mairal J, Elad M, Sapiro G (2008) Sparse representation for color image restoration. IEEE Trans Image Process 17: 53–69. doi:10.1109/TIP.2007.911828

2. Donoho DL (2006) Compressed sensing. IEEE Trans Inf Theory 52: 1289–1306. doi: 10.1109/TIT.2006.871582

3. Zhang HC, Nasrabadi NM, Zhang YN, Huang TS (2012) Joint dynamic sparse representation for multi-view face recognition. Pattern Recognit 45: 1290–1298. doi: 10.1016/j.patcog.2011.09.009

4. Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y (2009) Robust face recognition via sparse representation. IEEE Trans Pattern Anal Mach Intell 31: 210–227. doi: 10.1109/TPAMI.2008.79

5. Li KL, Zhang GY, Li X, Xie J (2010) Face recognition based on improved retinex and sparse representation. Pattern Recognit 15: 2010–2014. doi: 10.1016/j.proeng.2011.08.375

6. Mairal J, Bach F, Ponce J, Sapiro G, Zisserman A (2009) Supervised dictionary learning, Adv Neural Inf Process Syst 1033–1040.

7. Huang J, Huang X, Metaxas D (2009) Learning with dynamic group sparsity. Proc IEEE Int Conf Comput Vis 64–71. doi: 10.1109/ICCV.2009.5459202

8. Liu Y, Wu F, Zhang Z, Zhuang Y, Yan S (2010) Sparse representation using nonnegative curds and whey. Proc IEEE Conf Comput Vis Pattern Recognit 3578–3585. doi: 10.1109/CVPR.2010.5539934

9. Yang M, Zhang L, Yang J, Zhang D (2011) Robust sparse coding for face recognition. Proc IEEE Conf Comput Vis Pattern Recognit 625–632. doi: 10.1109/CVPR.2011.5995393

10. Aharon M, Elad M, Bruckstein A (2006) K-SVD: An algorithm for designing over-complete dictionaries for sparse representation. IEEE Trans Signal Process 54: 4311–4322. doi:10.1109/TSP.2006.881199

11. Mairal J, Bach F, Ponce J, Sapiro G, Zisserman A (2008) Discriminative learned dictionaries for local image analysis. Proc IEEE Conf Comput Vis Pattern Recognit 1–8. doi: 10.1109/CVPR.2008.4587652.

12. Mairal J, Bach F, Ponce J, Sapiro G (2009) Online dictionary learning for sparse coding. Proc Annual International Conference on Machine Learning 689–696. doi:10.1145/1553374.1553463

13. Yang M, Zhang L, Feng X, Zhang D (2011) Fisher discrimination dictionary learning for sparse representation. Proc IEEE Int Conf Comput Vis 543–550. doi: 10.1109/ICCV.2011.6126286

14. Candes EJ, Li XD, Ma Y, Wright J (2011) Robust Principal Component Analysis,J ACM. 58. doi:10.1145/1970392.1970395

15. Liu G, Lin Z, Yan S, Sun J, Yu Y, et al. (2013) Robust recovery of subspace structures by low-rank representation. IEEE Trans Pattern Anal Mach Intell 171–184. doi: 10.1109/TPAMI.2012.88

16. Ma L, Wang CH, Xiao BH, Zhou W (2012) Sparse Representation for Face Recognition based on Discriminative Low-Rank Dictionary Learning. Proc IEEE Conf Comput Vis Pattern Recognit 2586–2593. doi: 10.1109/CVPR.2012.6247977

17. Zhou TY, Tao DC (2011) GoDec: Randomized Low-rank & Sparse Matrix Decomposition in Noisy Case. Proc Int Conf Mach Learn USA 33–40.

18. Wang HQ, Huang DS (2006) Regulation Probability Method for Gene Selection. Pattern Recognit Lett 27: 116–122. doi: 10.1016/j.patrec.2005.07.007

19. Liebermeister W (2002) Linear Modes of Gene Expression Determined by Independent Component Analysis. J Bioinform 18: 51–60. doi:10.1093/bioinformatics/18.1.51

20. Zheng CH, Zhang L, Ng To-Yee, Shiu S CK, Huang DS (2011) Metasample-Based Sparse Representation for Tumor Classification. IEEE/ACM Trans Compu Biol Bioinform 8: 1273–1282. doi: 10.1109/TCBB.2011.20

21. De la Torre F, Black M (2001) Robust principal component analysis for computer vision. Proc IEEE Int Conf Comput Vis 362–369. doi: 10.1109/ICCV.2001.937541

22. Gnanadesikan R, Kettenring J (1972) Robust estimates, residuals, and outlier detection with multiresponse data. Int J Biom 28: 81–124.

23. Ke Q, Kanade T (2005) Robust'l-norm factorization in the presence of outliers and missing data by alternative convex programming. Proc IEEE Int Conf Comput Vis Pattern Recognit 1: 739–746. doi: 10.1109/CVPR.2005.309

24. Fischler M, Bolles R (1981) Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM (1981) 381–385. doi:10.1145/358669.358692

25. Bertsekas DP (1982) Constrained optimization and Lagrange multiplier methods. Computer Science and Applied Mathematics, Boston: Academic Press.

26. Bertsekas DP (1999) Nonlinear programming. Athena Scientific Press. 708 p.

27. Lin Z, Chen M, Wu L, Ma Y (2009) The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices. doi:10.1016/j.jsb.2012.10.010 arXiv:1009.5055 [math.OC].

28. Koh S, Kim K, Boyd S (2007) An Interior-Point Method for Large-Scale l1-Regularized Least Squares. IEEE J Sel Top Signal Process 1: 606–617. doi: 10.1109/JSTSP.2007.910971

29. Alter O, Brown PO, Botstein D (2000) Singular Value Decomposition for Genome-Wide Expression Data Processing and Modeling. Proc Natl Acad Sci 97: 10101–10106. doi: 10.1073/pnas.97.18.10101

30. Huang DS, Zheng CH (2006) Independent Component Analysis-Based Penalized Discriminant Method for Tumor Classification Using Gene Expression Data. Bioinform 22: 1855–1862. doi:10.1093/bioinformatics/btl190

31. Brunet JP, Tamayo P, Golub TR, Jill PM (2004) Metagenes and Molecular Pattern Discovery Using Matrix Factorization. Proc Natl Acad Sci 101: 4164–4169. doi: 10.1073/pnas.0308531101

32. Lee K, Ho J, Kriegman D (2005) Acquiring linear subspaces for face recognition under variable lighting. IEEE Trans Pattern Anal Mach Intell 27: 684–698. doi: 10.1109/TPAMI.2005.92

33. Martinez A (1998) The AR face database. CVC Tech Rep 24.

34. Samaria F, Harter A (1994) Parameterisation of a stochastic model for human face identification. Proc IEEE Workshop Appl Comput Vis 138–142. doi: 10.1109/ACV.1994.341300

35. Terence S, Simon B, Maan B (2002) The CMU Pose, Illumination, and Expression (PIE) database. Proc IEEE Int Conf Autom Face Gesture Recognit 46–51. doi: 10.1109/AFGR.2002.1004130

36. You D, Hamsici OC, Martinez AM (2011) Kernel optimization in discriminant analysis. IEEE Trans Pattern Anal Mach Intell 33: 631–638. doi: 10.1109/TPAMI.2010.173

37. Pujol O, Masip D (2009) Geometry-based ensembles: Towards a structural characterization of the classification boundary. IEEE Trans Pattern Anal Mach Intell 31: 1140–1146. doi: 10.1109/TPAMI.2009.31

38. Shan SG, Yang P, Chen XL, Gao W (2005) AdaBoost Gabor Fisher Classifier for Face Recognition. Lecture notes in Computer Science, 3723: 279–292. doi: 10.1007/11564386_22

39. Graham D, Allinson N (1998) Characterising virtual eigen signatures for general purpose face recognition. NATO ASI Series F: Comput Syst Sci 163: 446–456.