Pattern Recognition 60 (2016) 13–24

Contents lists available at ScienceDirect

Pattern Recognition

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

Robust face recognition based on dynamic rank representation

Hongjun Li [a,b,*], Ching Y. Suen [b]

[a] School of Electronic Information Engineering, Nantong University, Nantong 226019, PR China
[b] Centre for Pattern Recognition and Machine Intelligence, Concordia University, Montreal, Quebec, Canada H3G 1M8

ARTICLE INFO

Article history:
Received 11 September 2015
Accepted 10 May 2016
Available online 20 May 2016

Keywords:
Face recognition
Low-rank representation
Dynamic subspace
Discriminative component
Occlusion

ABSTRACT

Robust face recognition is an active topic in computer vision, while face occlusion is one of the most challenging problems for robust face recognition algorithm. The latest research on low-rank representation demonstrated its high efficiency to subspace segmentation and feature extraction. Motivated by previous work, in this paper, we consider the problem of human face recognition from frontal views with varying illumination, as well as occlusion and disguise. We present a novel approach for face recognition by extracting dynamic subspace of images and obtaining the discriminative parts in each individual. We use these parts to represent the characteristic of discriminative components, give a recognition protocol to classify face images. The experiments carried on publicly available databases (i.e., AR, Extended Yale B, and ORL) vilidate its accuray, robustness and speed. The proposed method needs lower dimensions training samples but gains a higher recognition rate than other popular approaches.

1. Introduction

During the past few years, a low-rank model is often used to capture the important components of an observed data matrix and yield a low-dimensional representation. To deal with small sample size problem like face recognition, the dimensionality of the image is larger than the training sample size. As a result, low-dimensional representations could make the data easier to manipulate than the original data. Low-rank matrix recovery problem has been widely studied and successfully applied to many areas [1–5], such as face recognition [3], motion segmentation [4], and image segmentation [5]. More recently, there is a new extension of low-rank and sparse decomposition called the low-rank representation (LRR) [6]. With respect to the case that the data samples are contaminated by outliers or arbitrary errors, LRR can recover the row space of the original data exactly and detect the outliers under certain conditions as well [7]. For handling the case where the number of observation data samples is insufficient or samples themselves are too badly corrupted in face recognition, this process becomes tedious as the class number grows. Liu and Yan [8] proposed a latent low-rank representation approach, and applied to subspace segmentation and feature extraction. Based on LRR, a fixed-rank representation method [9] is proposed for unsupervised learning on image features. In Ref. [10], a non-negative low-rank and sparse graph is constructed for semi-supervised learning. Also, low-rank constraints have been introduced to the visual domain adaption in Ref. [11] and transfer subspace learning in Ref. [12]. Only recently, some interest in deep learning to face recognition has emerged [13–16]. Taigman et al. [13] showed excellent results using convolutional neural network to learn the face representation in recognition. Convolutional neural network is good, but need big data to train; it is slowly not only training, but also testing. In contrast, we try to learn a dynamic subspace, it is simple and different from theses in the term of the way the network is used.

In particular, images of a human face can be well approximated by a low-dimensional subspace. Being able to correctly retrieve this subspace is crucial in many applications such as face recognition and alignment. Recently, face recognition approaches based on low-rank matrix have led to the state-of-the-art performance [17–21]. Liu et al. [10] proposed that the task of clustering face images into their respective classes can be converted into a subspace segmentation problem if we assume that face images from different individuals lie near independent subspaces. Ma et al. [17] integrated rank minimization into sparse representation and achieved impressive results especially when corruption existed. Though method in [17] was claimed to be able to handle noisy samples as well, it may suffer from certain information loss because of the low-rank regularization. Ming et al. [18] proposed a double low-rank matrix recovery method to learn low-rank subspaces from face images, where it takes the recovery of row space and column space information into account simultaneously. Li

* Corresponding author at: School of Electronic Information Engineering, Nantong University, 9 Seyuan Road, Nantong 226019, PR China.
E-mail address: lihongjun103@126.com (H. Li).

http://dx.doi.org/10.1016/j.patcog.2016.05.014
0031-3203/

et al. [19] proposed discriminative dictionary learning with low rank regularization algorithm. It can well cope with training samples containing large noise and can still achieve impressive performance due to the low-rank regularization on the sub-dictionaries. Zhang and Yang [20] presented a LRR-based discriminative projection method for robust feature extraction, by virtue of the underlying low-rank structure of data representation revealed by LRR. In Ref. [21], a low-rank approximation method with structural incoherence is proposed for face recognition. However, realistic face images often suffer from self-shadowing, specularities, or saturations in brightness, which make it a difficult task and subsequently compromise the recognition performance.

Inspired by the previous work, we know that discriminative information is essential to face recognition in the presence of illumination variations, noise corruption and block occlusion. In this paper, we propose a face recognition method based on dynamic subspace analysis. First, we discuss several subspace algorithms and analyze the relation between rank of the subspace and information of samples. Second, we observe how the rank of samples influence the recognition accuracy and design a dynamic subspace model by statistic, which can acquire more exact information and is faster than traditional methods. Then, we obtain discriminative features by dynamic low-rank representation. Finally, we use K-nearest neighbor (K-NN) classifier [22] based on Euclidean distance to evaluate the quality of various transformed features.

In this paper, we use exactly low-rank to reduce the influence of occlusion in data. We propose a simple yet effective model for simultaneous feature learning and data classification. By choosing dynamic rank as an optimal subspace, our model achieves an overall optimality in recognition in same sense. Several advantages of our method are summarized as follows: (1) we try to find a dynamic rank that can more precisely predict the low-rank representation model. Therefore, we recover the face image from corruption more accurately. (2) The dynamic rank representation is an optimal solution to low-rank representation. It is robust in real face database, especially, when data sampling is insufficient. Furthermore, dynamic rank representation solves the compute complexity. (3) The algorithm in this paper is designed to decompose the corrupted data into two parts: low-rank representation of data and residual parts. Then, we design a new feature extraction method that produces solution by dynamic rank representation, which obtains a simple unique analysis solution. (4) The proposed method enhances the performance of low-rank representation in feature extraction, which can improve the efficiency of discriminative feature selection. It can achieve a competitive result with accuracy, robustness and efficiency.

To demonstrate the feasibility and effectiveness of our algorithm, we test it on some well-known face databases, including ORL [23], AR [24], and Extended Yale B [25]. The experimental results show that proposed method outperforms other methods, such as SRC [26], FDDL [27], LRR [11], Latent-LRR [12], and Deep Learning [28], with varying expressions, poses and illuminations. The remainder of the paper is organized as follows. The related works are presented in Section 2. Section 3 analyzes the dynamic subspace by low-rank representation and shows some interest of dynamic rank representation and presents the proposed method. Section 4 is devoted to experimental results and analysis, and Section 5 concludes the paper.

2. Review of previous work

2.1. Robust PCA

Robust PCA [29] is a low-rank matrix recovery method. It aims to decompose a data matrix X into $Z+E$, where Z is a low-rank matrix that we want to recover, which stands for clean data lying on a low-dimensional subspace, and E is a sparse error matrix. The separation is achieved by solving the following principal component pursuit problem:

$$\min_{A,E} \|Z\|_* + \lambda\|E\|_1 \quad s.t. \quad X = Z + E \tag{1}$$

where the nuclear norm and the l_1 norm are convex surrogates of the rank function and the l_0 pseudo norm, i.e., the number of nonzero entries, respectively. λ is a positive parameter trading off between low rankness and sparsity. Z can be exactly recovered from X as long as the rank of Z is sufficiently low and E is sufficiently spare.

2.2. Low-rank representation

More recently, low-rank representation exists that integrate subspace segmentation and noise correction into a unified framework for multiple subspace recovery problems. Generally, the LRR problem can be formulated as follows:

$$\min_{Z,E} rank(Z) + \lambda\|E\|_{2,1} \quad s.t. \quad X = DZ + E \tag{2}$$

where $Z \in R^{d\times n}$ is called the low-rank representation of data X, and $D \in R^{m\times d}$ is a dictionary that linearly spans the data space. $E \in R^{m\times n}$ is the noise and λ is a parameter that takes off the rank term and sparse term. m, n are the size of data X, d is the rank of data X. Unfortunately, the optimization problem mentioned in Eq. (2) above is NP-hard in general due to the discrete nature of the rank function. An efficient and acknowledged approach for solving rank minimization problems is to relax the matrix rank into the matrix nuclear norm. Moreover, this relaxation technique achieves a convex optimization problem, which is tractably solved. To this end, the nuclear norm minimization problem for LRR can be reformulated as follows:

$$\min_{Z,E} \|Z\|_* + \lambda\|E\|_{2,1} \quad s.t. \quad X = XZ + E \tag{3}$$

where $\|Z\|_*$ is the nuclear norm of matrix Z, and the data X itself is used as the dictionary.

The advantage of LRR mainly comes from its ability of correcting the corruptions in data automatically [30,31]. Several LRR-based algorithms [32] have been proposed to address the challenging recognition task on heavily corrupted face databases, and they can remove the shadows or specular lights from images and then significantly improve clustering accuracy. Fig. 1 shows an example of low-rank representation with face images taken from a database. We can find that the low-rank components (Fig. 1(b)) among different images look very similar to the original (Fig. 1(a)), even in the presence of expression and occlusion. It can also be seen from the error components (Fig. 1(c)) that spare errors depict the major difference between original and corresponding low-rank face images. Compared with the sparse error image from the correct individual, error images generated from the incorrect subjects do not only contain the intra-class difference, but also include a lot of detailed information, such as the edges of mouth, nose and chin, that is the inter-class difference.

If the test and training images belong to the same subject, the low-rank component can describe the common feature of this subject correctly, and the corresponding error component only reflects the sparsely intra-class variability. By contrast, if the test and training images come from different subjects, the low-rank

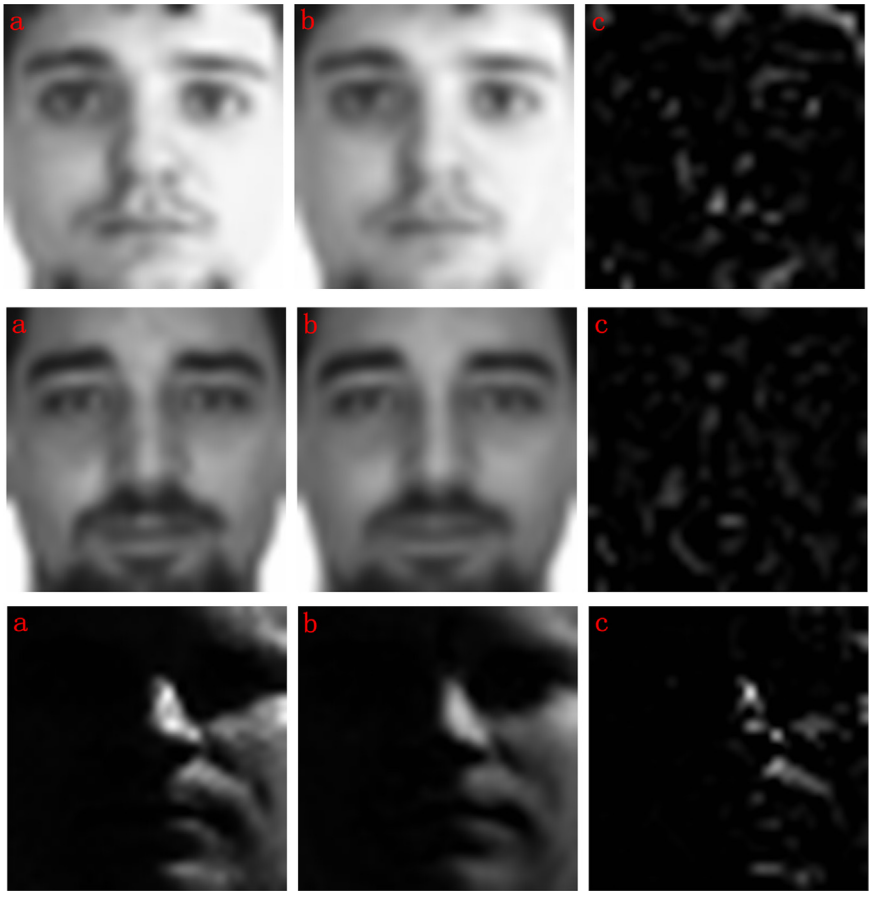

Fig. 1. Some examples of using LRR to correct the corruptions in face. Left: the original data (X); middle: the corrected data (XZ); and right: the error (E).

component may be incapable of extracting the intrinsic features of the image; hence the error component is not as sparse as the one in the same subject case. Due to the impact of nonzero entries is not absolutely sparse, and the intensity values of some pixels that supposed to be equal to zero are approximately equal to zero.

2.3. Latent low-rank representations

Latent-LRR (LatLRR) [12] is an extension of LRR that can recover the effects of unobserved hidden data. LatLRR can also extract salient features from corrupted images for classification tasks. LatLRR method is adequate to recover the hidden effects by solving the following convex optimization problem:

$$\min_{Z,L,E} \quad \|Z\|_* + \|L\|_* + \lambda\|E\|_1 \quad s.t. \quad X = XZ + LX + E \tag{4}$$

where $\lambda > 0$ is a parameter, L is low-rank part of X and $\|\cdot\|_1$ is the $l_1$norm chosen for characterizing sparse noise.

As a subspace segmentation algorithm, LatLRR could be regarded as an enhanced version of LRR, and thus obtains more accurate segmentation results. Being an unsupervised feature extraction algorithm, LatLRR can automatically extract salient features from corrupted data so as to produce effective features for classification. Fig. 2 shows an example of LatLRR with face images taken from the database. We can find that the low-rank components (Fig. 2(b)) of different images looks very similar to the original one (Fig. 1(a)), Fig. 2(c) is the salient feature and Fig. 2(d) is a sparse part. Compared with LRR, the LatLRR salient feature (Fig. 2 (c)) includes more detailed information while the sparse part (Fig. 2(d)) is sparser.

LatLRR aims to construct the dictionary by employing both observed and unobserved hidden data. As a result, LatLRR is capable of resolving the insufficient data sample problem supposing that the given dictionary is always sufficient to represent the subspaces. It is also elaborated that the hidden effects of unobserved data can also be recovered through minimizing a nuclear norm problem approximately.

LatLRR introduces hidden effects to overcome this drawback. However, there are two problems in Eq. (4). Firstly, an explanation for its observed capability to identify salient feature is unknown. And, the complexity of LatLRR depends on the dimensionality of the feature vector [33,34]. So a more accurate recovery of low-rank method is required.

3. Dynamic rank representation

3.1. Motivation

In particular, when the data is only contaminated by outliers, LRR or LatLRR may have good recovery performance. However, this choice cannot ensure the validity while the data contains other types of error. Note that such a type of data corruption commonly exists because of sensor failures, uncontrolled environment, etc. We still need to seek other approaches to overcome the recognition against occlusions. A natural question comes out, given a test image, is there any subspace which can restore most of the discriminative information? If yes, it is possible to explicitly use the optimal subspace to keep the discriminative features. As mentioned above, the goal of our proposed method is to choose the optimal subspace from different classes in which they are as distant as possible. Thus, we can obtain stable inter-class representation.

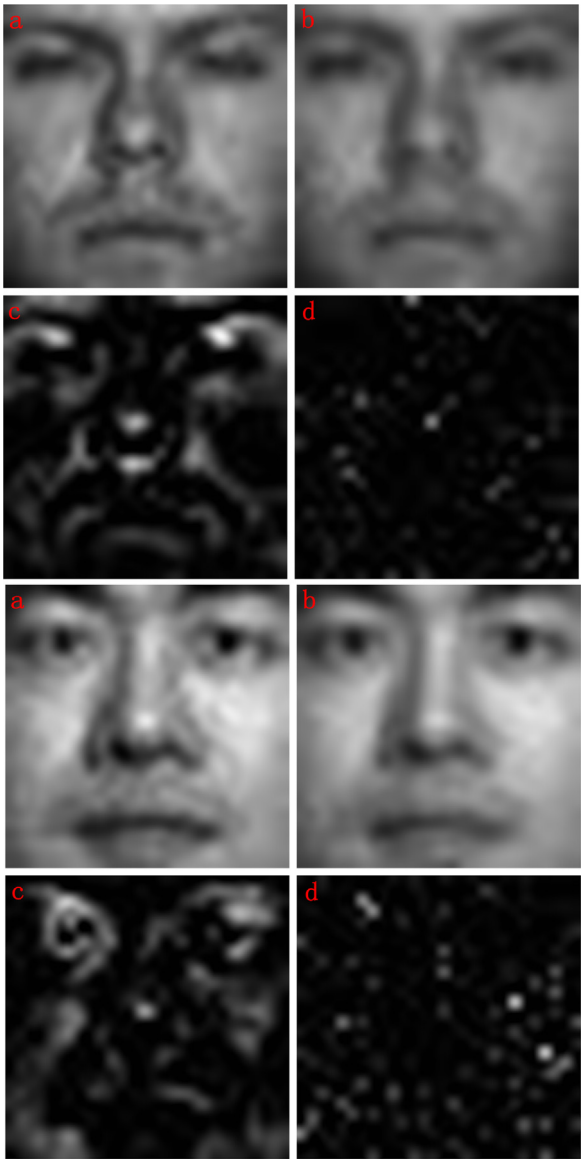

Fig. 2. Illustrates LatLRR's mechanism of decomposing data. Given a data matrix X, LatLRR decomposes it into a low-rank part XZ that represents the principal feature, a low-rank part LX that encodes the salient feature, and a sparse part E that fits the noise.

We interpret the feature extraction effect of LatLRR, mainly from the perspective of the singular value decomposition (SVD) decomposition. Suppose X and L have a skinny SVD: $X = U_X \sum_X V_X^T$ and $L = U_L W_L V_L^T$. The meaning of "skinny" is that $\sum_X$ and W_L are square matrices of size rank (X) and rank (L) respectively. In the presence of outliers, the LatLRR model is defined as problem (4). When the data is noisy, LatLRR uses the contaminated data as the dictionary (XZ) and also extracts features from noisy (LX). However, contaminated data as the dictionary is valid only when the percentage of corruption is relatively low and the noise level is also low. To overcome this limitation, Zhang et al. [34] proposed to denoise X first, and then apply the noiseless LRR or the LatLRR to the denoised data. This leads to the following model:

$$\min_{Z,L,E} \quad \|Z\|_* + \|L\|_* + \lambda\|E\|_1 \quad s.t. \quad X - E = (X - E)Z + L(X - E) \tag{5}$$

The complete solutions to problem (5) are as follows:

$$Z^* = V_A(I - S)V_A^T \tag{6}$$

$$L^* = U_A S U_A^T \tag{7}$$

Where $U_A \sum_A V_A^T$ is the skinny SVD of A (denoised X) and S is any block-diagonal matrix, $diag\{s_1, s_2, \ldots, s_r\}$ with $0 \le s_i \le 1$ for all i.

In fact, the rank r = rank (A) is unknown. Can we recover the row space of r efficiently and exactly? In particular we do not know rank r and the number of subspace are unknown either. Hence, to obtain optimal subspace is still of interest.

3.2. Dynamic rank

The upper bound of the estimated rank r is an important parameter and a suitable value of r is critical to the proposed approach. However, the determination of the reduced rank is also an open problem. Fortunately, several researchers [35-37] have provided some rank estimation strategies to compute a good value for the rank of X. Ma et al. [38] proposed the fixed point continuation with approximate (FPCA) that can solve very large matrix rank minimization problems. FPCA can recover them with a relative error of 10^{-5} in about 3 min by sampling only 20% of the matrix elements. Candès et al. [39] finds a perfect recovery as long as r obeys $rank \le \rho_r n \mu^{-1} (\log n)^{-2}$ and $m \le \rho_s n^2$, which for large values of r, matches the best result available. ρ_r, ρ_s are positive numerical constants and μ is the incoherence condition. m, n are the size of sample, so we can find that rank is mainly dependent on the samples. In real word applications the observations are often corrupted by noise, the paper [40] studies the problem of recovering a low-rank matrix from a high dimensional data matrix despite both small entry-wise noise and gross sparse errors. The experimental results show that rank r is also relative to dimension of the samples.

3.3. The basic model

The rank r approximation of X is given by:

$$\begin{aligned} &\min \ \|Z\|_* + \lambda\|X - XZ\|_F^2 \\ &s.t. \ X = XZ + LX \\ &Z = V_X W_Z V_X^T, \ L = U_X(I - W_Z)U_X^T \\ &rank(Z) = r \end{aligned} \tag{8}$$

where $\|\cdot\|$ is any unitary invariant norm. The optimal approximant is given by a truncated singular value decomposition of X. If $X = U \sum_r V^T$, where the first r diagonal entries of $\sum_r$ are largest r singular values, the approximation error $\|X - XZ\|_F^2$ is small enough, then the amount of data is $r(m + n - r)$, which can be much smaller than the mn required to transmit the value of all the entries. XZ is only allowed to contain the most principal features, which means that the columns of XZ must be similar to each other. Thus $\|Z\|_*$ cannot be large, since the nuclear norm reflects low-rank self-expressiveness, which is a very good similarity measure for multiple samples.

Indeed, when (Z^*, L^*) is an optimal solution to Eq. (8), the column space of Z^* is a subspace of V_X. Hence, for Z^* to exactly recover V_Z. If there are outliers that exactly lie on the subspace, then X will contain more samples than Z. In order to exactly recover Z from X. We suppose input X contains two different parts: M_1 and M_2. M_1 is the main part of M, candidates as the low-rank part, while M_2 is the residual part. M_1 and M_2 have the following close form:

$$\begin{aligned} &\min \ \|X - M_1\|_2^2 + \lambda\|M_2\|_1 \\ &s.t. \quad M_1 = V_r V_r^T \quad M_2 = V_{n-r} V_{n-r}^T \end{aligned} \tag{9}$$

where V_r and V_{n-r} are calculated by computing the skinny SVD of

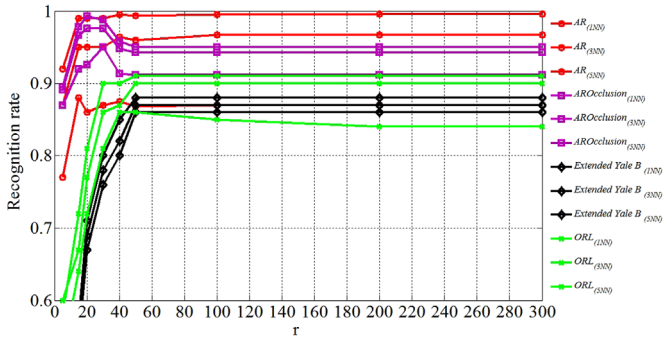

Fig. 3. Relation between rank and recognition rate.

M_1 and M_2. n is positive number of singular values, r is the number of first few singular values. So M_1 can be defined as subspace from different classes.

3.4. Basic model analysis

We first select rank r randomly, and carry out experiments on publicly databases. The total dimension of training image is $20 \times 15 = 300$. We change the dimension of r between [0 300] and compare the recognition rate among different databases. Since the total dimension of training image is $20 \times 15 = 300$, we do not consider higher dimensional space for evaluation. In Fig. 3, the recognition rate is increasing as the rank increases, and then the rate of recognition tends to be steady. Via the experiments carried on different databases, we find that rank r has the similar characteristic in different databases. An optimal r can achieve the best performance. As a result, selecting a good value of rank r is important and is beneficial to recognition, it is also an interesting and challenging task. From Fig. 3 we know that the rank r mainly depends on the size of training image.

For more intuitive observation, we analyze the relation of rank and image size in Fig. 4. The x-axis represents ratio of r to image size ($S_1 \times S_2$) and y-axis represents the recognition rate. We can find that recognition rate may tend to be steady, and the ratio of r to image size ($S_1 \times S_2$) in different databases almost has the same optimal value. So the r can be defined as follows:

$$r = \rho \times S_1 \times S_2 \tag{10}$$

Where ρ is ratio of r to image size, ρ is a parameter in the interval [0 1]. $S_1 \times S_2$ is the image size.

Fig. 4. illustrates the relation between ρ and face recognition rate in different database. We can find that when the value of ρ fall into the interval [0.16–0.2], the recognition rate approximates best performance. To simplify the experiments, we choose $\rho = 0.2$ as the candidate ratio of r to image size. This observation leads us to obtain the dynamic value of r and construct optimal subspace.

3.5. Feature extraction

Unsupervised feature extraction is a fundamental step in pattern recognition. In the above, we choose M_1 as the optimal subspace from different classes. So our model is defined as the following optimization problem:

$$\begin{aligned} &\min_{Z^*,L^*,E^*} \|Z\|_* + \lambda\|M_1 - M_1Z\|_F^2 \\ &s.\,t.\ M_1 = M_1Z + LM_1 + E \\ &Z = V_{M_1}W_Z V_{M_1}^T,\ L = U_{M_1}(I - W_Z)U_{M_1}^T \\ &rank(Z) = r \end{aligned} \tag{11}$$

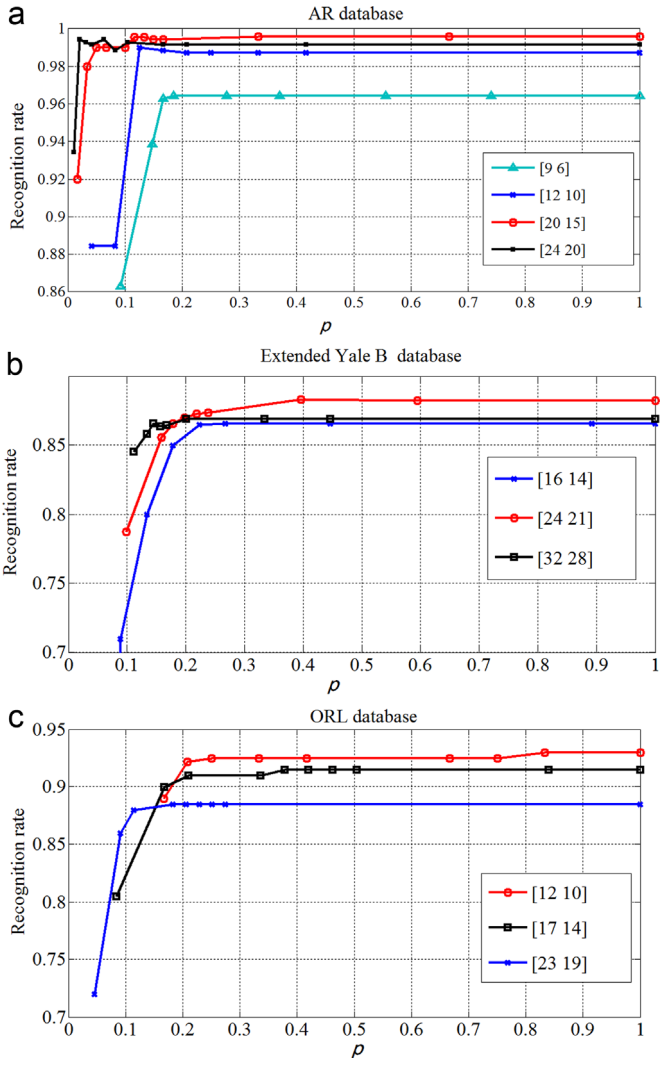

Fig. 4. Relations between rank and image size.

where M_1Z is a low-rank part which represents the principal feature. LM_1 is a low-rank part that encodes the salient features and E is a sparse part that represents noise. The regularization parameter λ depends on quantities which in practice are not known beforehand. In many real datasets, low-rank assumption is realistic [41].

Fig. 5 shows some examples of face images taken from Extended Yale B database. We can find that the SVD components (Fig. 5(b)) among different images look very similar to the original one (Fig. 5 (a)), Fig. 5(d) is the salient feature and Fig. 5(e) is a sparse part. Compared with LRR (Fig. 2(c)) and LatLRR (Fig. 3(c)), decomposed by our proposed method (Fig. 5(d)) is more accurate in salient feature extraction. It is obvious that Fig. 5(d) includes a lot of detailed information while the sparse part (Fig. 5(e)) is more sparsely.

In Fig. 5, LM_1 may be useful for feature extraction (Fig. 5(d)). We experimentally find that LM_1 is able to extract discrimination features from the original image. After learning LM_1 from a set of training image, it is straightforward to generalize the learnt model to testing image. Our proposed method finds the most discriminative feature and reduce feature dimension.

3.6. Classification

In this stage, any type of classifier can be used, such as the

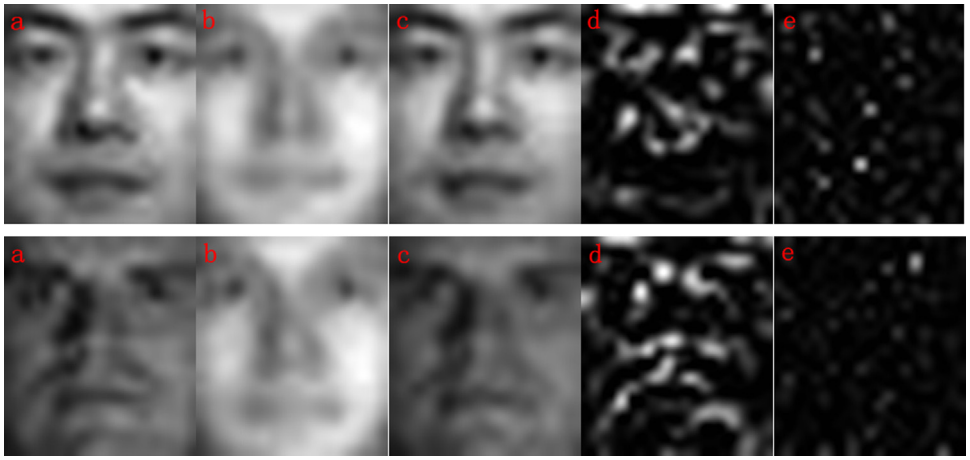

Fig. 5. Illustrates the data decomposing mechanism of our proposed method. Given a data matrix X and extract M_1, decompose it into a low-rank part M_1Z that represents the principal feature, a low-rank part LM_1 that encodes the salient feature and a sparse part E that fits noise.

posterior probabilities in Bayes classification [42], or the decision value in the SVM [43]. In this paper, we propose to use the K-NN algorithm [44] because it is simple. In pattern recognition, the k-nearest neighbor algorithm is a widely used classifier for classifying objects based on closest training examples in the feature space. The k-nearest neighbor algorithm is the simplest classifier of all machine learning algorithms. Via this classifier, image is classified by a majority vote of its neighbors. In K-NN classifier the Euclidean distance between the testing image feature and each training image feature is determined to form a distance matrix. The summation value of distance matrix is estimated and sorted in an increasing order. The first K elements are selected and majority class value is determined for classifying the image accurately.

K-NN is a lazy learning method for classifying objects based on the closest training examples in the feature space. A case is classified by a majority vote of its neighbors, with the case being assigned to the class most common amongst its K nearest neighbors measured by a distance function. If $K = 1$, then the case is simply assigned to the class of its nearest neighbor. Choosing the optimal value for K is best done by first inspecting the data. In general, a large K value is more precise as it reduces the overall noise but there is no guarantee. Cross-validation is another way to retrospectively determine a good K value by using an independent dataset to validate the K value [45].

3.7. The proposed algorithm

The main steps of the proposed method are described below:

Algorithm 1. Proposed method

Input: The training data X, the testing data Y. The size of training and testing data is $S_1 \times S_2$. The parameters λ, ρ. A matrix of normalized training samples $D = [D_1, D_2, \ldots, D_k] \in R^{m \times n}$ for K classes, a test sample $y \in R^m$.

1 Dynamic rank estimation

2 for each subject i **do**

3 Conduct Singular Value Decomposition on training data and obtain the clean data M_{i1}

4 Solve (11) to obtain L_i^*

5 Calculate the discriminative information of face image by $L_i^* M_{i1}$

6 Calculate the residual associated with i-th class by:

$$e_i = \|L_i^* y - L_i^* M_{i1}\|_2^2$$

7 end for

Output: $identify(y) = \arg \min_i \{e_i\}$

4. Experimental results

In this section, we conduct experiments on publicly available databases for frontal-face recognition, which serve to illustrate the efficacy of the proposed approach. We randomly split the database into training set and testing set first, and then resize the image and normalize the pixel values to [0, 1]. The experiments are implemented by MATLAB R2013b on a computer with Intel(R) Xeon (R) CPU X3450@2.67 GHz 2.66 GHz, and windows 7 operating system. We repeat each experiment for 10 times and the accuracy is averaged. We will examine the performance of the proposed method on dealing with varying occultation, and evaluate our algorithm on three databases (AR, Extended Yale B and ORL). Our approach is compared with several state-of-the-art methods, such as SRC [26], FDDL [27], LRR [12], Latent-LRR [14], and Deep Learning [28]. According to the experimental results, the proposed algorithm achieves a state-of-the-art performance.

4.1. AR face database

AR database consists of over 4000 frontal images from 126 individuals. For each individual, 26 pictures were taken in two separate sessions. A subset that contains 50 males and 50 females was chosen from AR database. Fig. 6 presents some face images taken from the AR database. We adopt K-NN ($K = 1, 3, 5$) classifier to classify various features, the original face image is down-sampled to 9×6, 12×10, 20×15, 24×20, 28×20, 42×30, according to the dimensions of 54, 120, 300, 480, 560 and 1260 respectively. Table 1 shows the recognition rate of each method at different dimensions, we can see that the proposed method achieves a stable recognition rate for each dimension. For example, when the dimension is equal to 54, our method achieves a high recognition rate at 96%, and those for Latent-LRR, LR, SRC, FDDL,

Fig. 6. Samples in different illuminations and expressions from the AR database.

Table 1
Recognition rates on the AR database.

Method	Dimension	[6,9]	[10,12]	[15,20]	[20,24]	[20,28]	[30,42]
Proposed method	acc(1 nn)	0.96	0.99	0.99	0.99	0.98	0.97
	acc(3 nn)	0.92	0.96	0.97	0.97	0.95	0.95
	acc(5 nn)	0.83	0.85	0.90	0.90	0.88	0.89
	Time(S)	2.79	6.81	25.02	56.00	80.56	150.00
LRR	acc(1 nn)	0.57	0.62	0.68	0.69	0.69	0.71
	acc(3 nn)	0.56	0.60	0.65	0.66	0.67	0.69
	acc(5 nn)	0.53	0.59	0.64	0.65	0.66	0.67
	Time(S)	0.59	0.68	0.89	1.34	1.39	3.46
Latent-LRR	acc(1 nn)	0.86	0.88	0.87	0.85	0.84	0.83
	acc(3 nn)	0.85	0.85	0.85	0.83	0.83	0.81
	acc(5 nn)	0.81	0.82	0.80	0.80	0.80	0.80
	Time(S)	5.91	8.67	29.89	58.80	81.67	187.03
FDDL	acc	0.41	0.74	0.81	0.86	0.89	0.91
	Time(S)	1836	2256	3348	4762	7775	11500
SRC	acc	0.63	0.84	0.92	0.94	0.94	0.95
	Time(S)	5.49	9.08	22.89	40.67	50.88	208.00
Deep learning	acc	0.35	0.54	0.66	0.72	0.73	0.81
	Time(S)	104.06	104.91	142.76	152.81	154.77	155.24

Deep Learning are merely 86%, 57%, 63%, 41%, and 35% respectively. It can be seen that when the dimension is equal to 300 (20×15), the performance is good and stable. So due to the limited space, we will choose this dimension in the following experiments on AR database.

4.1.1. Robustness to noise

This part of the experiment aims to test the robustness of our method in the presence of noise. We randomly add some noise to the images from the database. Fig. 7 shows some test images under different levels of noise. We choose the dimension 300 as the candidate, and conduct experiments to evaluate the performance of the algorithm dealing with different density of noise in AR database. Table 2 shows the recognition rate under different noise levels of each method. It can be seen that both LRR, Latent-LRR are stable with noise but our proposed method generates a higher recognition rate and more stable than others.

4.1.2. Robustness to occlusion

In this experiment, test images which are occluded by randomly selected blocks are used to test the robustness of our algorithm against occlusion. We use the cropped and normalized 20×15 face images. Fig. 8 shows an example of images with different types of block occlusion and different levels of occlusion from the AR database. The location of occlusion is randomly chosen for images and is unknown to the algorithms.

Compared to the noise corruption, contiguous occlusion is obviously worse for the face recognition algorithm. Fig. 9 shows the recognition rates with different types and levels of block occlusions. In general, the recognition rate reduces by the increasing block size, recognition rates of LRR and Latent-LRR fall sharply when the block size increases. The recognition rate of the proposed method is the highest. Those of SRC and FDDL are approximate to the recognition rate of ours. It is obvious that our method significantly outperforms others in Fig. 9. Up to half a face occluded, the proposed method performs by correctly identifying over 90% of the test subjects.

4.1.3. Recognition with real face disguise

This section shows the performance of the algorithm dealing with real face disguise. We carry out experiments on the AR database, occluded by sunglasses and scarves. In this experiment, the first test set contains images of the subjects wearing sunglasses. The second test set contains images of the subjects wearing a scarf. All images are converted into gray-scale, cropped and aligned by the centers of eyes and mouth, and then normalized with different resolution, as shown in Fig. 10.

4.1.3.1. Sunglasses. We consider corrupted training images due to the occlusion of sunglasses. We only choose one image with sunglasses (randomly chosen) for training, and the remaining images with sunglasses for testing. In other words, we have a total of one training image and three test images per person.

4.1.3.2. Scarf. We consider the training images corrupted by scarf. We apply a similar training/test set choice, and have a total of one training image (randomly selected image with scarf) and the remainder as test images.

4.1.3.3. Sunglasses+Scarf. Finally, we consider the case where images with sunglasses and scarf are presented during training. We choose two corrupted images (one with sunglasses and the other with scarf) for training. A total of two test images (one with sunglasses and the other with scarf) are available for this case.

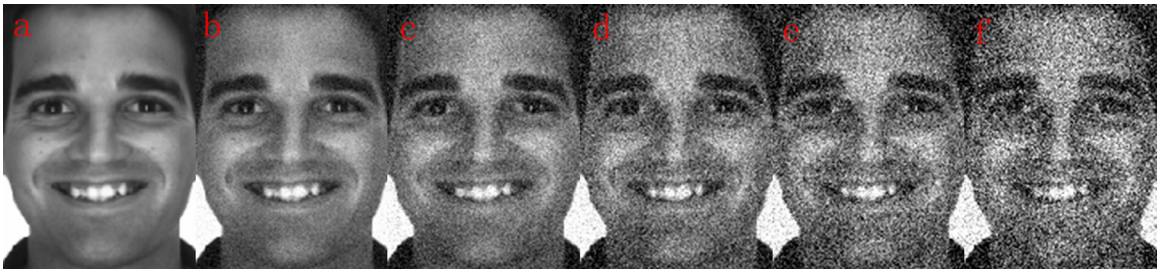

Fig. 7. Images from the AR database with noise corruption.

Table 2
Recognition rates with different levels of noise.

σ	SRC	FDDL	LRR			Latent-LRR			Proposed method		
			1 nn	3 nn	5 nn	1 nn	3 nn	5 nn	1 nn	3 nn	5 nn
10	0.51	0.80	0.69	0.68	0.68	0.74	0.72	0.70	0.99	0.96	0.90
20	0.30	0.60	0.67	0.66	0.66	0.73	0.71	0.68	0.99	0.95	0.88
30	0.18	0.44	0.68	0.68	0.68	0.72	0.70	0.67	0.99	0.96	0.89
40	0.10	0.34	0.64	0.64	0.64	0.71	0.69	0.65	0.98	0.95	0.85
50	0.08	0.26	0.63	0.63	0.63	0.69	0.66	0.63	0.91	0.90	0.83

Fig. 11 summarizes the performance compared with different approaches under three different scenarios and shows that our approach significantly outperforms some state-of-the-art methods. The proposed method achieves a high recognition rate and stable to different dimensions. For example, when the data dimension is equal to 300, we achieve recognition rates at 98%, 70% and 92% for the scenarios of Sunglasses, Scarf, and Sunglasses+Scarf respectively. By using Latent-LRR, which is the most relevant method to ours, it obtains the rates of 65%, 25% and 45% respectively. In other words, we outperform the method of Latent-LRR by about 32–47%. FDDL and SRC, which are close to ours, it obtains 85%, 58% and 70% respectively. Furthermore, our proposed method is less complex, Fig. 12 shows the time complexity of different algorithms, LR, SRC and our method are similar and much faster than FDDL.

4.2. Extended yale B face database

The Extended Yale B database consists of 2414 frontal-face images of 38 individuals. The images were captured under various laboratory-controlled illumination conditions. For each class, there are about 64 images, all test images used in the experiments are manually aligned, cropped, and then resized to 64×56. The original face image is down-sampled to 8×7,16×14,24×21,32×28, according to the dimensions of 56, 224, 504 and 896 respectively. Fig. 13 shows some face images from the Extended Yale B database.

We first randomly select images for training and the remaining for test. We vary the dimension of 56, 224, 504, and 896 to compare the recognition rate between different methods. Since the size of training image is 1216, we do not consider higher dimensional space for evaluation. We compare the proposed method with SRC, FDDL, LRR, Latent-LRR, Deep Learning and Table 3 shows the experimental results.

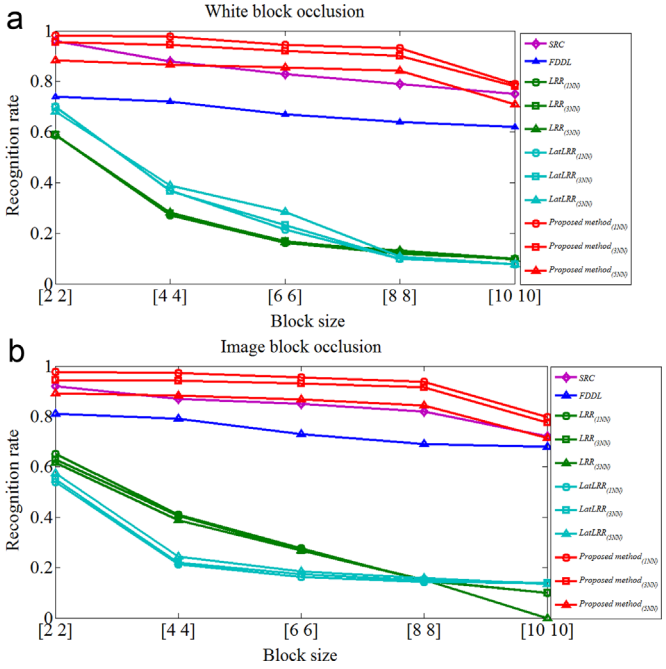

Fig. 9. Recognition rates with different types and levels of block occlusions.

From Table 3, the proposed method consistently produced higher recognition rates than those of LRR and Latent-LRR, especially in low dimension. Although, SRC, FDDL, and Deep Learning obtain higher recognition rates in some dimensions, our proposed method is more robust, lower in computational complexity and it

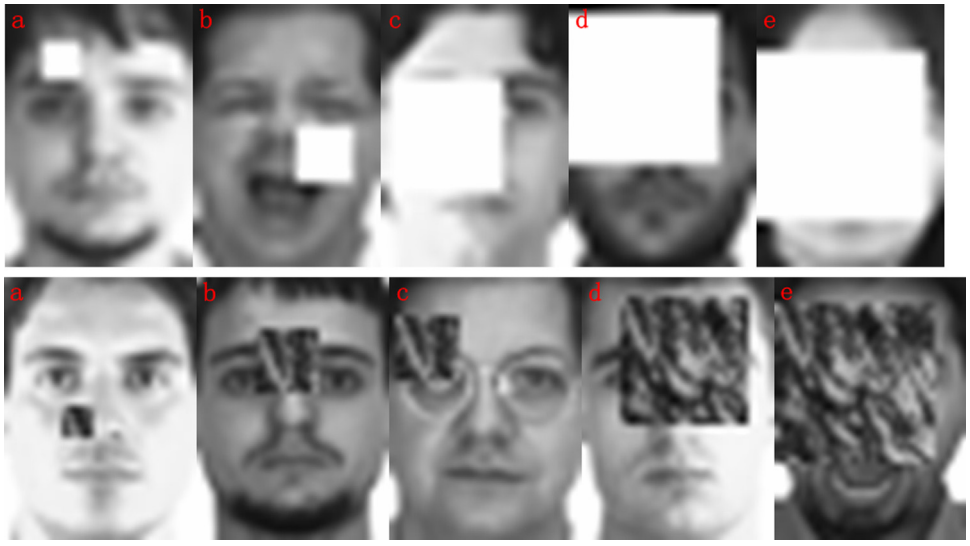

Fig. 8. An example of different random block occlusions: white block occlusions (top) and Image block occlusions (bottom).

Fig. 10. Real face disguise from the AR database.

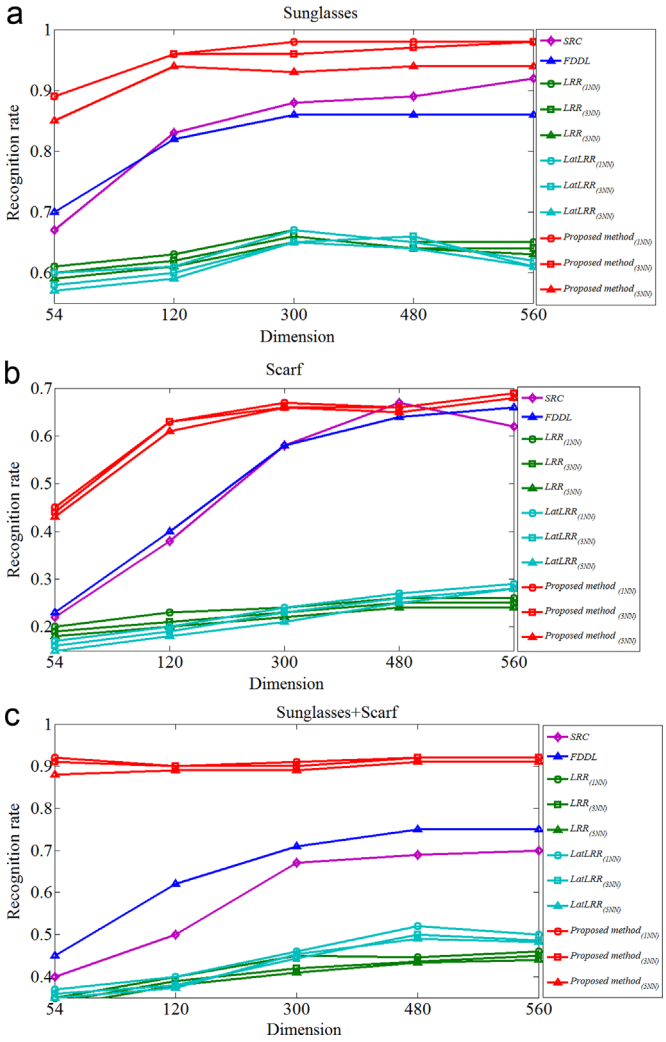

Fig. 11. Recognition rates on AR database with disguise occlusion.

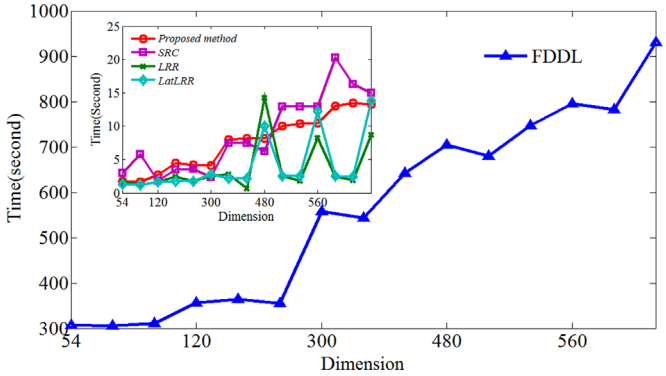

Fig. 12. Time complexity of algorithms on AR database with disguise occlusion.

has better performances overall.

4.2.1. Robustness to noise

We choose the dimension 224 (16×14) as a candidate, and carry out experiments with different densities of noise on the Extended Yale B Face Database. We compare the performance of the proposed method with LRR, Latent-LRR, SRC, and FDDL. Table 4 shows the recognition rate under different noise levels of each method. From Table 4 we can see that our proposed method shows excellent performance than others. LRR and Latent-LRR are robust to noise, but the recognition rate is low. The recognition rates of SRC and FDDL dramatically decrease while the noise variance increases. Our method achieves a higher recognition rate and more stable.

4.2.2. Robustness to Occlusion

In order to demonstrate the robustness of our algorithm to occlusion, we choose the dimension 224 (16×14) as a candidate. Test images which are occluded by randomly chosen block are used in the experiments. Fig. 14 shows the recognition rate with different types and levels of block occlusions. We can see that in general the recognition rates drop as the block size increased. The recognition rates of LRR and Latent-LRR fall sharply when the block size increases, the results of SRC and FDDL are approximately the same as the proposed method. It is obvious that our proposed method significantly outperforms others and is stable under different types of block corruption. Up to 65% of image area is occluded by randomly chosen block, the proposed method performs by correctly identifying over 75% of the test subjects.

4.3. ORL face database

The ORL database contains 400 images in total, ten different images for each of the 40 different subjects. The background of the images is uniform and dark while the subjects are in frontal, upright posture. The images were shot under different lighting conditions and with various facial expressions and details. The original face image is down-sampled to 6×5, 12×10, 17×14, 23×19, according to the dimensions of 30, 120, 238 and 437 respectively. Fig. 15 shows some face images from the ORL database.

We carry out some experiments on the ORL database and the results are illustrated in Table 5. From this table, we find that LRR method conducted on the ORL database can achieve nearly 100% recognition rate in some dimensions. However, in other dimensions it does not perform well. Latent-LRR method consistently produces higher recognition rates than FDDL, SRC and Deep Learning approaches, our proposed method is more stable than other approaches and it can also achieve high recognition rates. For example, when the dimension is 30, our method achieves a high recognition rate at 95%, and those of Latent-LRR, LR, SRC, FDDL, Deep Learning are 93%, 57%, 75%, 89%, and 14% respectively.

4.3.1. Robustness to Noise

We choose the dimension 120 (12×10) as a candidate, and carry out experiments with different densities of noise on the ORL database to evaluate the robustness of the algorithm. We compare

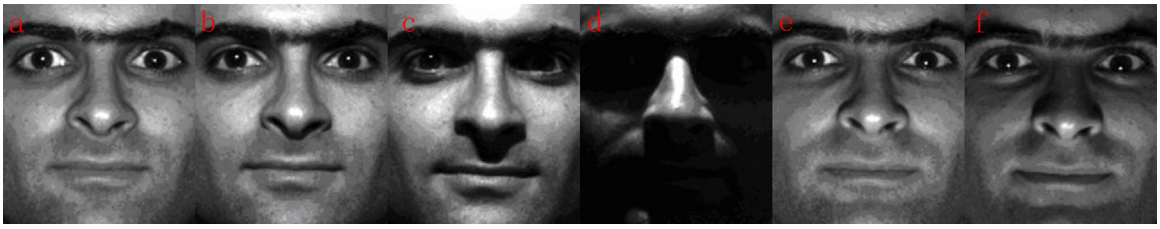

Fig. 13. Samples of different illuminations from the Extended Yale B database.

Table 3
Recognition rates on Extended Yale B Face Database.

Method	Dimension	[7,8]	[14,16]	[21,24]	[28,32]
Proposed method	acc(1 nn)	0.76	0.87	0.89	0.90
	acc(3 nn)	0.75	0.87	0.88	0.87
	acc(5 nn)	0.75	0.86	0.86	0.85
	Time(S)	2.09	13.67	50.04	150.00
LRR	acc(1 nn)	0.35	0.51	0.57	0.61
	acc(3 nn)	0.34	0.50	0.56	0.60
	acc(5 nn)	0.33	0.48	0.54	0.58
	Time(S)	0.29	0.48	3.37	4.08
Latent-LRR	acc(1 nn)	0.66	0.85	0.87	0.88
	acc(3 nn)	0.67	0.84	0.86	0.87
	acc(5 nn)	0.68	0.83	0.85	0.85
	Time(S)	20.98	80.65	100.09	200.01
FDDL	acc	0.40	0.90	0.97	0.97
	Time(S)	25530	27700	31207	31790
SRC	acc	0.43	0.78	0.83	0.87
	Time(S)	3.89	18.93	57. 49	200.07
Deep learning	acc	0.79	0.86	0.88	0.92
	Time(S)	100.16	124.58	155.60	233.74

the results of our proposed method with those of LRR, Latent-LRR, SRC and FDDL. Fig. 16 shows the recognition rates under different noise levels for each method. The proposed method shows a much better performance than others.

4.3.2. Robustness to occlusion

In this experiment, test images which are occluded by randomly selected blocks are used to test the robustness of our algorithm to occlusion. Fig. 17 shows the recognition rates with different types and levels of block occlusions. As shown in Fig. 17, the recognition rates of LRR and Latent-LRR fall sharply with increasing block size, the recognition rates of SRC and FDDL are approximate to the recognition rates of proposed method. In general, our method significantly outperforms others and is the most stable among all methods in different types of block corruption due to its robustness in handling errors. When more than half of the face is occluded, the proposed method can still achieve a recognition rate of 80%.

The complexity of proposed algorithm is similar to SRC, and much faster than FDDL. As a rough comparison, the execution times for face recognition show in Table 1, Table 3 and Table 5 in different databases. We choose the dimension 300 in AR database, 224 in Extended Yale B Face Database, and 120 in ORL database. The execution times for face recognition is 0.4–3.2 s for LRR, for

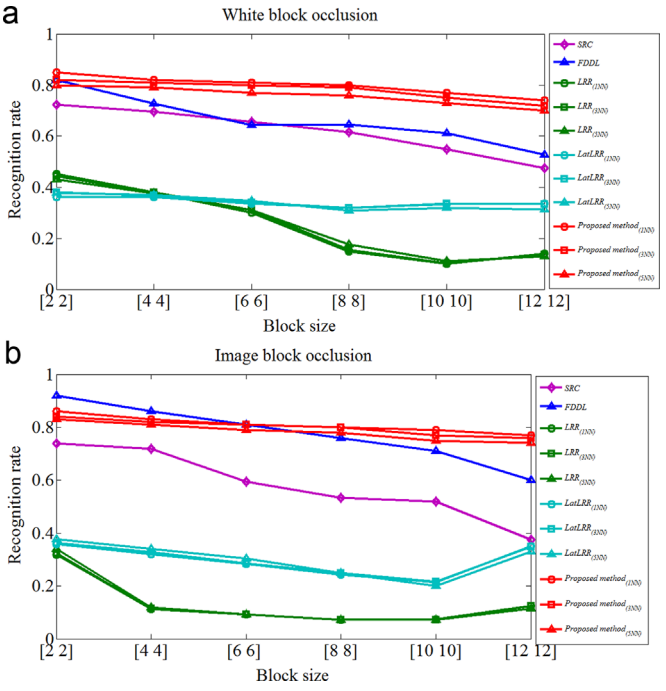

Fig. 14. Recognition rates with different types and levels of block occlusions.

3.7–80 s for Latent-LRR, 40–234 s for deep learning, and 0.7–22 s for SRC. The execution time of proposed algorithm is 1.7–26 s. From the experimental results, we can find that our method could get a more robust performance with faster speed. In most cases, our method gets the top performance.

5. Conclusions

In this paper, motivated by recent success of LRR, we present a novel method to recognize face image in the presence of illumination variations, noise corruption and block occlusion. In particular, an exact low subspace of image and discriminative parts are defined to characterize the features of face image of different individuals. We use discriminative parts to represent the characteristic of face image, and give a recognition protocol for classification. Our extensive experimental results on different face

Table 4
Recognition rates with different levels of noise.

σ	SRC	FDDL	LRR			Latent-LRR			Proposed method		
			1 nn	3 nn	5 nn	1 nn	3 nn	5 nn	1 nn	3 nn	5 nn
10	0.64	0.87	0.57	0.55	0.53	0.83	0.80	0.78	0.85	0.83	0.81
20	0.41	0.76	0.56	0.54	0.52	0.82	0.80	0.78	0.84	0.82	0.80
30	0.27	0.61	0.55	0.54	0.52	0.79	0.78	0.77	0.81	0.80	0.78
40	0.18	0.48	0.54	0.53	0.51	0.77	0.75	0.73	0.79	0.77	0.76
50	0.13	0.35	0.53	0.53	0.51	0.70	0.70	0.69	0.75	0.73	0.71

Fig. 15. Samples in different illuminations from the ORL database.

Table 5
Recognition rates on the ORL database.

Method	Dimension	[5,6]	[10,12]	[14,17]	[19,23]
Proposed method	acc(1 nn)	0.95	0.99	0.95	0.93
	acc(3 nn)	0.93	0.91	0.91	0.90
	acc(5 nn)	0.88	0.87	0.86	0.82
	Time(S)	0.50	1.70	3.90	5.40
LRR	acc(1 nn)	0.57	0.99	0.99	0.99
	acc(3 nn)	0.54	0.89	0.90	0.91
	acc(5 nn)	0.42	0.80	0.79	0.85
	Time(S)	2.10	3.20	5.30	8.20
Latent-LRR	acc(1 nn)	0.93	0.94	0.95	0.93
	acc(3 nn)	0.90	0.90	0.90	0.90
	acc(5 nn)	0.86	0.88	0.88	0.85
	Time(S)	2.10	3.74	5.63	9.32
FDDL	acc	0.89	0.94	0.95	0.95
	Time(S)	45.01	57.06	70.47	120.43
SRC	acc	0.745	0.92	0.90	0.90
	Time(S)	0.32	0.65	1.85	5.61
Deep Learning	acc	0.14	0.72	0.81	0.87
	Time(S)	44.30	57.24	75.42	60.29

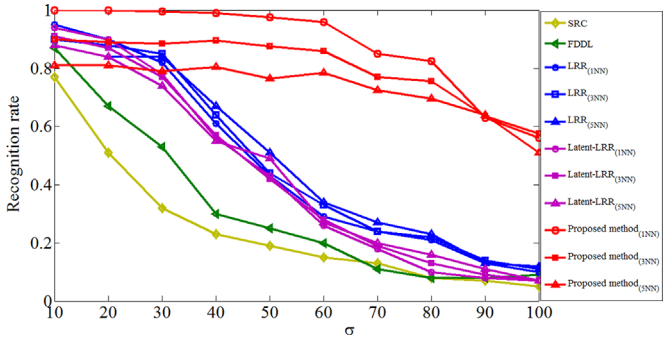

Fig. 16. Recognition rates on ORL database with noise.

databases show that the proposed method is competitive with some state-of-the-art methods in terms of high recognition rate, low computational complexity and robustness to complicated occasions such as illumination variations and corruptions, occlusion and disguise. It has been proved to be suitable for low dimension face recognition applications.

Acknowledgments

The authors would like to thank the anonymous reviewers and the editors for their valuable suggestions. This work was supported by the National Natural Science Foundation of China (No. 61171077), the University Science Research Project of Jiangsu Province (No.12KJB510025), and the National Sciences and Engineering Research Council of Canada.

Appendix

Note: Let $\|M\|_*:=\sum_i \sigma_i(M)$ denote the nuclear norm of the matrix M, i.e. the sum of the singular values of M, and let

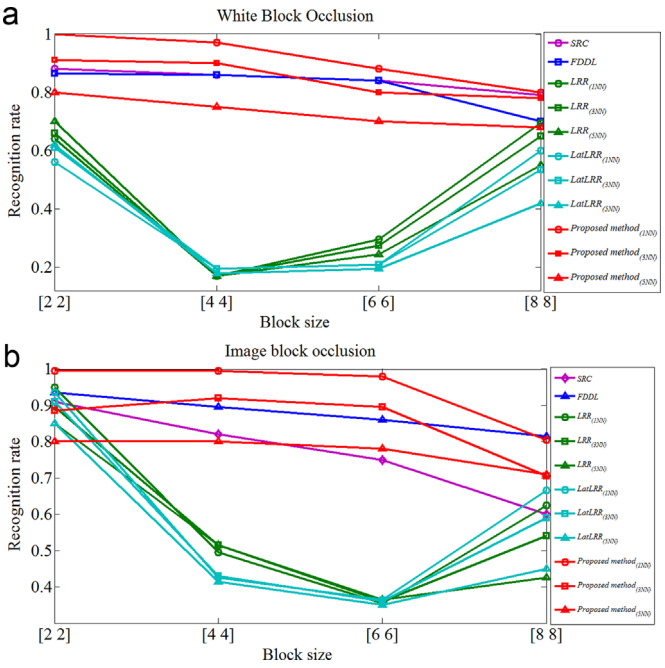

Fig. 17. Recognition rates with different types and levels of block occlusions.

$\|M\|_1:=\sum_{ij}(M_{ij})$ denote the l_1-norm of M seen as a long vector in $R^{n1\times n2}$.

Independent subspace: A collection of k subspaces $\{S_1, S_2, S_3, \ldots, S_k\}$ are independent if and only if $S_i \cap \sum_{j\neq i} S_j = \{0\}$. When the subspaces are of low-rank and the ambient dimension is high, the independent assumption is roughly equal to the pair wise disjoint assumption; that is $S_i \cap S_j = \{0\}, \forall\, i \neq j$.

Full SVD and Skinny SVD: for an $m \times n$ matrix M; its Singular Value Decomposition is defined by $M = U[\sum, 0]V^T$,where U and V are orthogonal matrices and $\sum = diag(\sigma_1, \sigma_2, \ldots, \sigma_m)$ with $\{\sigma_i\}_{i=1}^m$ being singular values. The SVD defined in this way is also called the full SVD. If we only keep the positive singular values, the reduced form is called skinny SVD. For a matrix M of rank$r$, its skinny SVD is computed by $M = U_r \sum_r V_r^T$,where $\sum_r = diag(\sigma_1, \sigma_2, \ldots, \sigma_r)$ with $\{\sigma_i\}_{i=1}^r$ being positive singular values. More precisely, U_r and V_r are formed by taking the first r columns of U and V,respectively.

References

[1] Liu Yuanyuan, L.C. Jiao, Fanhua Shang, An efficient matrix factorization based low-rank representation for subspace clustering, Pattern Recognit. 46 (1) (2013) 284–292.
[2] Yue Deng, Qionghai Dai, Risheng Liu, Zengke Zhang, Sanqing Hu, Low-rank structure learning via nonconvex heuristic recovery, IEEE Trans. Neural Netw. Learn. Syst. 24 (3) (2013) 383–396.
[3] Y. Mu, J. Dong, X. Yuan, S. Yan, Accelerated low-rank visual recovery by random projection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 2609–-2616.
[4] X. Wang, C. Gan, Y. Wang, Human motion segmentation based on low-rank representation, in: Proceedings of the IEEE Conference on Audio, Language and Image Processing, 2012, pp. 936–941.
[5] Guangcan Liu, Huan Xu, Shuicheng Yan, Exact subspace segmentation and

outlier detection by low-rank representation, J. Mach. Learn. Res. Proc. Track 22 (2012) 703–711.
[6] G. Liu, Z. Lin, Y. Yu, Robust subspace segmentation by low-rank representation, in: Proceedings of the 27th International Conference on Machine Learning, 2010, pp. 663–670.
[7] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, Y. Ma, Robust recovery of subspace structures by low-rank representation, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 171–184.
[8] Guangcan Liu, Shuicheng Yan, Latent low-rank representation for subspace segmentation and feature extraction, in: Proceedings of the IEEE International Conference on Computer Vision, 2011, pp.1615–1622.
[9] R.S. Liu, Z.C. Lin, F.D. Torre, Z.X. Su, Fixed-rank Representation for Unsupervised Visual Learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 598–605.
[10] L.S. Zhuang, H.Y. Gao, Z.C. Lin, Y. Ma, X. Zhang, N.H. Yu, Non-negative Low Rank and Sparse Graph for Semi-supervised Learning, In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2328-2335.
[11] I.H. Jhuo, D. Liu, D.T. Lee, S.F. Chang, Robust visual domain adaptation with low-rank reconstruction, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2168–2175.
[12] M. Shao, C. Castillo, Z. Gu, Y. Fu, Low-rank transfer subspace learning, in: Proceedings of the IEEE 12th International Conference on Data Mining, 2012, pp. 1104–1109.
[13] Yaniv Taigman, Ming Yang, Marc'Aurelio Rantzazo, Lior Wolf, Deepface: closing the gap to human-level performance in face verification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp.1701–1708.
[14] B. Chen, G. Polatkan, G. Sapiro, D. Blei, D. Dunson, L. Carin, Deep learning with hierarchical convolutional factor analysis, IEEE Trans. Pattern Anal. Mach. Intell. 35 (8) (2013) 1887–1901.
[15] L. Li, X. Zhang, M. Zhou, L. Carin, Nested dictionary learning for hierarchical organization of imagery and text, in: Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI2012), Catalina Island, CA, Aug. 2012.
[16] Quoc V. Le, Marc'Aurelio Ranzato, Rajat Monga, Matthieu Devin, Kai Chen, Greg S. Corrado, Jeff Dean, Andrew Y. Ng, Building high-level features using large scale unsupervised learning, in: Proceedings of the 29th International Conference on Machine Learning, ICML 2012, 1, 81–88.
[17] L. Ma, C. Wang, B. Xiao, W. Zhou. Sparse representation for face recognition based on discriminative low-rank dictionary learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2586–2593.
[18] Yin Ming, Shuting Cai, Junbin Gao. Robust face recognition via double low-rank matrix recovery for feature extraction, in: Proceedings of the IEEE Conference on image processing, 2013, pp. 3770–3774.
[19] L. Li, S. Li, Y. Fu. Discriminative dictionary learning with low-rank regularization for face recognition, in: Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, 2013, pp.1–6.
[20] Nan Zhang, Jian Yang, Low-rank representation based discriminative projection for robust feature extraction, Neurocomputing 111 (2) (2013) 13–20.
[21] C.F. Chen, C.P. Wei, Y.F. Wang, Low-rank matrix recovery with structural incoherence for robust face recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2618–2625.
[22] D.T Larose. k-nearest neighbor algorithm. Discovering knowledge in data: an introduction to data mining, 2005, pp. 90–106.
[23] F. Samaria, A. Harter. Parameterisation of a stochastic model for human face identification, in: Proceedings of the Second IEEE Workshop on Applications of Computer Vision, 1994, pp. 138–142.
[24] A.M. Martinez, R. benavente. The AR face database, Technical Report CVC 24, 1998.
[25] K.-C. Lee, J. Ho, D. Kriegman, Acquiring linear subspaces for face recognition under variable lighting, IEEE Trans. Pattern Anal. Mach. Intell. 27 (5) (2005) 684–698.
[26] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 210–227.
[27] M. Yang, L. Zhang, X. Feng, D. Zhang, Fisher discrimination dictionary learning for sparse representation, in: Proceedings of the IEEE International Conference on Computer Vision, 2011, pp. 543–550.
[28] A. Vedaldi, K. Lenc, MatConvNet - Convolutional Neural Networks for MATLAB, arXiv:1412.4564, 2014.
[29] E.J. Candès, X. Li, Y. Ma, et al., Robust principal component analysis? J. Am. Stat. Assoc. 8 (1) (2009) 1–73.
[30] S. Wang, X. Yuan, T. Yao, S. Yan, J. Shen. Efficient subspace segmentation via quadratic programming, in: Proceedings of the 25th AAAI Conference on Artificial Intelligence, 2011, pp. 519–524.
[31] P. Favaro, R. Vidal, A. Ravichandran. A closed form solution to robust subspace estimation and clustering, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 1801–1807.
[32] Y. Ni, J. Sun, X. Yuan, S. Yan, L. Cheong, Robust low-rank subspace segmentation with semidefinite guarantees, in: Proceedings of the IEEE International Conference on Data Mining, 2010, pp. 1179–1188.
[33] Z. Lin, M. Chen, Y. Ma, The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices, arXiv preprint arXiv:1009.5055, 2010.
[34] D. Bertsekas, Constrained Optimization and Lagrange Multiplier Method, Academic Press, New York, 1982.
[35] Hongyang Zhang, Zhouchen Lin, Chao Zhang, Junbin Gao, Relations among Some low rank subspace recovery models, Neural Comput. (2015).
[36] Z. Wen, W. Yin, Y. Zhang, Solving a low-rank factorization model for matrix completion by a non-linear successive over-relaxation algorithm, Math. Program. Comput. 4 (4) (2012) 333–361.
[37] Y. Shen, Z. Wen, Y. Zhang, Augmented Lagrangian alternating direction method for matrix separation based on low-rank factorization, Optim. Methods Softw. 29 (2) (2014) 239–263.
[38] Shiqian Ma, Donald Goldfarb, Lifen Chen, Fixed point and Bregman iterative methods for matrix rank minimization, Math. Program. 28 (1-2) (2011) 321–353.
[39] E.J. Candès, X. Li, Y. Ma, J. Wright, Robust Principal Component Analysis? Technical Report, Stanford, 2009.
[40] Z. Zhou, X. Li, J. Wright, E. Candes, Y. Ma, Stable principal component pursuit, in: Proceedings of the IEEE International Conference on Information Theory Proceedings, 2010, pp.1518–1522.
[41] V. Chandrasekaran, S. Sanghavi, P. Parrilo, A. Willsky, Rank-sparsity incoherence for matrix decomposition, SIAM J. Optimization 21 (2) (2011) 572–596.
[42] D.L. Poole, A.K. Mackworth, Artificial Intelligence: Foundations of Computational Agents, Cambridge University Press, England, 2010.
[43] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, Cambridge university press, England, 2000.
[44] N.S. Altman, An introduction to kernel and nearest-neighbor nonparametric regression, Am. Statistician 46 (3) (1992) 175–185.
[45] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification., 2nd ed., Wiley, New York, 2001.

Hongjun Li received Ph.D. degree from Nanjing University of Aeronautics and Astronautics, China. He works in School of Electronic Information Engineering, Nantong University, China. He is currently a Visit Scientist of Center for Pattern Recognition and Machine Intelligence, Concordia University, Canada. His research interests include image processing and pattern recognition. He has coauthored over 20 technical publications in these areas.

Ching Y. Suen received the M.Sc. degree in engineering from the University of Hong Kong and PhD degree from the University of British Columbia, Vancouver, BC, Canada. He is the Director of Center for Pattern Recognition and Machine Intelligence, Concordia University, Montreal, QC, Canada, and the Concordia Chair on Artificial Intelligence and Pattern Recognition. He has guided/hosted 80 visiting scientists and professors, and has supervised 82 doctoral and master's graduates. Currently, he is the editor-in-chief of the journal of Pattern Recognition and an advisory or associate editor of four journals. He has founded and organized numerous international conferences on pattern recognition, handwriting recognition, and document analysis. He has also founded the IAPR ICDAR Awards. He has served numerous professional societies as president, vice-president, governor, and director. He has given 180 invited talks at various industries and academic institutions around the world, and has been the principal investigator or consultant of 30 industrial projects. His publications include four conference proceedings, 12 books, and more than 480 papers, of which many have been widely cited. He is a fellow of the IAPR, and the Academy of Sciences of the Royal Society of Canada. He is a life fellow of the IEEE.