Low-Rank Matrix Recovery with Discriminant Regularization

Zhonglong Zheng, Haixin Zhang, Jiong Jia, Jianmin Zhao, Li Guo, Fangmei Fu, and Mudan Yu

Department of Computer Science, Zhejiang Normal University, Jinhua Zhejiang 321004, China

Abstract. Recently, image classification has been an active research topic due to the urgent need to retrieve and browse digital images via semantic keywords. Based on the success of low-rank matrix recovery which has been applied to statistical learning, computer vision and signal processing, this paper presents a novel low-rank matrix recovery algorithm with discriminant regularization. Standard low-rank matrix recovery algorithm decomposes the original dataset into a set of representative basis with a corresponding sparse error for modeling the raw data. Motivated by the Fisher criterion, the proposed method executes low-rank matrix recovery in a supervised manner, i.e., taking the with-class scatter and between-class scatter into account when the whole label information is available. The paper shows that the formulated model can be solved by the augmented Lagrange multipliers, and provide additional discriminating ability to the standard low-rank models for improved performance. The representative bases learned by the proposed method are encouraged to be structural coherence within the same class, and as independent as possible between classes. Numerical simulations on face recognition tasks demonstrate that the proposed algorithm is competitive with the state-of-the-art alternatives.

1 Introduction

With the ever-growing amount of digital image data in multimedia databases, there is a great requirement for algorithms that can provide effective semantic indexing. Categorizing digital images only using keywords is the quintessential, but not always executable example in image classification tasks. Face recognition (FR) is one typical image classification problem. Several aspects contribute to the difficulty of FR problem including the large variability in variance, illumination, pose, occlusion and even disguise of different subjects.

To design realistic FR systems, researchers usually focus on feature extraction of facial images and the generalization of classifiers. The testing sample from the same subjects will be used to evaluate the associated identification or verification performance. Although the testing sample might be corrupted, the training data sets are commonly assumed to be well taken in some desired conditions including reasonable illumination, pose, variations and without occlusion or disguise. When applying existing face recognition methods for practical scenarios, we will need to throw away the corrupted training images, and we might thus encounter small sample size and overfitting problems. Moreover, the disregard of corrupted training face images might give

J. Pei et al. (Eds.): PAKDD 2013, Part II, LNAI 7819, pp. 437–448, 2013.
© Springer-Verlag Berlin Heidelberg 2013

up some valuable information for recognition. Inspired by the sparse coding mechanism of human vision system [1][2], and with the rapid development of ℓ_1-norm minimization techniques in recent years, the sparse representation classification (SRC) ideas have been successfully used in various machine vision and pattern recognition applications [3][4][5][6]. Though interesting classification results have been reported in documentations, more investigations need to be made in order for a clearer understanding about the relationship between object representation and classification. Since SRC requires the training images to be well aligned for reconstruction purposes, [7] and [8] further extend it to deal with face misalignment and illumination variations. [5] also proposes modified SRC-based framework to handle outliers such as occlusions in face images. However, the above methods might not generalize well if both training and testing images are corrupted.

To address this issue, we propose formulating the face recognition problem under a matrix completion framework fueled by the recent advances in low-rank (LR) matrix recovery [9][10][11], together with the discriminant regularization denoted by within-class scatter and between-class scatter [12]. In this paradigm, low-rank matrix approximation is solved in a supervised manner as the whole label information of the training database is accessible. That is, we regularize the representative basis derived from standard LR matrix recovery using class-specific discriminant criterion which is motivated by Fisher criterion, and plays an important role in face recognition tasks [12][13][14]. By introducing this type of regularization, our matrix completion algorithm is able to capture discriminative portions extracted from different classes.

2 Related Works

2.1 Discrimination in Face Recognition

The face recognition literature is fairly dense and diverse and thus cannot be surveyed in its entirety in this limited space. In this paper, we focus on the class of face recognition approaches called subspace methods that are more closely related to our method. A prime instance of such methods is Eigenfaces [15], which attempts to group images by minimizing data variance. Fisherfaces [12], due to finding a subspace that minimizes the within-class distances while maximizing the between-class distances at the same time, achieves much better classification performance than Eigenfaces in face recognition problem. Some other subspace methods are geometrically inspired where the emphasis is on identifying a low dimensional sub-manifold on which the face images lie. The most successful of these methods include those which seek to project images to a lower dimensional subspace such that the local neighborhood structure present in the training set is maintained. These include Laplacianfaces [16], Locality Preserving Projections (LPP) [17], Orthogonal Laplacianfaces [18], Marginal Fisher Analysis (MFA)[19] etc.. Over time, improvements on discrimination of these methods have appeared in [20][21][22][23][24]. These generalizations seriously make the discriminant regularization as an indispensable part of their models, and therefore great improvements can be witnessed.

2.2 Sparse Representation-Based Classification

Recently, Wright et al [4] proposed a sparse representation-based classification algorithm for face recognition. In SRC-based algorithms, each testing image is regarded as a sparse linear combination of the whole training data by solving an ℓ_1 minimization problem, and very impressive results were reported in [4]. Several works have been proposed to further extend SRC-based algorithms for improved performance. For example, [25] utilizes a LASSO type regularization for computing the joint sparse representation of different features for visual signals. Jenatton et al. [26] utilizes a tree-structured sparse regularization for hierarchical sparse coding. Although promising face recognition results were reported by SRC-based algorithm, it still requires clean face images for training and thus might not be preferable for real-world scenarios. If corrupted training data is presented, SRC-based algorithms tend to recognize testing images with the same type of corruption and thus lead to poor performance. In the following section, we will introduce our proposed method for robust face recognition, in which both training and testing data can be corrupted.

2.3 Matrix Recovery via Rank Minimization

Low-rank matrix recovery is a procedure for reconstructing an unknown matrix with low-rank or approximately low-rank constraints from a sampling of its entries. This problem is motivated by the requirement of inferring global structure from a small number of local observations. [10], a breakthrough in matrix completion algorithms, states that the minimization of the rank function under broad conditions can be achieved using the minimizer obtained with the nuclear norm (sum of singular values). Since the natural reformulation of the nuclear norm gives rise to a semi-definite program, existing interior point methods can only handle problems with a number of variables in the order of the hundreds. Recently, Robust PCA method [9] has been proved to achieve the state-of-the-art performance using Augmented Lagrange Multipliers (ALM) method [11]. The proposed algorithm is also solved within the framework of ALM due to its fast efficiency. In the context of computer vision and pattern recognition, minimization of the nuclear norm in matrix completion has been applied to several problems: structure from motion [27], RPCA [9][28], subspace alignment [29], subspace segmentation [30] and signal denoising [31] etc..

3 Proposed Algorithm

3.1 Problem Setting

Given the original dataset $X = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^{D\times n}$ consists of n columns, each column denotes a sample. Low-rank matrix recovery decomposes X into the following form

$$X = A + E, \tag{1}$$

where A is a low-rank matrix, and E is a sparse matrix. The dimension of matrices A and E is the same as X. According to [10], the solution of eq(1) can be solved by ALM

[11] method by optimizing the following model

$$\arg\min_{A,E} \|A\|_* + \lambda\|E\|_1, \quad s.t.\ X = A + E, \tag{2}$$

where $\|\|_*$ denotes nuclear norm, and $\|\|_1$ denotes ℓ_1 norm.

3.2 Within-Class and Between-Class Scatters

Assume that all the labels of data X are available. Specifically, let x_i^s denote the i-th sample of the s-th class. We derived with-in class scatter and between class scatter matrices in the following manner which is different from Fisherfaces [12].

Let w_s denote the within-class scatter of class s. Define it as

$$w_s = \sum_{i=1}^{c_s} \|x_i^s - \bar{x^s}\|_2^2, \quad s = 1, \ldots, c. \tag{3}$$

Let $X_s = [x_1^s, x_2^s, \ldots, x_{c_s}^s]$ denote the s-th class data matrix, c_s is the number of samples in class s, and e_{c_s} denote all-one column vector of length c_s. Then we have $\bar{x^s} = \frac{1}{c_s} X_s e_{c_s}$. Rewriting eq(3) shows

$$\begin{aligned} w_s &= \sum_{i=1}^{c_s} (x_i^s - \bar{x^s})(x_i^s - \bar{x^s})^T \\ &= Tr\{\sum_{i=1}^{c_s} x_i^s (x_i^s)^T\} - 2Tr\{\sum_{i=1}^{c_s} x_i^s (\frac{1}{c_s} X^s e_{c_s})^T\} + Tr\{(\frac{1}{c_s} X^s e_{c_s})(\frac{1}{c_s} X^s e_{c_s})^T\} \\ &= Tr(X_s X_s^T) - \frac{2}{c_s} Tr\{X_s e_{c_s} (e_{c_s})^T X_s^T\} + \frac{(e_{c_s})^T e_{c_s}}{c_s^2} Tr\{X_s e_{c_s} (e_{c_s})^T X_s^T\}, \end{aligned} \tag{4}$$

where Tr denotes trace operator of matrix. Thus we have

$$w_s = Tr\{X_s D_s X_s^T\}, \tag{5}$$

where $D_s = I_s - \frac{2}{c_s} e_{c_s} (e_{c_s})^T + \frac{(e_{c_s})^T e_{c_s}}{c_s^2} e_{c_s} (e_{c_s})^T$.

Next, we can define the between-class scatter of s-th class with the other classes

$$\beta_s = \sum_{j=1, j \neq s}^{c} \|\bar{x^s} - \bar{x^j}\|_2^2, \tag{6}$$

where c is the number of classes. Following similar formulations from eq(3) to (5), we can rewrite eq(6) as

$$\begin{aligned} \beta_s &= \sum_{j=1, j \neq s}^{c} (\bar{x^s} - \bar{x^j})(\bar{x^s} - \bar{x^j})^T \\ &= \sum_{j=1, j \neq s}^{c} Tr\{\bar{x^s}(\bar{x^s})^T - 2\bar{x^s}(\bar{x^j})^T + \bar{x^j}(\bar{x^j})^T\} \\ &= \frac{c-1}{c_s^2} Tr\{X_s e_{c_s} (e_{c_s})^T X_s^T\} - 2Tr\{\bar{x^s} \sum_{j=1, j \neq s}^{c} \bar{x^j}\} + Tr\{\sum_{j=1, j \neq s}^{c} \bar{x^j}(\bar{x^j})^T\} \\ &= Tr\{X_s B_1 X_s^T\} - Tr\{X_s B_2\} + B_3, \end{aligned} \tag{7}$$

where $B_1 = \frac{c-1}{c_s^2} e_{c_s}(e_{c_s})^T$, $B_2 = \frac{2}{c_s} e_{c_s} \sum_{j=1,j\neq s}^{c} \bar{x^j}$ and $B_3 = Tr\{\sum_{j=1,j\neq s}^{c} x^j (x^j)^T\}$.

3.3 Low-Rank Matrix Recovery Discrimination

Although low-rank matrix recovery decomposes the original data X and produces a low-rank matrix A together with a sparse error matrix E for better representation purpose, as shown in eq(1), the derived low-rank matrix A might not contain sufficient discriminating information. Assume that the original X represents face image data, we can rewrite it into class-wise form $X = [X_1, X_2, \ldots, X_c]$.

Based on the within-class scatter and between-class scatter matrices shown in eq(5) and (7), it is a natural idea of adding a discriminant regularization to the low-rank matrix recovery problem shown in eq(1)

$$\begin{array}{c} \arg\min_{A,E} \sum_{s=1}^{c} \{\|A_s\|_* + \lambda\|E_s\|_1 + \gamma(w_s(A_s) - \beta_s(A_s))\} \\ s.t.\ X_s = A_s + E_s, \end{array} \tag{8}$$

which is a class-wise optimization problem. In eq(8), $w_s(A_s)$ and $\beta_s(A_s)$ are the within-class scatter and between-class scatter of s-th class, respectively. Like LDA or Fisherfaces [12], to make projected samples favor of classification in feature space, we expect that the samples within the same class cluster as close as possible and samples between classes separate as far as possible in the learned low-rank matrix A. The term $\|A_s\|_* + \lambda\|E_s\|_1$ shown in eq(8) performs the standard low-rank decomposition of the data matrix X. The term $\gamma(w_s(A_s) - \beta_s(A_s))$ is our discriminant regularizer based on within-class and between-class scatters, which is penalized by the parameter γ balancing the low-rank matrix approximation and discrimination. We refer to eq(8) as low-rank matrix recovery with discriminant regularization.

Meanwhile, we can rewrite $(w_s(A_s) - \beta_s(A_s))$ into the following form

$$\begin{aligned} w_s(A_s) - \beta_s(A_s) &= Tr\{A_s D_s A_s^T\} - Tr\{A_s B_1 A_s^T\} + Tr\{A_s B_2\} - B_3 \\ &= Tr\{A_s (D_s - B_1) A_s^T\} + Tr\{A_s B_2\} - B_3 \\ &\leq \|A_s\|_F \|(D_s - B_1)\|_F \|A_s\|_F + \|A_s\|_* \|B_2\|_2 - B_3 \\ &= b_1 < A_s, A_s > + b_2 \|A_s\|_* - b_3, \end{aligned} \tag{9}$$

where

$$b_1 = \|(D_s - B_1)\|_F,\ \ b_2 = \|B_2\|_2\ \ and\ \ b_3 = B_3. \tag{10}$$

As b_3 is irrelevant to A_s, the optimization of eq(8) can be rewritten as

$$\begin{array}{c} \arg\min_{A_s,E_s} \|A_s\|_* + \lambda\|E_s\|_1 + \gamma(b_1 < A_s, A_s > + b_2\|A_s\|_*) \\ s.t.\ X_s = A_s + E_s. \end{array} \tag{11}$$

The optimization of eq(11) can be solved by ALM [11]. The general method of ALM is introduced for solving the following constrained optimization problem

$$\min f(X) \quad s.t. \quad h(X) = 0. \tag{12}$$

The corresponding ALM function of eq(12) is defined as

$$L(X, Y, \mu) = f(X) + < Y, h(X) > + \frac{\mu}{2}\|h(X)\|_F^2, \tag{13}$$

Algorithm 1 General Method of ALM

1: $\rho \geq 1$.
2: **while** not converged **do**
3: solve $X_{k+1} = \arg\min_X L(X_k, Y_k, \mu_k)$
4: $Y_{k+1} = Y_k + \mu_k h(X_k)$
5: update μ_k to μ_{k+1}
6: **end while**
7: Output: X_k

where Y is a Lagrange multiplier matrix and μ is a positive scalar. The solution to eq(13) is outlined as Algorithm1.

In the proposed eq(11), let $X = (A_s, E_s)$, then

$$\begin{aligned} f(X) &= \|A_s\|_* + \lambda\|E_s\|_1 + \gamma(b_1 < A_s, A_s > +b_2\|A_s\|_*), \\ h(X) &= X_s - A_s - E_s \end{aligned} \tag{14}$$

respectively. The ALM function of our eq(11) is

$$\begin{aligned} L(A_s, E_s, Y_s, \mu, \gamma) = \|A_s\|_* + \lambda\|E_s\|_1 + \gamma(b_1 < A_s, A_s > +b_2\|A_s\|_*) \\ + < Y_s, X_s - A_s - E_s > + \frac{\mu}{2}\|X_s - A_s - E_s\|_F^2. \end{aligned} \tag{15}$$

To solve eq(15), we can optimize A_s, E_s and Y_s iteratively.

– Updating A_s:
When updating A_s, we have to fix E_s and Y_s to solve the following problem based on eq(15), and the 3rd iteration of Algorithm(1) evolves

$$\begin{aligned} A_s^{k+1} &= \arg\min_{A_s^k} L(A_s^k, E_s^k, Y_s^k, \mu^k, \gamma) \\ &= \arg\min_{A_s^k} (1 + b_2)\|A_s^k\|_* + (\gamma b_1 + \frac{\mu^k}{2}) < A_s, A_s > +\mu^k < X_s^k - E_s^k + \frac{1}{\mu^k} Y_s^k, A_s^k > \\ &= \arg\min_{A_s^k} \epsilon\|A_s^k\|_* + \frac{1}{2}\|X_a - A_s^k\|_F^2, \end{aligned} \tag{16}$$

where $\epsilon = \frac{1+b_2}{2\gamma b_1 + \mu^k}$ and $X_a = \frac{\mu^k}{2\gamma b_1 + \mu^k}(X_s^k - E_s^k + \frac{1}{\mu^k} Y_s^k)$. Introducing the following soft-thresholding operator

$$S_\epsilon[x] \doteq \begin{cases} x - \epsilon, if \;\; x > \epsilon \\ x + \epsilon, if \;\; x < -\epsilon \\ 0, \qquad otherwise \end{cases}, \tag{17}$$

then we have the solution of eq(16) [11]

$$A_s^{k+1} = US_s[S]V^T, \tag{18}$$

where USV^T is the SVD of X_a.

– Updating E_s:
When updating E_s, we have to fix A_s and Y_s. The eq(15) can be derived as

$$E_s^{k+1} = \arg\min_{E_s^k} \eta \|E_s^k\|_1 + \frac{1}{2}\|X_e - E_s^k\|_F^2, \tag{19}$$

where $\eta = \lambda \frac{1}{\mu^k}$ and $X_e = \frac{\mu^k}{2\gamma b_1 + \mu^k}(X_s^k - A_s^{k+1} + \frac{1}{\mu^k} Y_s^k)$.

Once we obtain A_s and E_s, Y_s can be updated using the 4th iteration of Algorithm1. The whole method we proposed is described in Algorithm2.

Algorithm 2 Low-rank Matrix Recovery with Discrimination

```
1: Input observation matrix X, λ.
2: Input μ_0 > 0, ρ > 1 and η.
3: Compute Y_0^* = sgn(X)/J(sgn(D)).
4: while not converged do
5:    A_0^{k+1} = A_*^k, E_0^{k+1} = E_*^k, j = 0 and b_{1,2} shown in eq(10);
6:       while not converged do
7:          (U, S, V) = svd(X_s^k − E_s^k + (1/μ^k) Y_s^k);
8:          A_{j+1}^{k+1} = U S_ε[S] V^T;
9:          E_{j+1}^{k+1} = S_η(X_s^k − A_s^{k+1} + (1/μ^k) Y_s^k);
10:         j ← j + 1
11:      end while
12:   Y_{k+1}^* = Y_k^* + μ_k(X_s − A_*^{k+1} − E_*^{k+1})
13:   update μ_k to μ_{k+1}
14: end while
15: Output: (A_*^k, E_*^k).
```

3.4 LR with Discrimination for Face Recognition

Occlusion is a common challenging encountered in face recognition tasks, such as eyeglasses, sunglasses, scarves and some objects placed in front of the faces. Moreover, even in the absence of an occluding object, violations of an assumed model for face appearance may act like occlusions: e.g., shadows due to extreme illumination. Robustness to occlusion is therefore essential to practical face recognition system. If the face images are partially occluded, popular recognition methods based on holistic features such Eigenfaces [15], Fisherfaces [12] and Laplacianfaces [16] would lead to unacceptable performance due to the corruption of the extracted features. Although SRC-based algorithm [28] achieves better results in recognizing occluded testing images, it still requires unoccluded face images for training and thus might not be preferable for real application scenarios.

Low-rank matrix recovery has been applied to alleviate the aforementioned problems by decomposing the collected data matrix into two different parts, one is a representation basis matrix of low rank and the other is the corresponding sparse error, as shown in Fig.1.

(a) The original face images

(b) The standard low-rank recovery of (a)

(c) The standard sparse error of (a)

(d) The low-rank recovery of our method

(e) The sparse error of our method

Fig. 1. The results of low-rank matrix recovery with and without discrimination

We can find out from Fig.1 that when the standard low-rank matrix recovery is combined with discrimination, the face images within the recovered representation basis matrix tend to be more similar to each other for the same subject, which means more compactness exists within the same classes and dissimilarity between different classes. In addition, we also can conclude from Fig.1 that the sparse error with discrimination can remove more sparse noise. As a result, the representation basis matrix of low-rank recovery with discrimination has a better representative ability than the original version. Since the face images usually lie in high dimensional spaces, traditional dimensionality reduction techniques, like PCA or LDA, can be performed on the recovered representation basis matrix. As a result, the derived subspace can be applied as the dictionary for training and the testing purposes. In the recognition stage, one can also use SRC-based classification strategy to identify the input image. Our scheme for face recognition is described as Algorithm3.

Algorithm 3 LR with Discrimination for Face Recognition

1: Input training data $X = [X_1, X_2, \ldots, X_c]$ and a testing image y.
2: Use Algorithm2 on X to compute the representation basis matrix A.
3: Calculate the projection matrix of P of A.
4: Compute the projection of X and y:
 $X_p = P^T X$, and $y_p = P^T y$.
5: Perform SRC-based classification on y_p:
 $\arg\min_\alpha ||y_p - X_p\alpha||_2^2 + \lambda||\alpha||_1$,
 for $i = 1 : c$
  $err(i) = ||y_p - X_p^i \alpha_i||_2^2$
 end for
6: Output: $label(y) = \min_i err(i)$.

4 Experiments

In this section, we perform the proposed method shown in Algorithm3 on publicly available databases for face recognition to demonstrate the efficacy of the proposed classification algorithm. We will first examine the role of feature extraction within our framework, comparing performance across various feature spaces and feature dimensions, and comparing to several popular methods. Meanwhile, We will then demonstrate the robustness of the proposed algorithm to corruption and occlusion. Finally, the experimental results demonstrate the effectiveness of sparsity as a means of validating testing images.

Besides the standard low-rank matrix recovery without discrimination and our proposed method, we also consider Nearest Neighbor (NN), SRC [4], and LLC [32] for comparisons. Note that LLC can be regarded as an extended version of SRC exploiting data locality for improved sparse coding, and the classification rule is the same as that of SRC. To evaluate our recognition performance using data with different dimensions, we project the data onto the eigenspace derived by PCA using our LR with discrimination models. For the standard LR approach, the eigenspace spanned by LR matrices without discrimination is considered, while those of other SRC based methods are derived by the data matrix X directly. We vary the dimension of the eigenspace and compare the results in this section.

4.1 Two Databases

- The Extended Yale B database consists of $2,414$ frontal face images of 38 individuals around $59-64$ images for each person [33]. The cropped and normalized 192×168 face images were captured under various laboratory-controlled lighting conditions. Some sample face images of the Extended Yale B are shown in Fig.2(a).
- The AR database consists of over $4,000$ frontal images for 126 individuals [34]. For each individual, 26 pictures were taken in two separate sessions. In Fig.2(b), the left and right ones are some images collected in two sessions.

(a) (b)

Fig. 2. Some sample images of Extended Yale B and AR

4.2 Results

On the Extended Yale B Database. For each subject, we randomly select $10, 20$ and 30 images of each subject for training respectively, and the left images for testing. Randomly choosing the training set ensures that our results and conclusions will not depend on any special choice of the training data. We vary the dimension of the eigenspace as $25, 50, 75, 100, 150, 200, 300$ and 400 to compare the recognition performance between

Table 1. 10 training samples

METHOD	25	50	75	100	150	200	300
LR+SRC	67.3	75.7	79.1	82.2	83.4	84.6	86.8
SRC	61.8	69.6	76.9	82.1	83.3	85.8	86.2
LLC+SRC	44.6	62.5	68.7	73.4	75.7	77.3	78.6
NN	30.7	42.8	45.6	49.7	53.4	56.1	58.3
OURS	72.4	77.2	81.4	83.8	86.5	86.9	86.8

Table 2. 20 training samples

METHOD	25	50	75	100	150	200	300
LR+SRC	75.4	82.3	85.7	87.2	89.6	90.5	91.9
SRC	67.3	74.9	80.6	85.8	87.2	90.5	91.1
LLC+SRC	51.7	66.8	72.4	78.3	81.6	85.8	87.2
NN	38.1	45.7	52.2	58.4	62.4	66.3	69.8
OURS	82.4	84.6	86.3	89.1	92.9	93.1	93.5

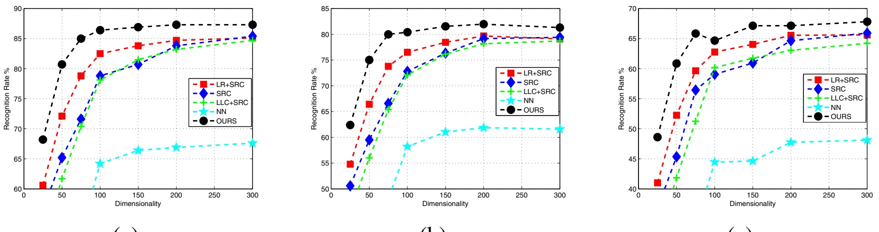
Table 3. 30 training samples

METHOD	25	50	75	100	150	200	300
LR+SRC	86.9	93.3	94.7	95.4	96.1	96.5	96.4
SRC	80.3	88.6	92.9	94.5	95.4	96.1	96.6
LLC+SRC	62.4	79.6	86.4	89.7	92.5	93.8	94.5
NN	45.1	56.0	63.3	66.8	69.4	73.2	76.7
OURS	90.5	94.3	95.8	96.2	97.1	97.5	97.5

different methods. All experiments run ten times and the average results are shown in Table1-3.

It is clear from those Tables mentioned above that the proposed method consistently achieves higher recognition rates than other NN and SRC-based approaches. For example, at dimension = 100, our method achieves a better recognition rate at 96.2%, and those for LR, SRC, LLC, and NN are 95.4%, 94.5%, 89.7%, and 66.8%, respectively (see Table3). Repeating the above experiments using different training images for each person, we can confirm from these empirical results that the use of LR method alleviates the problem of severe illumination variations even when such noise is presented in both training and testing data. Furthermore, when discrimination is taken into account as proposed in the paper, LR method exhibits enhanced classification capability and thus outperforms the standard LR algorithm.

On AR Database. In the experiment, a subset of the dataset consisting of 50 male subjects and 50 female subjects was chosen. The images are cropped with dimension 165 × 120. Different from [4], for each subject, both neutral (four neutral faces with different lighting conditions and three faces with different expressions) and corrupted images (three faces with sunglasses and three faces with scarfs) taken at session 1 are used for training, and session 2 for testing. Specifically, we consider the following sample selection for training: 7 neutral images plus 3 sunglass images; 7 neutral images plus 3 scarf images; 7 neutral images plus 3 sunglass images and 3 scarf images. We vary the dimension of the eigenspace as 25, 50, 75, 100, 150, 200, 300 and 400 to compare the recognition performance between different methods. The experimental results are visualized in Fig.3.



(a) (b) (c)

Fig. 3. Recognition rate on AR. (a)7 neutral + 3 sunglass images. (b)7 neutral + 3 scarf images. (c)session 1 as training set.

From these three figures, we see that the proposed method outperforms all other algorithms across different dimensions. It is worth noting that with the increase of occlusion (from sunglass ro scarf), the recognition rates of all the approaches are severely degraded, which can be seen from Fig.3(a) and Fig.3(b). In addition, with the increase of occluded images in the training set, the performances of all the approaches are also severely degraded which can be seen from Fig.3(c). These two cases indicate that the direct use of corrupted training image data will remarkably make the recognition results worse.

5 Conclusions

In this paper, a low-rank matrix recovery algorithm with discriminant regularization is proposed. The discrimination regularizer is motivated by Fisher criterion which plays an important role in classification tasks. The introduction of this kind of regularizer into low-rank matrix recovery promotes the discrimination power in the learned representation basis. We also show that the proposed optimization algorithm can be formulated by augmented Lagrange multipliers. When applied to face recognition problem, the proposed algorithm demonstrates robustness to severe occlusions of face images even in the training set. The experiments has shown that our method achieves the state-of-the-art recognition results.

Acknowledgments. The authors confirm that the research was supported by National Natural Science Foundation (No.61170109, No.61100119 and No.61272468), and Science and Technology Planning Project of Zhejiang Province (No.2012C21021), China.

References

1. Olshausen, B.A., Field, D.J.: Sparse coding with an over-complete basis ser: a strategy employed by v1? Vision Research 37(23), 3311–3325 (1997)
2. Vinje, W.E., Gallant, J.L.: Sparse coding and decorrelation in primary visual cortex during natural vision. Science 287(5456), 1273–1276 (2000)
3. Wright, J., Ma, Y., Mairal, J., Spairo, G., Huang, T., Yan, S.C.: Sparse representation for computer vision and pattern recognition. Proceedings of the IEEE 98(6), 1031–1044 (2010)
4. Wright, J., Yang, A.Y., Sastry, A.G.S.S., Ma, Y.: Robust face recognition via sparse representation. IEEE PAMI 31(2), 210–227 (2009)
5. Yang, M., Zhang, L., Yang, J., Zhang, D.: Robust sparse coding for face recognition. In: CVPR (2011)
6. Huang, K., Aviyente, S.: Sparse representation for signal classification. In: NIPS (2006)
7. Wagner, A., Wright, J., Ganesh, A., Zhou, Z.H., Ma, Y.: Towards a practical face recognition system: Robust registration and illumination by sparse representation. In: CVPR (2009)
8. Wagner, A., Wright, J., Ganesh, A., Zhou, Z.H., Ma, Y.: Towards a practical face recognition system: Robust registration and illumination by sparse representation. IEEE PAMI 34(2), 372–386 (2012)
9. Candes, E., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? Journal of ACM 58(1), 1–37 (2009)

10. Candes, E., Recht, B.: Exact low rank matrix completion via convex optimization. In: Allerton (2008)
11. Lin, Z., Chen, M., Ma, Y.: The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrix. UIUC Technical Report UILU-ENG-09-2215 (2009)
12. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs fisherfaces: recognition using class specific linear projection. IEEE PAMI 19(7), 711–720 (1997)
13. Li, Z., Lin, D., Tang, X.: Nonparametric discriminant analysis for face recognition. IEEE PAMI 31(4), 755–761 (2009)
14. Lu, J., Tan, Y., Wang, G.: Discriminaive multi-manifold analysis for face recognition from a single trainning sample per person. IEEE PAMI pp(99), 1 (2012)
15. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscinces 3, 72–86 (1991)
16. He, X., Yan, S., Hu, Y., Niyogi, P., Zhang, H.: For recognition using laplacianfaces. IEEE PAMI 27(3), 328–340 (2005)
17. He, X., Cai, D., Niyogi, P.: Locality preserving projections. In: NIPS (2003)
18. Cai, D., He, X., Han, J., Zhang, H.: Orthogonal laplacianfaces for face recognition. IEEE TIP 15(11), 3608–3614 (2006)
19. Yan, S., Xu, D., Zhang, B., Zhang, H., Yang, Q., Lin, S.: Graph embedding and extension: A general framework for dimensionality reduction. IEEE PAMI 29(1), 40–51 (2007)
20. Hua, G., Viola, P., Drucker, S.: Face recognition using discriminatively trained orthogonal rank one tensor projections. In: CVPR (2007)
21. Xue, H., Chen, S., Yang, Q.: Discriminatively regularized least-squares classification. Pattern Recognition 42(1), 93–104 (2009)
22. Si, S., Tao, D., Geng, B.: Bregman divergence-dased regularization for transfer subspace learning. IEEE TKDE 22(7), 929–942 (2010)
23. Lu, J., Tan, Y.: Cost-sensitive subspace learning for face recognition. In: CVPR (2010)
24. Lu, J., Tan, Y.: Regularized locality preserving projections and its extensions for face recognition. IEEE SMCB 40(3), 958–963 (2010)
25. Yuan, X., Yan, S.: classification with multi-task joint sparse representation. In: CVPR (2010)
26. Jenatton, R., Mairal, J., Obozinski, G., Bach, F.: Proximal methods for hierarchical sparse coding. Journal of Machine Learning Research 12, 2297–2334 (2011)
27. Cabral, R., Costeira, J., Torre, F., Bernardino, A.: Fast incremental method for matrix completion: an application to trajectory correction. In: ICIP (2011)
28. Wright, J., Ganesh, A., Rao, S., Ma, Y.: Robust principal component analysis: exact recovery of corrupted low rank matrices by convex optimization. In: NIPS (2009)
29. Peng, Y., Ganesh, A., Wright, J., Xu, W., Ma, Y.: Robust alignment by sparse and low-rank decomposition for linearly correlated images. In: CVPR (2010)
30. Liu, G., Lin, Z., Yu, Y.: Robust subspace segmentation by low-rank representation. In: ICML (2010)
31. Ji, H., Liu, C., Shen, Z., Xu, Y.: Robust video denoising using low rank matrix completion. In: CVPR (2010)
32. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality constrained linear coding for image classification. In: CVPR (2010)
33. Georghiades, A., Belhumeur, P., Kriegman, D.: From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE PAMI 23(6), 643–660 (2001)
34. Martinez, A., Benavente, R.: The ar face database. CVC Technical Report 24 (1998)