Learning Low-Rank Representations with Classwise Block-Diagonal Structure for Robust Face Recognition

Yong Li[1], **Jing Liu**[1], **Zechao Li**[2], **Yangmuzi Zhang**[3], **Hanqing Lu**[1], **Songde Ma**[1]
[1]National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
[2]School of Computer Science, Nanjing University of Science and Technology
[3]University of Maryland, College Park
{yong.li,jliu,luhq}@nlpr.ia.ac.cn, zechao.li@gmail.com, ymzhang@umiacs.umd.edu, masd@most.cn

Abstract

Face recognition has been widely studied due to its importance in various applications. However, the case that both training images and testing images are corrupted is not well addressed. Motivated by the success of low-rank matrix recovery, we propose a novel semi-supervised low-rank matrix recovery algorithm for robust face recognition. The proposed method can learn robust discriminative representations for both training images and testing images simultaneously by exploiting the classwise block-diagonal structure. Specifically, low-rank matrix approximation can handle the possible contamination of data. Moreover, the classwise block-diagonal structure is exploited to promote discrimination of representations for robust recognition. The above issues are formulated into a unified objective function and we design an efficient optimization procedure based on augmented Lagrange multiplier method to solve it. Extensive experiments on three public databases are performed to validate the effectiveness of our approach. The strong identification capability of representations with block-diagonal structure is verified.

Introduction

Face recognition has been actively studied over the past years due to its important role in a number of applications (i.e., access control, visual surveillance). Representations of faces play a important role in the face recognition system. Existing methods like Eigenfaces (Turk and Pentland 1991), Fisherfaces (Belhumeur, Hespanha, and Kriegman 1997), and Laplacianfaces (He et al. 2005) are proposed to learn low dimension representations to improve recognition efficiency. However, these methods are not robust to outliers or sparse noises such as occlusion, illumination changes, pixel corruption. To alleviate the aforementioned problems and learn robust representations, methods based on low-rank matrix recovery have been proposed and shown to achieve promising results (Chen, Wei, and Wang 2012)(Ma et al. 2012).

Low-rank matrix recovery is proposed to recover a low-rank matrix from corrupted observations, also known as Robust Principal Component Analysis (RPCA) (Wright et al. 2009b). It efficiently removes sparse noises from corrupted observations and has been widely applied to applications like background modeling (Cui et al. 2012), saliency detection (Shen and Wu 2012)(Peng et al. 2013), image classification (Zhang et al. 2011). However, the previous work only deals with unsupervised case and does not take advantage of structural information when the label information is available.

To leverage structural information in the face recognition problem, some methods are proposed based on low-rank matrix recovery (Chen, Wei, and Wang 2012)(Ma et al. 2012)(Zhang, Jiang, and Davis 2013). Chen et al. proposed a low-rank matrix approximation algorithm with structural incoherence (LRSI). The introduction of such incoherence would prefer the resulting low-rank matrix of different classes to be independent. Ma et al. presented a discriminative low-rank dictionary learning algorithm (DLRD_SR). The proposed algorithm tries to optimize the sub-dictionaries for each class to be low-rank. However, both of them explore structural information class by class, which do not capture the global structure.

To explore the global structure, Zhang et al. proposed a discriminative, structured low-rank method (DSLR). It tries to regularize all the training images of the same class to have the same representation code. However, it is not usually the case. Though images from the same class usually lie in the same subspace, it is unreasonable to assume representations of the same class images to be the same.

To this end, we focus on capturing the global data structure among the whole database. Since training images of the same class usually lie in the same subspace, and should be represented by the base elements of corresponding class, then the ideal representations of training images will have classwise block-diagonal structure as follows,

$$\begin{bmatrix} \bar{Z}_1^* & 0 & 0 & 0 \\ 0 & \bar{Z}_2^* & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \bar{Z}_C^* \end{bmatrix}$$

where $\bar{Z}_i^*$ is the representation matrix of the i-th class training images corresponding to the i-th class bases, and C is the class number. Representations with classwise block-diagonal structure have high intra-class similarities and large inter-class differences, which is critical for the recognition problem. Moreover, representations of training images and testing images should be consistent, which is important for the recognition process. Representations of training images and testing images should be learnt in a unified framework.

For this purpose, we propose a semi-supervised framework to learn low-rank representations with classwise block-diagonal structure (LR-CBDS). Representations of training images and testing images are learnt simultaneously, leveraging the global structure over the whole database. We introduce a classwise regularization term to capture classwise block-diagonal structure for representations of training images. A sparse error component is introduced to remove sparse noises from corrupted observations for robust recognition. Extensive experiments show that the learnt representations of testing images also have classwise block-diagonal structure, due to the fact that the learnt representations of testing images are relevant to representations of training images because of low-rank property. The main contributions of this paper are summarized as follows.

- We propose a semi-supervised framework to learn low-rank representations with classwise block-diagonal structure, which have strong identification capability.
- Our approach learns robust representations of training images and testing images simultaneously, which leverages the global structure over the whole database.
- Our approach is robust even when training images and testing images are both badly corrupted. Furthermore, it outperforms state-of-the-art methods both in time efficiency and recognition accuracy.

Learning Low-Rank Representations with Classwise Block-Diagonal Structure

Notations. Throughout this paper, notations with a bar denote symbols of training images, while notations with a hat denote symbols of testing images. $\bar{X} = [\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_C]$ is the feature matrix of C classes training images. It is rearranged based on the class label, and $\bar{X}_i \in R^{d\times n_i}$ corresponds to the feature matrix of i-th class images with n_i samples of dimension d. Meanwhile, $\hat{X} = [\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_{n_s}] \in R^{d\times n_s}$ denotes the feature matrix of n_s testing images, where each column of $\hat{X}$ denotes the feature of a testing image. Moreover, $X = [\bar{X}, \hat{X}]$ is the feature matrix of whole database. Each sample of X can be represented by the linear combination of base elements of dictionary A,

$$X = AZ, \tag{1}$$

where $Z = [\bar{Z}, \hat{Z}]$, and $\bar{Z}$ corresponds to the representation matrix of training images, while $\hat{Z}$ corresponds to the representation matrix of testing images.

Generally, this paper is to address the following problem.

Problem Definition. Given a set of corrupted images from multiple classes, the problem is to learn robust representations of training images and testing images simultaneously. The representations should have classwise block-diagonal structure.

Low rankness is an appropriate criterion to capture low-dimensional structure in high-dimensional data. For the clean database X, Learning low-rank representations can be solved as follows,

$$\begin{aligned} &\min_{Z} rank(Z) \\ &s.t., X = AZ. \end{aligned} \tag{2}$$

Without loss of generality, we assume that the training feature matrix $\bar{X}$ in X is rearranged based on the class label, since interchanging columns does not change the value of rank function. The discrete nature of the rank function makes it difficult to solve, Liu et al.(2013) shows that the following convex optimization provides a good surrogate for problem (2),

$$\begin{aligned} &\min_{Z} \|Z\|_* \\ &s.t., X = AZ, \end{aligned} \tag{3}$$

where $\|Z\|_*$ is the nuclear norm (i.e., the sum of the singular values) of Z. When we do not have any learnt dictionary, the training data $\bar{X}$ itself can be used as the dictionary.

Generally, training samples should be reconstructed by the bases of corresponding class, then the ideal representation $\bar{Z}$ of training images will have classwise block-diagonal structure as follows.

$$\bar{Z} = \begin{bmatrix} \bar{Z}_1^* & 0 & 0 & 0 \\ 0 & \bar{Z}_2^* & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \bar{Z}_C^* \end{bmatrix}$$

To capture the classwise block-diagonal structure, we introduce a classwise regularization term $\sum_i^C \|\bar{Z}_{-i}\|_F^2$, where $\|\cdot\|_F$ denotes the Frobenius norm of a matrix, and $\bar{Z}_{-i}$ is the representation matrix of the i-th class training images subject to the dictionary $\bar{X}_{-i}$. $\bar{X}_{-i}$ stands for the feature matrix excluding the i-th class training images. It is to regularize representations of training images to be represented by the bases of corresponding class. By taking consideration of sparse noises, we formulate the final target function as follows,

$$\begin{aligned} &\min_{Z,E} \|Z\|_* + \lambda\|E\|_1 + \frac{\alpha}{2}(\sum_{i=1}^{C} \|\bar{Z}_{-i}\|_F^2 + \|\hat{Z}\|_F^2) \\ &s.t., X = \bar{X}Z + E, \end{aligned} \tag{4}$$

where E is corresponding to the sparse noise matrix and $\|E\|_1 = \sum_{i,j} |E_{i,j}|$. λ controls the sparsities of the noise matrix E, and α controls the contribution of structural regularization term. Representations of training images and testing images are learnt in a unified framework, which leverages the global structure over the whole database. Representations of testing images learnt by (4) will have classwise block-diagonal structure because of the low-rank property of Z and structural regularization to Z.

Optimization Algorithm

To solve optimization problem (4), we first introduce an auxiliary variable J and convert (4) to the following equivalent problem,

$$\begin{aligned}\min_{Z,E,J} &\|J\|_*+\lambda\|E\|_1+\frac{\alpha}{2}(\sum_{i=1}^{C}\|\bar{Z}_{-i}\|_F^2+\|\hat{Z}\|_F^2)\\ &s.t., X=\bar{X}Z+E\\ &\quad Z=J,\end{aligned} \tag{5}$$

which can be solved by the Augmented Lagrange Multiplier (ALM) method (Lin, Chen, and Ma 2010). Correspondingly, the augmented Lagrangian function form of (5) is as follows,

$$\begin{aligned}&L(Z,E,J,Y_1,Y_2,\mu)\\ &=\|J\|_*+\lambda\|E\|_1+\frac{\alpha}{2}(\sum_{i=1}^{C}\|\bar{Z}_{-i}\|_F^2+\|\hat{Z}\|_F^2)\\ &+<Y_1,X-\bar{X}Z-E>+<Y_2,Z-J>\\ &+\frac{\mu}{2}(\|X-\bar{X}Z-E\|_F^2+\|Z-J\|_F^2),\end{aligned} \tag{6}$$

where $<A,B>=trace(A^tB)$. Y_1 and Y_2 are Lagrange multipliers and $\mu>0$ is a penalty parameter. The optimization of (6) can be solved iteratively by updating J, Z and E once at a time. The updating scheme is as follows.

Updating J: Fix the other variables and solve the following problem,

$$\begin{aligned}J^{k+1}&=\arg\min_J\|J\|_*+<Y_2^k,Z^k-J>+\frac{\mu^k}{2}\|Z^k-J\|_F^2\\ &=\arg\min_J\frac{1}{\mu^k}\|J\|_*+\frac{1}{2}\|J-(Z^k+Y_2^k/\mu^k)\|_F^2\\ &=US_{\frac{1}{\mu^k}}[\Sigma]V^T.\end{aligned} \tag{7}$$

where $U\Sigma V^T$ is the singular value decomposition of the matrix $(Z^k+Y_2^k/\mu^k)$, and $S_\varepsilon[\cdot]$ is the soft-thresholding (shrinkage) operator defined as follows (Lin, Chen, and Ma 2010),

$$S_\varepsilon[x]=\begin{cases}x-\varepsilon, & \text{if } x>\varepsilon,\\ x+\varepsilon, & \text{if } x<-\varepsilon,\\ 0, & \text{otherwise.}\end{cases} \tag{8}$$

Updating Z: Fix the other variables and solve the following problem,

$$\begin{aligned}Z^{k+1}&=\arg\min_Z\frac{\alpha}{2}(\sum_{i=1}^{C}\|\bar{Z}_{-i}\|_F^2+\|\hat{Z}\|_F^2)\\ &+<Y_1^k,X-\bar{X}Z-E^k>+<Y_2^k,Z-J^{k+1}>\\ &+\frac{\mu^k}{2}(\|X-\bar{X}Z-E^k\|_F^2+\|Z-J^{k+1}\|_F^2).\end{aligned} \tag{9}$$

To solve (9) in a unified matrix form, we introduce an auxiliary matrix $Q^k=[\bar{Q}^k,\hat{Q}^k]$, defined as follows,

$$\bar{Q}^k=\begin{bmatrix}\bar{Z}_1^k & 0 & 0 & 0\\ 0 & \bar{Z}_2^k & 0 & 0\\ 0 & 0 & \ddots & 0\\ 0 & 0 & 0 & \bar{Z}_C^k\end{bmatrix}$$

and $\hat{Q}^k$ is a zero matrix with the same size as $\hat{Z}^k$, while $\hat{Z}^k$ is the sub-matrix of Z^k corresponding to the representation of testing images. Then problem (9) can be relaxed into the following formulation,

$$\begin{aligned}Z^{k+1}&=\arg\min_Z\frac{\alpha}{2}\|Z-Q^k\|_F^2\\ &+<Y_1^k,X-\bar{X}Z-E^k>+<Y_2^k,Z-J^{k+1}>\\ &+\frac{\mu^k}{2}(\|X-\bar{X}Z-E^k\|_F^2+\|Z-J^{k+1}\|_F^2).\end{aligned} \tag{10}$$

Since problem (10) is convex, it has closed form solution as follows,

$$\begin{aligned}Z=&((\alpha/\mu^k+1)I+\bar{X}^T\bar{X})^{-1}(\bar{X}^T(X-E^k)\\ &+J^{k+1}+(\alpha Q^k+\bar{X}^TY_1^k-Y_2^k)/\mu^k).\end{aligned} \tag{11}$$

Updating E: Fix the other variables and solve the following problem,

$$\begin{aligned}E^{k+1}&=\arg\min_E\lambda\|E\|_1+<Y_1^k,X-\bar{X}Z^{k+1}-E>\\ &+\frac{\mu^k}{2}(\|X-\bar{X}Z^{k+1}-E\|_F^2)\\ &=\arg\min_E\frac{\lambda}{\mu^k}\|E\|_1\\ &+\frac{1}{2}\|E-(X-\bar{X}Z^{k+1}+Y_1^k/\mu^k)\|_F^2\\ &=S_{\frac{\lambda}{\mu^k}}[X-\bar{X}Z^{k+1}+Y_1^k/\mu^k].\end{aligned} \tag{12}$$

Generally, we outline the optimization process in Algorithm 1.

Algorithm 1 Solving Problem (4) by Inexact ALM

Input: Feature Matrix X, parameter λ and α

1: **Initialize:** $Z^0=0$, $J^0=0$, $E^0=0$, $Y_1^0=0$, $Y_2^0=0$, $\mu^0=10^{-5}$, $\mu_{max}=10^8$, $\rho=1.1$, $\varepsilon=10^{-6}$
2: **While** not converged **do**
3: $J^{k+1}=\arg\min_J L(Z^k,E^k,J,Y_1^k,Y_2^k,\mu^k)$ of (7)
4: $Z^{k+1}=\arg\min_Z L(Z,E^k,J^{k+1},Y_1^k,Y_2^k,\mu^k)$ of (9)
5: $E^{k+1}=\arg\min_E L(Z^{k+1},E,J^{k+1},Y_1^k,Y_2^k,\mu^k)$ of (12)
6: update the multipliers:
$Y_1^{k+1}=Y_1^k+\mu^k(X-\bar{X}Z^{k+1}-E^{k+1})$
$Y_2^{k+1}=Y_2^k+\mu^k(Z^{k+1}-J^{k+1})$
7: update μ:
$\mu^{k+1}=min(\rho\mu^k,\mu_{max})$
8: check the convergence conditions:
$\|X-\bar{X}Z^{k+1}-E^{k+1}\|_\infty<\varepsilon$ and $\|Z^{k+1}-J^{k+1}\|_\infty<\varepsilon$
9: **End While**

Output: Z,E

Classification Based on LR-CBDS

We use a linear classifier for classification as in DSLR (Zhang, Jiang, and Davis 2013). Multivariate ridge regression model is used to obtain a linear classifier W^* based on the representations $\bar{Z}$ of training images and class label matrix H. The formulation to learn the linear classifier is as follows,

$$W^* = \arg\min_W \|H - W\bar{Z}\|_F^2 + \gamma\|W\|_F^2, \qquad (13)$$

where γ is the weight of regularization term. Formulation (13) is a convex problem, and it has closed form solution as follows,

$$W^* = H\bar{Z}^T(\bar{Z}\bar{Z}^T + \gamma I)^{-1}. \qquad (14)$$

Then the class label for a testing image i can be obtained as follows,

$$k^* = \arg\max_k W^* \hat{z}_i, \qquad (15)$$

where k^* is corresponding to the classifier with largest output.

Experiments and Comparisons

We evaluate our approach on three public databases: the Extended Yale B database (Georghiades, Belhumeur, and Kriegman 2001) (Lee, Ho, and Kriegman 2005), the AR database (Martinez and Benavente. 1998) and the ORL database (Samaria and Harter 1994). We deal with the case that both training and testing images are corrupted, including illumination changes, occlusions, uniform distributed noises and block noises. Extensive experiments are performed to validate the effectiveness of our method. We repeat each experiment 10 times and report the average accuracy. Our approach is compared with several related works, including SR (Wright et al. 2009a), SVM (Ma et al. 2012), RPCA (Wright et al. 2009b), LLC (Wang et al. 2010), LRSI (Chen, Wei, and Wang 2012) , DLRD_SR (Ma et al. 2012), DSLR (Zhang, Jiang, and Davis 2013). For fair comparison, we directly cite results of comparison methods reported in DSLR on the Extended Yale B database and AR database. and we cite results of comparison methods reported in DLRD_SR on the ORL database. All the experiments are performed on matlab with computer configuration as follows, CPU: i5-2400 3.10GHz RAM:14.0GB.

Extended Yale B Database

The Extended Yale B database consists of 2414 frontal-face images of 38 individuals. The images were taken under different poses and illumination conditions with size 192×168. For each person, there are about 59-64 images. We test our approach on the down-sampled images with sample rate $1/2, 1/4, 1/8$, and the corresponding feature dimension is 8064, 2016 and 504. Following the protocol in DSLR (Zhang, Jiang, and Davis 2013), we first randomly select N_c images for each person as training images ($N_c = 8, 32$), and the rest as testing images. For the case of 8 training images per class, the testing image set is randomly divided into four batches and each batch is dealt with parallelly to take advantage of classwise block-diagonal structure of training images. Comparison of different methods on Extended Yale B database is listed in Table 1.

Experimental results show that our approach outperforms DSLR, LRSI, RPCA, SR and LLC in both cases. It outperforms the state-of-art method DSLR by 1.5% improvement with 8 training images per person, and 5.4% improvement with 32 training images per person on average. Our approach improves obviously in the case of 32 training images per person, since face images with illumination changes usually lie in low-dimensional subspace, and feature matrix with sufficient training images exhibits good low-rank property. By taking advantage of classwise block-diagonal structure, our approach has better classification capability and is robust to illumination changes.

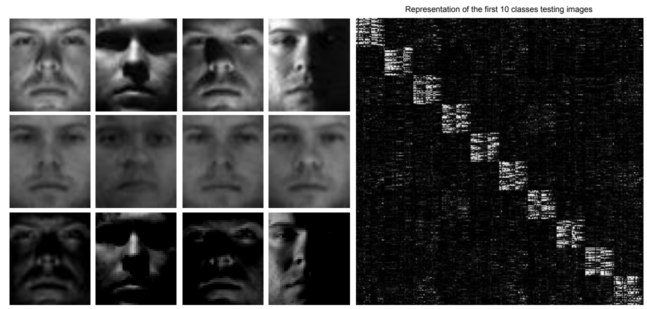

Figure 1: (a) Decomposition results of testing images; the top four images are original images with illumination changes, while the middle four images are the low-rank recovery images and the bottom four images are corresponding to sparse noises; (b) Representations of testing images from the first ten classes have classwise block-diagonal structure.

We present visualization of some decomposition results of testing images in Figure 1(a). The top four images shows the original images, while the middle four images and the bottom four images are corresponding to the low-rank recovery term $\bar{X}Z$ and the sparse error term E. It turns out that illumination changes can be removed as sparse noises. Figure 1(b) shows representations of the first ten class testing samples corresponding to the first ten class reconstruction bases. It is obtained with 32 training samples per class at sample rate $1/4$. The testing samples are mainly represented by the bases of corresponding class. The learnt representation has classwise block-diagonal structure, and shows strong identification capability.

Furthermore, we evaluate the computation time in Table 2 with 32 training samples per class at sample rate $1/8$. Table 2 shows that our approach is more efficient than DSLR and LRSI. Methods of LRSI and DSLR handle the face recognition problem as a two step approach. First, a noise-free dictionary is learnt based on the structure information among

Table 1: Recognition accuracy on Extended Yale B

No. per Class	N_c=8			N_c=32		
Sample Rate	1/8	1/4	1/2	1/8	1/4	1/2
LR-CBDS	**80.2**	**83.9**	**84.5**	**97.1**	**99.1**	**99.3**
DSLR	76.6	83.7	83.8	89.9	93.6	95.7
LRSI	73.3	80.9	80.8	89.5	93.3	94.5
RPCA	74.6	78.3	80.2	85.6	90.7	94.1
SR	79.3	83.0	83.8	87.2	89.5	90.7
LLC	65.7	70.6	76.1	76.4	80.0	85.6

training images. Then, sparse representation based classification method is adopted for face recognition based on the learnt dictionary (LRSI), or low-rank matrix recovery is performed to the testing images and the training images separately with the learnt dictionary and a linear classifier is adopted for recognition (DSLR). The dictionary learning process is highly time-consuming. The average computation time of DSLR and LRSI to update dictionary in each iteration is 289s and 72s. Generally, the final iteration number R_1 of DSLR is larger than 10, and the iteration number R_2 of LRSI is larger than R_1. Compared with DSLR and LRSI, our approach avoids the time-consuming dictionary learning process. Furthermore, our approach learns representations of training images and testing images simultaneously and use a linear classifier for classification, which is more efficient during testing process.

Table 2: Computation time on Extended Yale B database

Method	LR-CBDS	DSLR	LRSI
Dictionary learning (s)	-	$289\times R_1$	$72\times R_2$
Testing time (s)	0.435	0.527	3.354

AR Database

The AR database contains over 4000 face images of 126 individuals. For each individual, there are 26 images taken in two separate sessions under different illumination and expression changes. For each session, there are 13 images, in which 3 images are obscured by sunglasses, 3 images are obscured by scarves and the remaining 7 unobscured images are of different illumination and expression changes. Each image is with 165×120 pixels. Following the protocol in DSLR (Zhang, Jiang, and Davis 2013), we convert the color images to gray scale and down-sample at rate $1/3$, then the feature dimension of a sample is 2200. Experiments are performed under three different scenarios.

Sunglasses: In this scenario, we consider the obscured images due to occlusion of sunglass and unobscured images of illumination and expression changes. 7 unobscured images and 1 obscured image with sunglasses (randomly chosen) at session 1 are used for training (8 training images per class), and the remaining unobscured images and images with sunglasses from both sessions are used for testing (12 testing images per class). Sunglasses occlude about 20% area of the face image.

Scarf: In this scenario, we consider the obscured images due to occlusion of scarf and unobscured images of illumination and expression changes. 7 unobscured images and 1 obscured image with scarf (randomly chosen) at session 1 are used for training (8 training images per class), and the rest are used for testing (12 testing images per class). Scarf occludes about 40% area of the face image.

Mixed (Sunglasses+Scarf): In this scenario, we consider the obscured images with sunglasses and scarf, and the unobscured images of illumination and expression changes. 7 unobscured images, 2 corrupted images (one image with sunglass and one with scarf) at session 1 are used for training (9 training images per class), and the rest are used for testing (17 testing images per class).

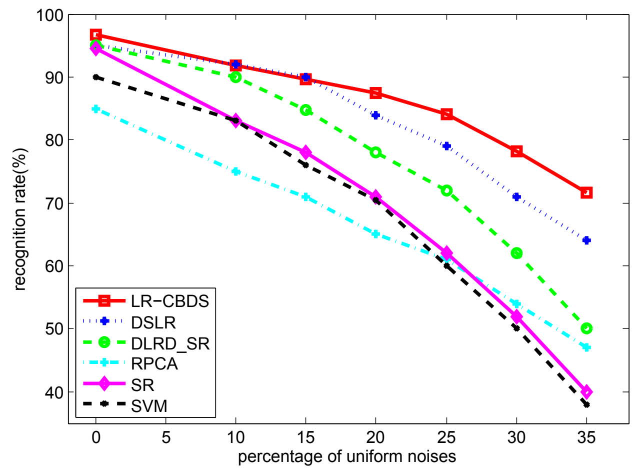

Figure 2: Recognition accuracy on AR database with different levels of pixel corruption.

The comparison of different methods on AR database is listed in Table 3. Our approach achieves the best result and outperforms the state-of-art method by 7.1% for the sunglasses scenario, 9.5% for the scarf scenario, and 10.4% for the mixed scenario. Our approach shows robustness to occlusions, illumination and expression changes.

Furthermore, we evaluate our algorithm on the AR database with uniform noises. Following the protocol in DSLR (Zhang, Jiang, and Davis 2013), 7 unobscured images with illumination and expression changes from session 1 are used for training, and the other 7 unobscured images from session 2 are used for testing. A percentage of randomly chosen pixels from both training images and testing images are replaced with samples from a uniform distribution over $[0, V_{max}]$, where V_{max} is the largest possible pixel value in the image. The recognition accuracy under different levels of corruption is shown in Figure 2. Our approach outperforms the state-of-art method by 3.5% improvement on average. Specially, our approach is robust to severe noises, and outperforms DSLR by 7.6% improvement with 35% uniform noises.

Figure 3(a) shows image decomposition examples with 10% uniform noises. Our approach separates uniform noises from the original images into the sparse error component. Figure 3(b) shows representations of testing images of the first ten classes, corresponding to the reconstruction bases

Table 3: Recognition accuracy on AR database

Scenario	Sunglasses	Scarf	Mixed
LR-CBDS	**94.4**	**92.9**	**92.8**
DSLR	87.3	83.4	82.4
LRSI	84.9	81.3	81.0
RPCA	83.2	75.8	78.9
SR	86.8	83.2	79.2
LLC	65.3	59.2	59.9

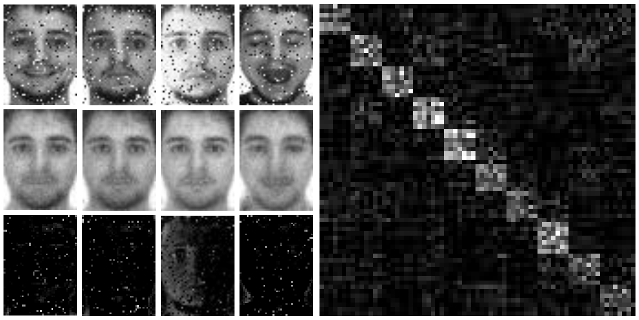

Figure 3: (a) Decomposition results of testing images on AR database; the top four images are original images with 10% uniform noises, while the middle four images are the low-rank recovery images and the bottom four images are corresponding to uniform noises and expression changes; (b) Representations of testing images from the first ten classes on AR database have classwise block-diagonal structure.

of the first ten classes. Representations of testing images have classwise block-diagonal structure, which increases the inter-class differences. Experimental results show that classwise block-diagonal structure is critical for the face recognition problem, especially when both training images and testing images are badly corrupted. Table 4 shows that our approach is more efficient than DSLR and LRSI in the sunglasses scenario. Moreover, the iteration number R_2 of LRSI during dictionary learning process improves obviously with the increase of class number.

ORL Database

The ORL database contains 400 images of 40 individuals. the images were taken at different time, with lighting variations, facial expression changes. All the images were taken against a dark homogeneous background with the individual in an upright, frontal position. Following the protocol in DLRD_SR (Ma et al. 2012), all the images are cropped to the size 28×23. For each individual, half images are randomly chosen as the training samples, and the rest are used as testing samples. A randomly located block of each image is replaced with an unrelated random image. Experiments are performed under different levels of block corruption, which is measured by the area proportion between the block and the whole image. Comparison of different methods on ORL database is listed in Table 5. Our approach achieves comparable performance with the state-of-art method with small percent block corruption, but our approach shows high robustness to the severe corruption and outperforms DLRD_SR by 2.9% improvement with 50% corruption.

Figure 4(a) shows image decomposition examples with

Table 4: Computation time on AR database

Method	LR-CBDS	DSLR	LRSI
Dictionary learning (s)	-	$167\times R_1$	$76\times R_2$
Testing time (s)	0.223	0.261	13.500

Figure 4: (a) Decomposition results of testing images on ORL database; the top five images are original images with 10% block corruption, while the middle five images are the low-rank recovery images and the bottom five images are corresponding to sparse noises including block corruption and facial expression changes; (b) Representations of testing images from the first ten classes on ORL database have classwise block-diagonal structure.

10% block occlusion. Our approach separates block occlusion and facial expression changes from the original images into the sparse noise component, and basic pattern for each person is left in the recovery image for robust recognition. Figure 4(b) shows representations of the testing images from the first ten classes, corresponding to the reconstruction bases of the first ten classes. Representations of testing images have classwise block-diagonal structure, which dominates the difference between different classes.

Conclusions

In this paper, we propose a generic semi-supervised framework to learn low-rank representations with classwise block-diagonal structure, which have strong identification capability for face recognition. It learns robust representations of training images and testing images simultaneously, which leverages the global data structure over the whole database. Moreover, by taking classwise block-diagonal structure into consideration, the proposed approach is robust even when training images and testing images are badly corrupted. Extensive experiments on three public databases indicate that the proposed approach achieves superior performance. Furthermore, the proposed approach can be applied but not limited to the face recognition problem. It can be extended to general classification problems to capture classwise structure information.

Table 5: Recognition accuracy on ORL database

Noise Percent	0	10	20	30	40	50
LR-CBDS	**96**	94.3	**91.4**	**86.5**	**79.4**	**72.8**
DLRD_SR	95.9	**94.4**	91.1	86	76.7	69.9
RPCA	89.3	88	83	76.6	72	66.2
SR	95.2	91.7	86	75.8	61.8	54
SVM	94.6	88.5	80.6	71.6	57.3	42

Acknowledgements

This work was supported by 973 Program (2010CB327905) and National Natural Science Foundation of China (61272329, 61273034, and 61070104)

References

Belhumeur, P.; Hespanha, J.; and Kriegman, D. 1997. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 19(7):711–720.

Chen, C.-F.; Wei, C.-P.; and Wang, Y.-C. 2012. Low-rank matrix recovery with structural incoherence for robust face recognition. In *Computer Vision and Pattern Recognition*, 2618–2625.

Cui, X.; Huang, J.; Zhang, S.; and Metaxas, D. 2012. Background subtraction using low rank and group sparsity constraints. In *ECCV*, volume 7572. 612–625.

Georghiades, A.; Belhumeur, P.; and Kriegman, D. 2001. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intelligence* 23(6):643–660.

He, X.; Yan, S.; Hu, Y.; Niyogi, P.; and Zhang, H.-J. 2005. Face recognition using laplacianfaces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 27(3):328–340.

Lee, K.-C.; Ho, J.; and Kriegman, D. 2005. Acquiring linear subspaces for face recognition under variable lighting. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 27(5):684–698.

Lin, Z.; Chen, M.; and Ma, Y. 2010. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *UIUC Technical Report UILU-ENG-09-2214*.

Liu, G.; Lin, Z.; Yan, S.; Sun, J.; Yu, Y.; and Ma, Y. 2013. Robust recovery of subspace structures by low-rank representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 35(1):171–184.

Ma, L.; Wang, C.; Xiao, B.; and Zhou, W. 2012. Sparse representation for face recognition based on discriminative low-rank dictionary learning. In *Computer Vision and Pattern Recognition*, 2586–2593.

Martinez, A., and Benavente., R. 1998. The ar face database. *CVC Technical Report 24*.

Peng, H.; Li, B.; Ji, R.; Hu, W.; Xiong, W.; and Lang, C. 2013. Salient object detection via low-rank and structured sparse matrix decomposition. In *AAAI Conference on Artificial Intelligence*.

Samaria, F. S., and Harter, A. C. 1994. Parameterisation of a stochastic model for human face identification. In *Applications of Computer Vision, Proceedings of the Second IEEE Workshop on*, 138–142.

Shen, X., and Wu, Y. 2012. A unified approach to salient object detection via low rank matrix recovery. In *Computer Vision and Pattern Recognition*, 853–860.

Turk, M., and Pentland, A. 1991. Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition*, 586–591.

Wang, J.; Yang, J.; Yu, K.; Lv, F.; Huang, T.; and Gong, Y. 2010. Locality-constrained linear coding for image classification. In *Computer Vision and Pattern Recognition*, 3360–3367.

Wright, J.; Yang, A.; Ganesh, A.; Sastry, S.; and Ma, Y. 2009a. Robust face recognition via sparse representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 31(2):210–227.

Wright, J.; Ganesh, A.; Rao, S.; Peng, Y.; and Ma, Y. 2009b. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In *Advances in Neural Information Processing Systems*. 2080–2088.

Zhang, C.; Liu, J.; Tian, Q.; Xu, C.; Lu, H.; and Ma, S. 2011. Image classification by non-negative sparse coding, low-rank and sparse decomposition. In *Computer Vision and Pattern Recognition*, 1673–1680.

Zhang, Y.; Jiang, Z.; and Davis, L. 2013. Learning structured low-rank representations for image classification. In *Computer Vision and Pattern Recognition*, 676–683.