Low-Rank Matrix Recovery with Structural Incoherence for Robust Face Recognition

Chih-Fan Chen Chia-Po Wei Yu-Chiang Frank Wang
Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan
{ryanchen, cpwei , ycwang}@citi.sinica.edu.tw

Abstract

We address the problem of robust face recognition, in which both training and test image data might be corrupted due to occlusion and disguise. From standard face recognition algorithms such as Eigenfaces to recently proposed sparse representation-based classification (SRC) methods, most prior works did not consider possible contamination of data during training, and thus the associated performance might be degraded. Based on the recent success of low-rank matrix recovery, we propose a novel low-rank matrix approximation algorithm with structural incoherence for robust face recognition. Our method not only decomposes raw training data into a set of representative basis with corresponding sparse errors for better modeling the face images, we further advocate the structural incoherence between the basis learned from different classes. These basis are encouraged to be as independent as possible due to the regularization on structural incoherence. We show that this provides additional discriminating ability to the original low-rank models for improved performance. Experimental results on public face databases verify the effectiveness and robustness of our method, which is also shown to outperform state-of-the-art SRC based approaches.

1. Introduction

Face recognition is among the most popular biometric approaches due to its low intrusiveness and high uniqueness. Unlike other physiological and behavioral biometric techniques like fingerprint or gait recognition which typically require cooperative subjects, face images can be acquired both actively or passively by surveillance cameras. With the increasing need for security-related applications, face recognition has been an active topic for researchers in the areas of computer vision and image processing.

To design a face recognition system, given training face image data, one typically focuses on the extraction of facial features and the learning of classification models. Unseen test data from the same subjects will be used to evaluate the associated identification or verification performance. To further assess the *robustness* of the designed face recognition algorithm, *occlusion* and *disguise* might be presented in the test image data. We note that while the test data might be corrupted, the training data is often assumed to be taken under a well controlled setting (i.e., under reasonable illumination, pose, etc. variations without occlusion or disguise). As a result, when applying existing face recognition methods for practical scenarios, we will need to discard corrupted training images, and we might thus encounter small sample size and over-fitting problems. Moreover, the disregard of corrupted training face images might give up some valuable information for recognition.

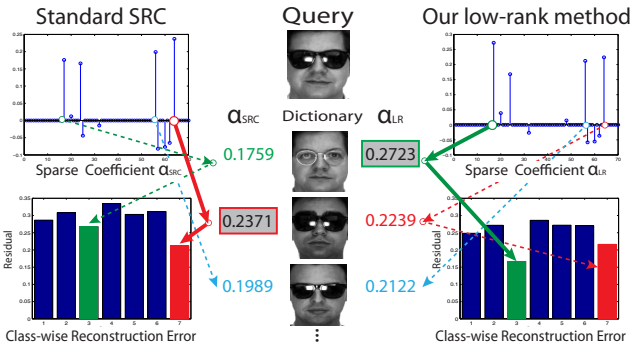


Figure 1. Comparison between the standard SRC and our method. Note that the standard SRC classifies the test input as the class with most similar training images even if they are corrupted (e.g. due to sunglasses), while our approach alleviates this problem and is robust to such occlusions presented in both training and test data.

For face recognition, it is common to apply existing techniques such as Eigenfaces [16], Fisherfaces [1], or Laplacianfaces [9] to reduce the dimension of the face images. As a result, the derived subspace is expected to achieve improved recognition performance. However, these techniques are not robust to outliers or sparse/extreme noise such as occlusion and disguise [5]. Some recent works on robust PCA have been proposed to alleviate the aforementioned problems [6, 11, 3]. Among them, low-rank matrix recovery can be solved in polynomial-time and has been

shown to provide promising results [3]. While this type of approach is able to identify a set of representative basis given corrupted training data, there is no guarantee that such a basis set would serve for classification purposes.

Recently, sparse representation-based classification (SRC) [19] has shown very promising results on face recognition. It considers each test image as a sparse linear combination of the training samples by solving an ℓ_1-minimization problem. If the test image is corrupted, SRC exhibits excellent robustness to face occlusion and corruption. Since SRC requires the training images to be well aligned for reconstruction purposes, Wagner *et al*. [17] further extends it to deal with face misalignment and illumination variations. Yang *et al.* [20, 21] also propose 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 test images are corrupted.

In this paper, we address the problem of robust face recognition, in which *both* training and test image data are corrupted, and we do *not* have the prior knowledge on the type of corruptions. We will show that the direct use of dimension reduction techniques such as Eigenfaces for training and testing would degenerate the performance with the presence of corrupted data (see the left half of Figure 1 for example). To overcome this problem, we propose a novel low-rank matrix approximation with structural incoherence. While our method decomposes the raw face image data into a set of representative basis and a sparse error matrix, we regularize the structural incoherence of the derived representative basis. The introduction of such incoherence between the basis extracted from different classes would provide additional discriminating ability to our framework. It is worth noting that, we are among the first applying low-rank techniques for face recognition problems, and our proposed method is able to further improve the recognition performance, as illustrated in the right half of Figure 1.

The remaining of this paper is organized as follows. Section 2 reviews related works on low-rank matrix recovery and SRC for face recognition. In Section 3, we present our face recognition algorithm based on low-rank matrix decomposition and structural incoherence. Experimental results on real-world face image data are presented in Section 4. Finally, Section 5 concludes this paper.

2. Related Work

2.1. Robust PCA and Low-Rank Matrix Recovery

Principal component analysis (PCA) has been widely used for data analysis and dimension reduction. In spite of its effectiveness, PCA is known to be sensitive to sparse errors with large magnitudes [5]. Aim at designing a robust PCA model while suppressing the effect of such sparse noise, a number of approaches have been proposed in the literatures, including the introduction of influence functions [6], alternating minimization techniques [11], and low-rank matrix recovery [3] (noted as LR in the remaining for this paper for conciseness). Among these methods, LR has been observed to be solved in polynomial time with performance guarantees [3]. Since our proposed algorithm is based on low-rank decomposition techniques, it is necessary for us to briefly review its formulation.

Low-rank matrix recovery seeks to decompose a data matrix $\mathbf{D}$ into $\mathbf{A} + \mathbf{E}$, where $\mathbf{A}$ is a low-rank matrix and $\mathbf{E}$ is the associated sparse error. More precisely, given the input data matrix $\mathbf{D}$, LR minimizes the rank of matrix $\mathbf{A}$ while reducing $\|\mathbf{E}\|_0$ to derive the low-rank approximation of $\mathbf{D}$. Since the aforementioned optimization problem is NP-hard, Candès *et al.* [3] solve the following formulation to make the original LR tractable:

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

In (1), the nuclear norm $\|\mathbf{A}\|_*$ (i.e., the sum of the singular values) approximates the rank of $\mathbf{A}$, and the ℓ_0-norm $\|\mathbf{E}\|_0$ is replaced by the ℓ_1-norm $\|\mathbf{E}\|_1$, which sums up the absolute values of entries in $\mathbf{E}$. It is shown in [3] that, solving this convex relaxation version is equivalent to solving the original low-rank matrix approximation problem, as long as the rank of $\mathbf{A}$ to be recovered is not too large and the number of errors in $\mathbf{E}$ is small (sparse). To solve the optimization problem of (1), the technique of inexact augmented Lagrange multipliers (ALM) [3, 13] has been applied due to its computational efficiency.

2.2. Sparse Representation-based Classification

Recently, Wright *et al*. [19] proposed a sparse representation-based classification (SRC) algorithm for face recognition. SRC considers each test image as a sparse linear combination of training image data by solving an ℓ_1-minimization problem, and very promising results were reported in [19]. Several works have been proposed to further extend SRC for improved performance. For example, Yuan and Yan [22] utilized an $\ell_{1,2}$ mixed-norm regularization for computing the joint sparse representation of different features for visual signals. Jenatton *et al*. [10] utilized a tree-structured sparse regularization for hierarchical sparse coding. Chao *et al*. [4] integrated the $\ell_{1,2}$ norm with a data locality constraint for improved face recognition.

Since our classification rule is based on SRC, we now briefly review this algorithm for the sake of clarity. Suppose that there exist N_T training images from N object classes, and each class j has N_j images. Let $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_N] \in \mathbb{R}^{d \times N_T}$ be the training set, where $\mathbf{D}_j \in \mathbb{R}^{d \times N_j}$ contains training images of the jth class as its columns, and d is the dimension of each image. Given a test image $\mathbf{y} \in \mathbb{R}^{d \times 1}$, the SRC algorithm calculates the sparse representation $\boldsymbol{\alpha}$ of $\mathbf{y}$, which is computed via the

ℓ_1 minimization process over the entire training image set. More precisely, SRC solves

$$\min_{\boldsymbol{\alpha}} \|\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1. \quad (2)$$

Once (2) is solved, let $\boldsymbol{\alpha}_j \in \mathbb{R}^{N_j \times 1}$ be the entries of $\boldsymbol{\alpha}$ associated with class j, i.e., $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1; \boldsymbol{\alpha}_2; \ldots; \boldsymbol{\alpha}_N]$, the test input $\mathbf{y}$ will be recognized as class j if it satisfies

$$j^* = \arg\min_j \|\mathbf{y} - \mathbf{D}_j\boldsymbol{\alpha}_j\|_2^2. \quad (3)$$

In other words, the test image $\mathbf{y}$ will be assigned to the class with the lowest reconstruction error. This is because that, the test image $\mathbf{y}$ should lie in the space spanned by the columns $\mathbf{D}_j$ of class j. As a result, most non-zero elements of $\boldsymbol{\alpha}$ will mainly be presented in $\boldsymbol{\alpha}_j$, which results in the minimum reconstruction error. The framework of SRC is depicted by the red arrows in Figure 3.

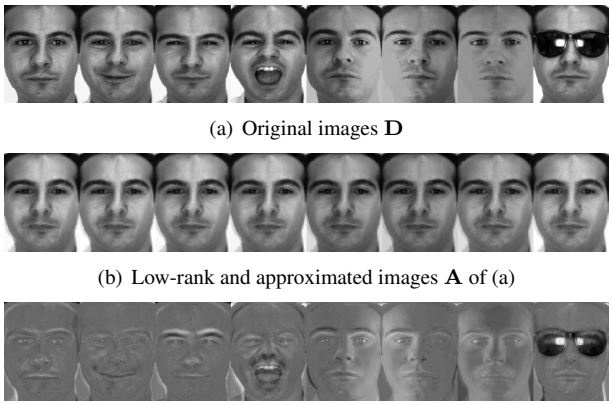
Although impressive face recognition results were reported by SRC [19], and it has been shown to be able to recognize occluded test images, SRC still requires *clean* (i.e., un-occluded) face images for training and thus might not be preferable for real-world scenarios. As later verified by our experiments, if corrupted training data is presented, SRC tends to recognize test images with the same type of corruption and thus results in poorer performance. In the following section, we will introduce our proposed method for robust face recognition, in which both training and test image data can be corrupted.

3. Low-Rank Matrix Recovery with Structural Incoherence

3.1. Face recognition by low-rank matrix recovery

For real-world face recognition problems, we cannot expect that the training image data can be always collected under a well-controlled setting. Besides illumination, pose, and expression variations, it is possible that one can be taking a scarf, gauze mask, or sunglasses when his/her face image is taken by the camera. When using such images for training, the learned face recognition algorithm might overfit the extreme noise of occlusion instead of modeling the face of the subject, and thus the performance will be degraded.

As discussed in Section 2.1, low-rank matrix recovery (LR) can be applied to alleviate the aforementioned problem by decomposing the collected data matrix into two different parts, one is a representative basis matrix of low rank and the other is the associated sparse error. It is worth noting that the data needs to be registered prior to the procedure of low-rank matrix decomposition, so that the extracted low-rank matrix would preserve the structure of the data (i.e., texture) and thus the corresponding error matrix will be sparse. When applying LR for face recognition with N subjects of interest, one can collect training data $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_N]$, where $\mathbf{D}_i$ is the training data matrix (with the presence of occlusion or disguise) for subject i, as shown in Figure 2(a). By performing low-rank matrix recovery, the data matrix data $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_N]$ will be decomposed into a low-rank matrix $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_N]$ and the sparse error matrix $\mathbf{E} = [\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_N]$. As shown in Figure 2(b), the error images in $\mathbf{A}$ can be considered as preprocessed data with sparse noise removed (see the corresponding images in Figure 2(c)). As a result, the low-rank matrix $\mathbf{A}$ has a better representative ability than the original data $\mathbf{D}$ does in describing the face images of the subject of interest.



(a) Original images $\mathbf{D}$

(b) Low-rank and approximated images $\mathbf{A}$ of (a)

(c) Sparse error images $\mathbf{E}$ of (a)

Figure 2. Example results of low-rank matrix recovery.

Since the face images are typically with high dimensionality, standard dimension reduction techniques such as PCA can be performed on the derived low-rank matrix $\mathbf{A}$. As a result, instead of using the Eigenfaces calculated by from the original data matrix $\mathbf{D}$, one can apply PCA on the low-rank matrix $\mathbf{A}$ (as shown in *Step 2* of Figure 3), and the resulting subspace can be applied as the dictionary for training and testing purposes (see *Step 3* in Figure 3). Finally, one can apply SRC and the derived dictionary to classify test inputs, which performs classification based on class-wise minimum reconstruction error (as depicted by *Step 4* in Figure 3). We will verify later that, compared with the direct use of raw data $\mathbf{D}$ for subspace and dictionary learning (as standard SRC does), LR better handles the problem in which the input training data is under severe illumination variations or is corrupted by occlusion or disguise. Algorithm 1 and Figure 3 sumamrize the procedure of integrating low-rank matrix recovery and SRC for face recognition.

3.2. Low-rank and structurally incoherent matrix decomposition

Although LR processes the original data $\mathbf{D}$ and produces a low-rank matrix $\mathbf{A}$ for better representation ability (with sparse noise removed), the face images from different subjects typically share common (correlated) features (e.g., the

Algorithm 1 LR for Face Recognition

Input: Training data $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_N]$ from N subjects and the test input $\mathbf{y}$

Step 1 : Perform LR on $\mathbf{D}$

(to be replaced by Algorithm 2 in Section 3.2)

for $i = 1 : N$ **do**

$\min_{\mathbf{A}_i, \mathbf{E}_i} \|\mathbf{A}_i\|_* + \lambda \|\mathbf{E}_i\|_1$ s.t. $\mathbf{D}_i = \mathbf{A}_i + \mathbf{E}_i$

end for

Step 2: Calculate principal components $\mathbb{W}$ of $\mathbf{A}$

$\mathbb{W} \leftarrow \mathcal{PCA}(\mathbf{A})$

Step 3: Project $\mathbf{D}$ and $\mathbf{y}$ onto $\mathbb{W}$

$\mathbf{D}_p = \mathbb{W}(\mathbf{D})$, and $\mathbf{y}_p = \mathbb{W}(\mathbf{y})$

Step 4: Perform SRC to classify $\mathbf{y}_p$

$\min_{\boldsymbol{\alpha}} \|\mathbf{y}_p - \mathbf{D}_p\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1$.

for $i = 1 : N$ **do**

$e(i) = \|\mathbf{y} - \mathbf{D}_{pi}\boldsymbol{\alpha}_i\|_2^2$

end for

Output: $y \leftarrow \arg\min_i e(i)$

locations of eyes, nose, etc.), and thus the derived matrix $\mathbf{A}$ might not contain sufficient discriminating information. Inspired by [15], we propose to promote the incoherence between different low-rank matrices. The introduction of such incoherence would prefer the resulting low-rank matrices to be as independent as possible. As a result, commonly shared features across difference classes will be suppressed while the independent/discriminating ones will be preserved. As illustrated in *Step 1* of Figure 3, our methods aims at providing additional discriminating ability to the original LR models by promoting their structural incoherence, and the recognition performance is expected to be improved.

Based on the LR formulation in (1), we add a regularization term to this objective function to enforce the incoherence between the low-rank matrices. We now solve the following optimization problem:

$$\min_{\mathbf{A},\mathbf{E}} \sum_{i=1}^{N} \{\|\mathbf{A}_i\|_* + \lambda \|\mathbf{E}_i\|_1\} + \eta \sum_{j \neq i} \left\|\mathbf{A}_j^T \mathbf{A}_i\right\|_F^2 \quad \text{s.t. } \mathbf{D}_i = \mathbf{A}_i + \mathbf{E}_i. \tag{4}$$

In (4), the first term performs the standard low-rank decomposition of the data matrix $\mathbf{D}$. The second term sums up the Frobenius norms between each pair of the low-rank matrices $\mathbf{A}_i$ and $\mathbf{A}_j$, which is penalized by the parameter η balancing the low-rank matrix approximation and matrix incoherence. We refer to (4) as low-rank matrix recovery with *structural incoherence*, aiming at providing improved discrimination ability to the original LR model. Since the error matrix $\mathbf{E}$ in (4) is sparse (the same as (1)) and represents extreme noise such as occlusion and disguise presented in face images, we do not enforce extra regularization on $\mathbf{E}$.

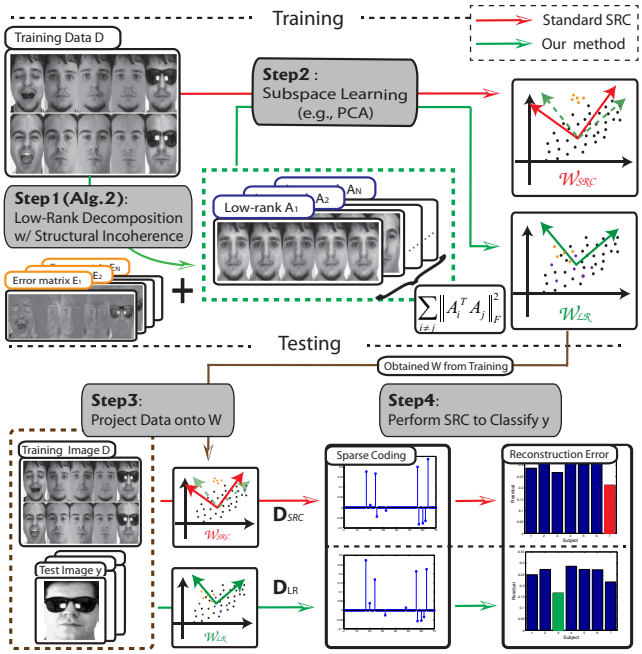


Figure 3. Illustration of our proposed method. Note that we promote the structural incoherence between low-rank matrices for better modeling and recognizing face images.

Instead of solving (4) directly, we solve the following class-wise optimization problem across different classes:

$$\min_{\mathbf{A}_i,\mathbf{E}_i} \|\mathbf{A}_i\|_* + \lambda \|\mathbf{E}_i\|_1 + \eta \sum_{j \neq i} \left\|\mathbf{A}_j^T \mathbf{A}_i\right\|_F^2 \quad \text{s.t. } \mathbf{D}_i = \mathbf{A}_i + \mathbf{E}_i. \tag{5}$$

In other words, we iteratively solve the above formulation for each class until the derived low-rank matrices converge. We note that, however, the above optimization problem is involved with Frobenius norms of different matrix pairs. To make the above problem more tractable, we advance the property that $\left\|\mathbf{A}_j^T \mathbf{A}_i\right\|_F^2 \leq \|\mathbf{A}_j\|_F^2 \|\mathbf{A}_i\|_F^2$ and relax (5) into the following formulation:

$$\min_{\mathbf{A}_i,\mathbf{E}_i} \|\mathbf{A}_i\|_* + \lambda \|\mathbf{E}_i\|_1 + \eta' \|\mathbf{A}_i\|_F^2 \quad \text{s.t. } \mathbf{D}_i = \mathbf{A}_i + \mathbf{E}_i. \tag{6}$$

where $\eta' = \eta \sum_{j \neq i} \|\mathbf{A}_j\|_F^2$ is a constant when deriving $\mathbf{A}_i$ and $\mathbf{E}_i$. We note that, from the above derivation, solving (6) will address low-rank matrix approximation with implication of structural incoherence between the derived low-rank matrices. In our proposed method, we choose to iteratively solve (6) for each class, as we discuss later in the following subsection and Algorithm 2.

3.3. Optimization via ALM

Augmented Lagrange multipliers (ALM) have been applied to solve the standard LR problem [3, 13]. In this sub-

section, we will detail how we extend ALM to solve our proposed LR formulation with regularization on structural incoherence.

For an optimization problem in which $f(\mathbf{X})$ is to be minimized with the constraint $h(\mathbf{X}) = \mathbf{0}$, its ALM function can be formulated as follows:

$$L(\mathbf{X}, \mathbf{Y}, \mu) = f(\mathbf{X}) + \langle \mathbf{Y}, h(\mathbf{X}) \rangle + \frac{\mu}{2} \|h(\mathbf{X})\|_F^2, \quad (7)$$

where each element of $\mathbf{Y}$ indicates a Lagrange multiplier. Let $\mathbf{X} = (\mathbf{A}_i, \mathbf{E}_i)$ in (7), we redefine $f(\mathbf{X})$ and $h(\mathbf{X})$ as

$$f(\mathbf{X}) = \|\mathbf{A}_i\|_* + \lambda \|\mathbf{E}_i\|_1 + \eta' \|\mathbf{A}_i\|_F^2 \text{ and}$$
$$h(\mathbf{X}) = \mathbf{D}_i - \mathbf{A}_i - \mathbf{E}_i.$$

As a result, our proposed LR problem in (6) can be reformulated as follows:

$$\begin{aligned} L(\mathbf{A}_i, \mathbf{E}_i, \mathbf{Y}_i, \mu, \eta') \quad = \quad & \|\mathbf{A}_i\|_* + \lambda \|\mathbf{E}_i\|_1 + \eta' \|\mathbf{A}_i\|_F^2 \\ & + \langle \mathbf{Y}_i, \mathbf{D}_i - \mathbf{A}_i - \mathbf{E}_i \rangle \quad (8) \\ & + \frac{\mu}{2} \|\mathbf{D}_i - \mathbf{A}_i - \mathbf{E}_i\|_F^2 \end{aligned}$$

To solve (8), we search for the optimal $\mathbf{A}_i$, $\mathbf{E}_i$, and $\mathbf{Y}$ iteratively. The pseudo code of our proposed algorithm is shown in Algorithm 2. We now discuss how we update/solve the above variables in each iteration.

Algorithm 2 Solving LR with Structural Incoherence

Input: Data matrix $\mathbf{D}$ and parameters η and ρ ($\rho > 1$)
  Use Step1 in Alg. 1 to initialize $\mathbf{A}^0$, $\mathbf{E}^0$, $\mathbf{Y}^0$, $\mu^0 > 0$
  for $i = 1$ **to** N **do**
    $\eta' \leftarrow \eta \sum_{j \neq i} \|\mathbf{A}_j\|_F^2$
    while not converged **do**
      $\mathbf{A}_i^{k+1} = \arg\min_{\mathbf{A}_i} L(\mathbf{A}_i, \mathbf{E}_i^k, \mathbf{Y}_i^k, \mu^k, \eta')$
      $\mathbf{E}_i^{k+1} = \arg\min_{\mathbf{E}_i} L(\mathbf{A}_i^{k+1}, \mathbf{E}_i, \mathbf{Y}_i^k, \mu^k, \eta')$
      $\mathbf{Y}_i^{k+1} = \mathbf{Y}_i^k + \mu^k(\mathbf{D}_i - \mathbf{A}_i^{k+1} - \mathbf{E}_i^{k+1})$
      $\mu^{k+1} = \rho\, \mu^k$
    end while
  end for
Output: $\mathbf{A}$ and $\mathbf{E}$

3.3.1 Updating $\mathbf{A}_i$

To update $\mathbf{A}_i^{k+1}$ for class i at the $(k+1)$*th* iteration in Algorithm 2, we have fixed $\mathbf{E}_i$ and $\mathbf{Y}_i$ and solve the following problem accordingly:

$$\begin{aligned} \mathbf{A}_i^{k+1} &= \arg\min_{\mathbf{A}_i} L(\mathbf{A}_i, \mathbf{E}_i^k, \mathbf{Y}_i^k, \mu^k, \eta') \\ &= \arg\min_{\mathbf{A}_i} \|\mathbf{A}_i\|_* + \eta' \|\mathbf{A}_i\|_F^2 + \langle \mathbf{Y}_i^k, \mathbf{D}_i - \mathbf{A}_i - \mathbf{E}_i^k \rangle \\ &\qquad + \frac{\mu^k}{2} \left\| \mathbf{D}_i - \mathbf{A}_i - \mathbf{E}_i^k \right\|_F^2 \\ &= \arg\min_{\mathbf{A}_i} \|\mathbf{A}_i\|_* + (\eta' + \frac{\mu^k}{2}) \langle \mathbf{A}_i, \mathbf{A}_i \rangle \\ &\qquad - \mu^k \langle \mathbf{D}_i - \mathbf{E}_i^k + (1/\mu)^k \mathbf{Y}_i^k, \mathbf{A}_i \rangle \\ &= \arg\min_{\mathbf{A}_i} \epsilon \|\mathbf{A}_i\|_* + \frac{1}{2} \|\mathbf{X}_a - \mathbf{A}_i\|_F^2, \end{aligned}$$

where $\epsilon = (2\eta' + \mu^k)^{-1}$ and $\mathbf{X}_a = \mu^k \epsilon (\mathbf{D}_i - \mathbf{E}_i + (1/\mu^k)\mathbf{Y}_i)$. As suggested by [2], the solution to the above problem can be solved as

$$\begin{aligned} & \mathbf{A}_i^{k+1} = \mathbf{U}\mathbf{S}_\epsilon\mathbf{V}^T = \mathbf{U}T_\epsilon[\mathbf{S}]\mathbf{V}^T \\ & \text{where } (\mathbf{U}, \mathbf{S}, \mathbf{V}^T) = \mathcal{SVD}(\mathbf{X}_a). \end{aligned} \quad (9)$$

Note that $\mathbf{S}$ is the singular value matrix of $\mathbf{X}_a$. The operator $T_\epsilon[\mathbf{S}]$ in (9) is defined by element-wise ϵ thresholding of $\mathbf{S}$, i.e., $diag(T_\epsilon[\mathbf{S}]) = [t\epsilon[s_1], t\epsilon[s_2], \ldots, t\epsilon[s_r]]$ for rank($\mathbf{S}$) = r, and each $t\epsilon[s]$ is determined as

$$t_\epsilon[s] = \begin{cases} s - \epsilon, & if\ s > \epsilon, \\ s + \epsilon, & if\ s < -\epsilon, \\ 0, & otherwise. \end{cases} \quad (10)$$

3.3.2 Updating $\mathbf{E}_i$

To update the error matrix $\mathbf{E}_i$ for class i, we derive (8) with fixed $\mathbf{A}_i$ and $\mathbf{Y}_i$ and obtain the following form:

$$\begin{aligned} \mathbf{E}_i^{k+1} &= \arg\min_{\mathbf{E}_i} L(\mathbf{A}_i^{k+1}, \mathbf{E}_i, \mathbf{Y}_i^k, \mu^k, \eta') \\ &= \arg\min_{\mathbf{E}_i} \|\mathbf{E}_i\|_1 + \langle \mathbf{Y}_i^k, \mathbf{A}_i^{k+1} + \mathbf{E}_i - \mathbf{D}_i \rangle \\ &\qquad + \frac{\mu^k}{2} \left\| \mathbf{A}_i^{k+1} + \mathbf{E}_i - \mathbf{D}_i \right\|_F^2 \\ &= \arg\min_{\mathbf{E}_i} \epsilon' \|\mathbf{E}_i\|_1 + \frac{1}{2} \|\mathbf{X}_e - \mathbf{E}_i\|_F^2. \end{aligned}$$

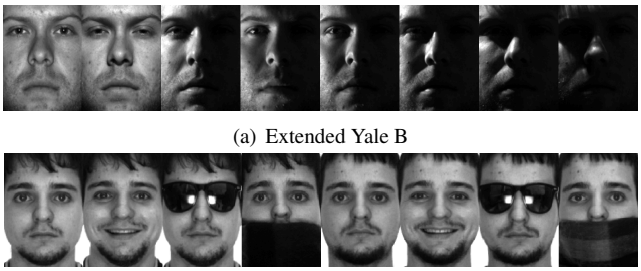
where $\epsilon' = (1/\mu^k)$ and $\mathbf{X}_e = \mathbf{D}_i - \mathbf{A}_i^{k+1} + (1/\mu^k)\mathbf{Y}_i^k$. The above optimization problem can be solved by ℓ_1-minimization techniques such as [8].

Once both $\mathbf{A}_i$ and $\mathbf{E}_i$ are obtained, the matrix $\mathbf{Y}_i$ can be simply updated by the last equation in Algorithm 2. The convergence of the three matrices indicates the termination of the optimization process for our proposed LR algorithm.

4. Experiments

4.1. Extended Yale B Database

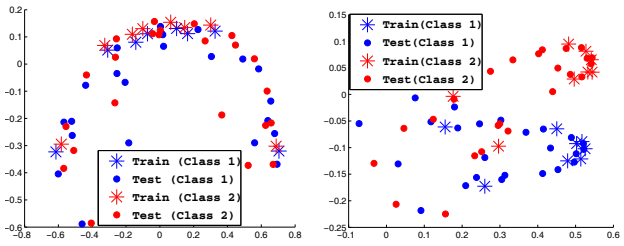
The extended Yale B database [7] consists of 2,414 frontal-face images of 38 subjects (around 59-64 images for



(a) Extended Yale B

(b) AR face databases

Figure 4. Example training and test images for our experiments.



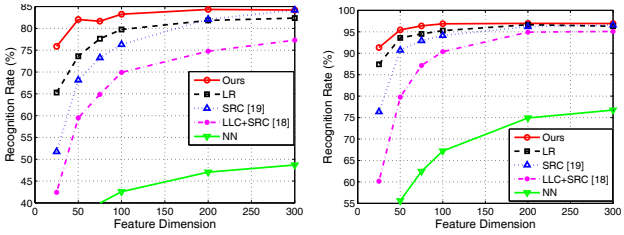
(a) (b)

Figure 5. Data distributions for 2 classes (in blue and red colors)ones). The 2D subspace is spanned by the first two eigenvectors derived by (a) the original data matrix $\mathbf{D}$ and (b) the LR matrices $\mathbf{A}$ with structural incoherence. Note that training and test instances are denoted as ($*$) and ($\bullet$), respectively.

each person). The cropped and normalized face images are of size $192\times168 = 32{,}256$ pixels, while each image is taken under various laboratory-controlled lighting conditions (see Figure 4(a) for example) [12]. Besides the standard LR (without structural incoherence) and our proposed method, we also consider nearest neighbor (NN), SRC [19], and LLC [18] for comparisons. Note that LLC is an extended version of SRC exploiting data locality for improved sparse coding, and the classification rule is also based on (3). To evaluate our recognition performance using data with different dimensions, we project the data onto the eigenspace derived PCA using our LR models (as shown in Figure 3). For the standard LR approach, the eigenspace spanned by LR matrices without structural incoherence is considered, while those of other SRC-based methods are derived by the data matrix $\mathbf{D}$ directly. We vary the dimension of the eigenspace and compare the results in this section.

We first visualize the effect of our proposed LR method with structural incoherence on the projected data. Figure 5(a) shows the distributions of training and test data from two classes, which are projected onto the first two eigenvectors determined by the data matrix $\mathbf{D}$ (as NN and SRC-based approaches do). On the other hand, we project the same data onto the subspace derived by our low-rank matrices, as shown in Figure 5(b). It is clear that the separation between the two classes (in red and blue colors) is significantly improved. More importantly, we see that the training data points (denoted as ($*$)) within the same class become closer to each other, while the separation between those from different classes becomes larger. This observation is consistent with our expectation that promoting structural incoherence on low-rank matrices will result in improved classification.

We first randomly select eight images for training and the remaining for test (per person). We vary the dimension of the eigenspace as 25, 50, 75, 100, 200, and 300 to compare the recognition performance between different methods. Since the total number of training image is $8 \times 38 = 304$, we do not consider higher dimensional space for evaluation. All experiments run ten times and the average results are shown in Figure 6(a). It is clear that, while the two LR methods consistently produced higher recognition rates than other NN and SRC-based approaches did, our proposed LR method was the best among all. For example, at dimension = 50, our method achieved a high recognition rate at 82.1%, and those for LR, SRC, LLC, and NN were 73.6%, 68.2%, 59.5%, and 32.5%, respectively (see Figure 6(a)). We repeat the above experiments using thirty-two training images per person. We compare the performance of different approaches in Figure 6(b) and observe the same advantages using our proposed method. From these empirical results, we confirm the use of our LR method alleviates the problem of severe illumination variations even when such noise is presented in both training and test data. And, due to the enforcement of structural incoherence between our LR matrices, our method exhibits additional classification capability and thus outperforms the standard LR algorithm. In the following subsection, in which sparse noise such as occlusion is presented in face images, we expect the improvement using our method will be more significant.

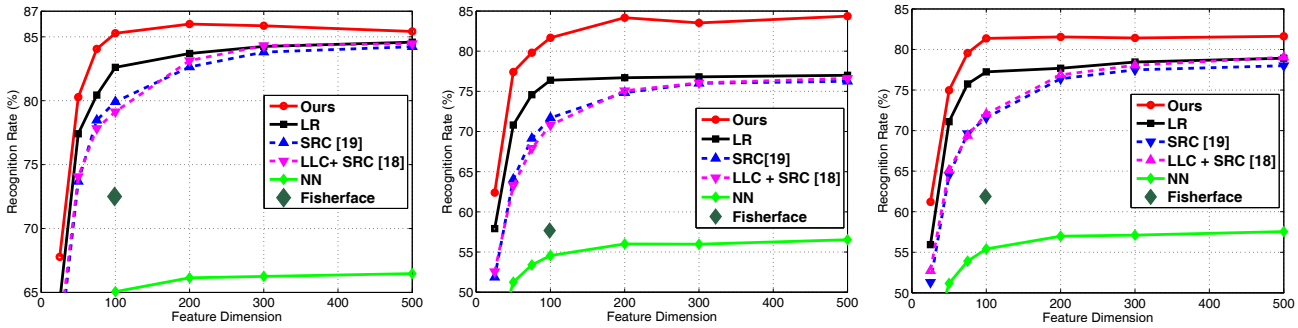


(a) $N = 8$ (b) $N = 32$

Figure 6. Performance comparisons on the Extended Yale B database with different numbers N of training images per person.

4.2. AR Database

The AR database [14] contains over 4,000 frontal images for 126 individuals. There are 26 face images available for each person, and the images are taken under different variations, including illumination, expression, and facial occlusion/disguise in two separate sessions. More specifically,



(a) **Sunglasses** (b) **Scarf** (c) **Sunglasses+Scarf**

Figure 7. Performance comparisons on AR database for different scenarios.

there are thirteen images for each session, in which three images with sunglasses, another three with scarfs, and the remaining seven are simply with illumination and expressions variations (and thus are considered as clean/neutral images). All images are with $165 \times 120 = 19800$ pixels and converted to gray scale (see Figure 4(b) for example). We note that, most prior works using this database only considered the use of neutral images for training. In our experiments, we choose a subset of the AR database consisting of 50 men and 50 women (as [19] did), and *both* neutral and corrupted images taken at session 1 (of a portion of it) are used for training. We consider the following scenarios:

Sunglasses: We consider corrupted training images due to the occlusion of sunglasses. We use seven neutral images plus one image with sunglasses (randomly chosen) at session 1 for training, and the remaining neutral images (all from session 2) plus the rest of the images with sunglasses (two taken at session 1 and three at session 2) for testing. In other words, we have a total of eight training images and twelve test images per person. Note that the presence of sunglasses occludes about 20% of the face image.

Scarf: We consider the training images are corrupted by disguise due to scarf, which occludes about 40% of the face image. We apply a similar training/test set choice, and have a total of eight training images (seven neutral plus one randomly selected image with scarf at session 1) and twelve test images (seven neutral images plus five images with scarfs) per person for this scenario.

Sunglasses+Scarf: Finally, we consider the case where images with sunglasses and scarfs are presented during training. We choose all seven neutral images at session 1 and two corrupted images (one with sunglasses and the other with scarf) for training. A total of nineteen test images (seven neutral images at session 2 plus the remaining ten occluded images) are available for this case.

Since there are at most three occluded images for each type of corruption available in session 1, we repeat our experiment for each scenario three times (i.e., randomly select one corrupted image with the remaining neural ones for training), and we report the averaged performance. Similar to the experiments on the Extended Yale B database, we vary the dimension of the face data from 25 up to 500, and we compare our method with the approaches of LR, NN, Fisherface, SRC, and LLC. Figures 7(a) and 7(b) show recognition results of scenarios **Sunglasses** and **Scarf**. From these two figures, we see that our method outperforms all other approaches across different dimensions. It is worth noting that, although Fisherfaces [1] also promote the separation between classes during its learning process, it did not achieve comparable performance as we did. For example, the recognition rates of Fisherfaces for **Sunglasses** and **Scarf** were only 72.5% and 57.7% at dimension $d = 99$, respectively. With the increase of occlusion (from sunglasses to scarf), it is observed that the recognition rate of Fisherfaces is severely degraded. This is because its direct use of corrupted training image data, and thus the associated performance is remarkably degraded due to the presence of occlusion and disguise. We achieved over 80% recognition rates at a comparable dimension at 100 for both cases. As for the last scenario in which the training data are corrupted by sunglasses and scarfs, we again confirm the robustness of our proposed method by performance comparisons shown in Figure 7(c).

Table 1 summarizes the performance comparisons with different approaches under three different scenarios. At lower dimensions, our approach significantly outperforms other baseline and state-of-the-art methods, especially when the percentage of occlusion increases. For example, when the data dimension is equal to 50, we achieve recognition rates at 77.41% and 80.27% for the scenarios of **Scarf** (40% occlusion) and **Sunglasses** (20% occlusion), respectively. Using LR, which is among the state-of-the-art and the most relevant method to ours, it obtains 70.81% and 77.41% for the above two cases. In other words, we improve the method of LR by about 3 to 7%, depending on the percentage of occlusion. We observe the same conclusion when larger dimensionality is of interest, and we still obtain comparable improvements over different scenarios. From both Figure 7 and Table 1, we successfully verify the effectiveness and robustness of our proposed method.

Table 1. Comparisons of recognition rates between our and other face recognition methods. (* : the dimension of Fisherfases is fixed at $N - 1 = 100 - 1 = 99$, where N is the number of subjects in AR database).

Methods	Dimension = 500			Dimension = 100			Dimension = 50		
	Sunglasses	Scarf	Sunglasses +Scarf	Sunglasses	Scarf	Sunglasses +Scarf	Sunglasses	Scarf	Sunglasses +Scarf
Fisherfaces [1]*	$--$	$--$	$--$	72.50	57.67	61.80	$--$	$--$	$--$
NN	66.47	56.53	57.55	65.06	54.56	55.41	60.89	51.25	51.15
LLC+SRC [18]	84.47	76.61	79.03	79.14	70.08	72.04	74.06	63.25	65.10
SRC [19]	84.22	76.25	78.00	79.92	71.70	71.59	73.68	64.05	64.51
LR	84.58	77.00	78.92	82.61	76.39	77.23	77.41	70.81	71.10
Ours	**85.42**	**84.36**	**81.62**	**85.27**	**81.67**	**81.37**	**80.27**	**77.41**	**74.96**

5. Conclusions

We presented a low-rank matrix approximation algorithm with structural incoherence for robust face recognition. The introduction of structural incoherence between low-rank matrices promotes the discrimination between different classes, and thus the associated models exhibit excellent discriminating ability. We showed that the proposed optimization problem can be easily solved by advancing augmented Lagrange multipliers. Our experiments confirmed that our proposed LR approach is robust to severe illumination variations or corruptions such as occlusion and disguise, while our method has been shown to outperform state-of-the-art face recognition algorithms.

Acknowledgement

This work is supported in part by the National Science Council of Taiwan via NSC101-2631-H-001-007 and NSC100-2221-E-001-018-MY2.

References

[1] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *PAMI*, 19(7):711–720, 1997.

[2] J. Cai, E. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal of Optimization*, 20(4):1956–1982, 2010.

[3] E. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM*, 58, 2011.

[4] Y.-W. Chao, Y.-R. Yeh, Y.-W. Chen, Y.-J. Lee, and Y.-C. F. Wang. Locality-constrained group sparse representation for robust face recognition. In *ICIP*, 2011.

[5] F. De la Torre and M. Black. Robust principal component analysis for computer vision. In *ICCV*, 2001.

[6] F. De la Torre and M. Black. A framework for robust subspace learning. *International Journal of Computer Vision*, 54(1):117–142, 2003.

[7] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *PAMI*, 23(6):643–660, 2001.

[8] E. T. Hale, W. Yin, and Y. Zhang. Fixed-point continuation for ℓ_1-minimization: Methodology and convergence. *SIAM Journal on Optimization*, 19:1107–1130, 2008.

[9] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang. Face recognition using Laplacianfaces. *PAMI*, 2005.

[10] R. Jenatton, J. Mairal, G. Obozinski, and F. Bach. Proximal methods for hierarchical sparse coding. *Journal of Machine Learning Research*, 12:2297–2334, 2011.

[11] Q. Ke and T. Kanade. Robust L_1 norm factorization in the presence of outliers and missing data by alternative convex programming. In *CVPR*, pages 739–746, 2005.

[12] K.-C. Lee, J. Ho, and D. J. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *PAMI*, 27(5):684–698, 2005.

[13] Z. Lin, M. Chen, and Y. Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *UIUC Tech. Rep. UILU-ENG-09-2214*, 2010.

[14] A. Martinez and R. Benavente. The AR face database. *CVC Technical Report*, 24, 1998.

[15] I. Ramirez, P. Sprechmann, and G. Sapiro. Classification and clustering via dictionary learning with structured incoherence and shared features. In *CVPR*, pages 3501–3508, 2010.

[16] M. Turk and A. Pentland. Face recognition using Eigenfaces. In *CVPR*, pages 586–591, 1991.

[17] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, and Y. Ma. Towards a practical face recognition system: Robust registration and illumination by sparse representation. In *CVPR*, pages 597–604, 2009.

[18] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-constrained linear coding for image classification. In *CVPR*, pages 3360–3367, 2010.

[19] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *PAMI*, 31(2):210–227, 2009.

[20] M. Yang and L. Zhang. Gabor feature based sparse representation for face recognition with Gabor occlusion dictionary. In *ECCV*, pages 448–461, 2010.

[21] M. Yang, L. Zhang, J. Yang, and D. Zhang. Robust sparse coding for face recognition. In *CVPR*, 2011.

[22] X. Yuan and S. Yan. Visual classification with multi-task joint sparse representation. In *CVPR*, 2010.