Low-Rank Laplacian-Uniform Mixed Model for Robust Face Recognition

Jiayu Dong, Huicheng Zheng, Lina Lian

School of Data and Computer Science, Sun Yat-sen University

Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, China

Guangdong Key Laboratory of Information Security Technology

Email: zhenghch@mail.sysu.edu.cn

Abstract

Sparse representation based methods have successfully put forward a general framework for robust face recognition through linear reconstruction and sparsity constraints. However, residual modeling in existing works is not yet robust enough when dealing with dense noise. In this paper, we aim at recognizing identities from faces with varying levels of noises of various forms such as occlusion, pixel corruption, or disguise, and take improving the fitting ability of the error model as the key to addressing this problem. To fully capture the characteristics of different noises, we propose a mixed model combining robust sparsity constraint and low-rank constraint, which can deal with random errors and structured errors simultaneously. For random noises such as pixel corruption, we adopt a Laplacian-uniform mixed function for fitting the error distribution. For structured errors like continuous occlusion or disguise, we utilize robust nuclear norm to constrain the rank of the error matrix. An effective iterative reweighted algorithm is then developed to solve the proposed model. Comprehensive experiments were conducted on several benchmark databases for robust face recognition, and the overall results demonstrate that our model is most robust against various kinds of noises, when compared with state-of-the-art methods.

1. Introduction

As one of the most important problems in computer vision, face recognition has claimed its importance in identity verification and security surveillance. Numerous face recognition methods have been proposed during the years. At the early beginning, geometric information has been used for face classification [3]. Then robust appearance features such as LBP [41, 1], SIFT [4], and Gabor features [29, 39] have been put forward successively. The geometric features or appearance features of facial images are concerned with the contents of the outer appearance, while ignoring the intrinsic attributes of the face. Subspace analy-sis [9, 18, 3] can effectively excavate the intrinsic characteristics of an image by mapping the original image to a lower dimension space. However, when the face image contains occlusion or illumination changes, the subspace projection methods can easily learn non-face feature components, especially when facing a small sample problem.

Deep learning has been applied widely in recent years due to its convenience and learning ability [14, 21]. Compared to early shallow back-propagation models, deeper models, such as DeepID2+ [30], DeepFace [31], VGG-Face [27], and FaceNet [28], demonstrate largely improved performance on face recognition. Although these methods have achieved great success in practical applications, existing deep neural networks often rely on a large number of training samples, and are not robust to inputs with anomalous distributions.

Sparse coding has achieved great performance in many fields, such as feature extraction, signal processing, image denoising and pattern recognition. Its core idea is to represent a test sample as a sparse linear combination of training samples, so that the smallest reconstruction error can be achieved using only those training samples of the same class as the test sample [35]. Usually, a test face image is denoted by a column vector $\boldsymbol{y}$ in sparse representation. Let $D \in \mathbb{R}^{m \times n}$ be a dictionary composed of n samples, each column of which is a training face sample. The general objective of sparse coding can be formulated as

$$\min_{\boldsymbol{x}} \sum_j \rho(e_j) + \lambda\|\boldsymbol{x}\|_1 \quad \text{s.t.} \quad \boldsymbol{y} - D\boldsymbol{x} = \boldsymbol{e} \quad (1)$$

where e_j denotes the j-th component in the residual $\boldsymbol{e}$. The objective is composed of two parts, i.e., the penalty for the reconstruction errors and the sparsity constraint for the representation.

Among numerous face recognition methods, sparse representation based methods have achieved great robustness against various types of noises. Most of these methods focus on finding reasonable penalty for the reconstruction residual. Many experimental results show that sparse repre-

sentation based methods have good robustness for dealing with external interference, such as occlusion and illumination change [26, 19, 33]. By exploring the intrinsic characteristics of the reconstruction errors, researchers have proposed a number of more robust algorithms. Some focus on the statistical analysis and others pay more attention to the structural information.

However, the error image in more challenging situations could be a complex integration that includes structured and non-structured parts at the same time. In this paper, we aim to provide a solution for face recognition with such complicated errors, and a low-rank Laplacian-uniform mixed (LR-LUM) model is proposed to tackle this problem. The main contributions of this paper can be summarized as follows:

1. By inspecting the compositions and characteristics of complex dense noises, we propose an LR-LUM model which integrates both sparsity and low-rank constraints for handling various types of noises simultaneously.

2. An iterative algorithm is developed to effectively excavate the discriminative ability and robustness of the proposed model.

3. We evaluate the proposed joint model on several representative face databases under various settings. The overall experimental results demonstrate that our method achieves the best robustness under many complex circumstances.

2. Related Work

Since sparse representation classification (SRC) was proposed and successfully applied to robust face recognition [35], many extended methods have emerged and made their contributions. ESRC [7] extended SRC by constructing an auxiliary dictionary to describe intra-class variations and improve generalization. Deng *et al*. further developed SLRC [8], which represents a test image as a superposition of class centroids and intra-class differences. To improve the robustness, He *et al*. [12] proposed maximum correntropy criterion. In the later work [13], they tried to integrate the error detection and error correction mechanism into a unified additive or multiplicative half-quadratic (HQ) framework. Inspired by robust regression theory, Yang *et al*. proposed robust sparse coding (RSC) [37], looking for a maximum likelihood estimation (MLE) solution for sparse representation. In a following work, they proposed regularized robust coding (RRC) [40]. Assume that reconstruction errors e_j and coding coefficients x_i are i.i.d with probability density functions $f(e_j)$ and $h(x_i)$, respectively. Let $\rho(e_j) = -\ln f(e_j)$ and $\pi(e_j) = -\ln h(e_j)$. RRC employs local quadratic approximation to minimize $\sum_{j=1}^{m} \rho(y_j - D_j \boldsymbol{x}) + \sum_{i=1}^{n} \pi(x_i)$ (D_j is the j-th row in

D) and leads to an iteratively reweighted least square solution, in which the weights are calculated by

$$w_j^t = \rho'(y_j - D_j \boldsymbol{x}^t)/(y_j - D_j \boldsymbol{x}^t) \qquad (2)$$

where ρ' is the first derivative of ρ. The minimization of such an objective function finally becomes ℓ_2- or ℓ_1-norm regularized problem, denoted as RRC-L2 or RRC-L1.

To address partial occlusion, Liao *et al*. [23] proposed an alignment-free approach based on multi-task multi-keypoint feature descriptor. In [25, 35], block-wise sparse representation has been demonstrated to be robust for occlusion. Jia *et al*. [19] attempted to recover the occluded part by the statistical law discovered from the training set.

The above methods stretch a two-dimensional face image into a vector, which obviously ignores the structured information of error images in the spatial domain useful for further improving the robustness. Therefore, some researchers proposed to exploit the structured information of error images to enhance the ability of models for error description, by using, e.g., Markov random fields and structured sparsity. Inspired by structured sparsity theory, Jia *et al*. [20] introduced structural information into SRC and proposed the structured sparse representation classifier (SS-RC). Face images suffering from continuous occlusion or illumination change would result in approximately low-rank error images through sparse reconstruction, as Zhang *et al*. revealed in [42]. Thus the rank function is used to restrain the structured characteristic of error images, as follows,

$$\min_{\boldsymbol{x}} \operatorname{rank}(E) \quad \text{s.t.} \quad E = T_M(\boldsymbol{y} - D\boldsymbol{x}) \qquad (3)$$

where $T_M(\cdot)$ transforms a vector back into a 2D image. Different from (1), the error image is represented as a matrix here for rank computation. But rank minimization is an NP-hard problem difficult to be optimized directly. Nuclear-norm based matrix regression (NMR) [38] was proposed to approximate a rank function and achieved great performance. The nuclear norm of an input matrix, denoted as $||\cdot||_*$, is equal to the sum of all singular values of the matrix. Compared to other methods relying on the structured information of error images, NMR is more concise and effective, which has attracted much attention in recent years. However, nuclear norm is sensitive to large singular values, which may increase the instability for face classification. Xie *et al*. [36] proposed a robust NMR, which takes a weighted form to enhance the approximation ability and guarantee an optimal solution. The robust NMR can be formulated as

$$\min_{\boldsymbol{x}} ||T_M(\boldsymbol{y} - D\boldsymbol{x})||_{w,*} = \sum_i \rho(d_i) \cdot d_i \qquad (4)$$

where d_i is the i-th largest singular value of $E = T_M(\boldsymbol{y} - D\boldsymbol{x})$. A weight function $\rho(d) = d^{-\kappa}$ is employed with a regulating factor $\kappa \geq 0$. When $\kappa = 0$, $||E||_{w,*}$ is equal to the regular nuclear-norm constraint $||E||_*$.

Figure 1. An example of a contaminated face image.

3. Our Proposed Method

3.1. Low-Rank Laplacian-Uniform Mixed Model

For robust face recognition, we aim to develop a method capable of recognizing the identity of a face image contaminated by dense noises as shown in Fig. 1. Although dense noises in real scenarios could be complex combinations induced by different sources, they can be generally categorized into two types according to the characteristics, *i.e.*, unpredictable random noises such as random pixel corruption and structured noises such as continuous occlusion. To handle errors induced by these two types of noises simultaneously, we construct an objective function as follows,

$$J(\boldsymbol{x}) = \alpha C_1(\boldsymbol{e}_1) + (1-\alpha)C_2(\boldsymbol{e}_2) + \beta\|\boldsymbol{x}\|_1$$
$$\text{s.t.} \quad \boldsymbol{y} - D\boldsymbol{x} = \boldsymbol{e}_1 + \boldsymbol{e}_2 \tag{5}$$

where $C_1(\cdot)$ and $C_2(\cdot)$ are penalty functions associated with these two types of errors, respectively. For the structured errors, *e.g.*, that corresponding to the sunglasses in Fig. 1, we can easily observe that the error matrix exhibits a low-rank structure in comparison to the matrix size. We utilize the robust nuclear norm proposed in [36] as a close relaxation for the rank function, *i.e.*, $C_1(\cdot)$ is a robust nuclear norm regularization term.

As for the random errors, the most significant characteristic that we can exploit is the statistical distribution. Specifically, different assumptions about the error distribution may lead to different constraints on the random errors. Assume that random errors $e_2^{(i)}$ (i-th element in $\boldsymbol{e}_2$) are i.i.d. with a probability density $f(e_2^{(i)})$, and let $\rho(e_2^{(i)}) = -\ln(f(e_2^{(i)}))$. The MLE for $\boldsymbol{e}_2$ would be

$$C_2(\boldsymbol{e}_2) = \sum_i \rho\left(e_2^{(i)}\right) \tag{6}$$

In order to better describe random errors in challenging situations, we adopt a Laplacian-uniform mixed (LUM) function for fitting the error distribution, which can be expressed as

$$f(e_2^{(i)}) = \alpha\left(\exp(-|e_2^{(i)}|/b) + c\right) \tag{7}$$

where $b > 0$ corresponds to the scale of the Laplacian component and $c > 0$ corresponds to a uniform distribution. $\alpha > 0$ is a distribution normalization factor.

The rationale behind this mixed distribution model comes from a more in-depth analysis of the causes of reconstruction residuals. First, no matter how well a test face can be reconstructed from the training set, there are bound to be minor position shift, expression variations, and even illumination changes between the original face and the reconstruction, which causes small coding residuals that take place at random positions. Second, in worse conditions, pixels at random positions might be corrupted to unpredictable values with unknown characteristics. The LUM model can appropriately fit such situation, with a Laplacian function addressing the first case and a uniform distribution addressing the latter case.

For a more intuitive demonstration, we computed the reconstruction residuals of an example face with 50% pixel corruption. The empirical distribution curve of reconstruction residuals and other fitted distribution curves are shown in Fig. 2. The peak around zero comes from inaccurate reconstruction on face position, pose, expression, *etc.*, as discussed above. A higher peak around zero is fitted by the LUM function for inclining to inlier pixels with small residuals. And the long tail is caused by randomly corrupted outliers. We can see that the LUM function is able to describe both characteristics well.
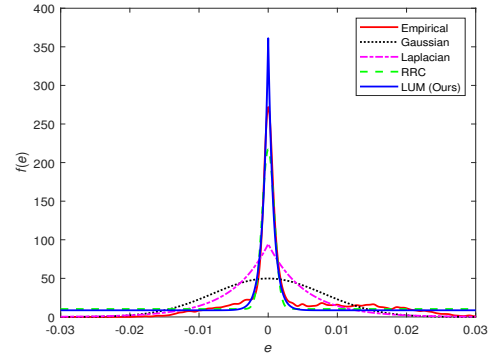

Figure 2. The empirical distribution of the reconstruction errors and different fitted distributions.

In order to optimize the $C_2(\boldsymbol{e}_2)$ term with the LUM prior, we need to make a relaxation by preserving the first order Taylor expansion of $\rho(e_2^{(i)})$, at the current estimation $e_2^{(i),t}$,

$$\rho(e_2^{(i)}) = \rho(|e_2^{(i),t}|) + \rho'(|e_2^{(i),t}|)(|e_2^{(i)}| - |e_2^{(i),t}|) \tag{8}$$

Inserting (8) into (6), and removing constant terms, the C_2 term becomes a reweighted ℓ_1-norm constraint,

$$C_2(\boldsymbol{e}_2) = \sum_i |w^{(i),t}e_2^{(i)}| = \|W\boldsymbol{e}_2\|_1 \tag{9}$$

where W is a diagonal matrix, with i-th diagonal element

$$w^{(i),t} = \frac{\exp(-|e_2^{(i),t}|/b)}{\exp(-|e_2^{(i),t}|/b) + c} \propto \rho'(|e_2^{(i),t}|) \tag{10}$$

After understanding the motivation of choosing the corresponding penalties for the two types of errors, we construct the final objective function as follows,

$$\min_{\boldsymbol{x}} \quad \alpha\|T_M(\boldsymbol{e}_1)\|_{w,*} + (1-\alpha)\|W\boldsymbol{e}_2\|_1 + \beta\|\boldsymbol{x}\|_1$$

$$\text{s.t.} \quad \boldsymbol{y} - D\boldsymbol{x} = \boldsymbol{e}_1 + \boldsymbol{e}_2 \tag{11}$$

where $0 < \alpha < 1$ is a weight for balancing the reweighted ℓ_1-norm penalty and the low-rank penalty corresponding to the two types of residuals. β is the weight of the sparsity penalty of coding coefficients.

The proposed model is denoted as low-rank Laplacian-uniform mixed model (LR-LUM). It is different from previous mixed models proposed in [36, 17]. Iliadis *et al.* proposed an objective function that utilized both sparsity and low-rank constraints for error penalties [17]. However, their model was designed against errors caused by occlusion, and they assigned both constraints to the same type of errors in their model. Moreover, they fitted the errors to a distribution based on a tailored loss function, rather than the LUM function adopted in our model. Nuclear norm was used for low-rank constraint rather than the robust (weighted) nuclear norm. In [36], ℓ_1-norm was utilized to penalize random errors. However, as demonstrated in Fig. 2 and argued in the above, the Laplacian distribution does not fit the random errors as well as the LUM function in practice.

3.2. Optimization

Now that we have the final objective function. We develop an algorithm to optimize it in the following. Let $B = [\frac{1}{\beta}WD, \frac{1}{1-\alpha}I]$, $\boldsymbol{c} = [\beta\boldsymbol{x}; (1-\alpha)W\boldsymbol{e}_2]$, then the minimization problem in (11) becomes

$$\min_{\boldsymbol{c}, \boldsymbol{e}_1} \quad \alpha\|T_M(\boldsymbol{e}_1)\|_{w,*} + \|\boldsymbol{c}\|_1$$

$$\text{s.t.} \quad W\boldsymbol{y} = B\boldsymbol{c} + W\boldsymbol{e}_1 \tag{12}$$

To solve the above optimization problem, we first make an equivalent transformation by introducing an auxiliary variable $\boldsymbol{d}$, to make this complex problem able to be decomposed into several easier subproblems. Let $\tilde{\boldsymbol{e}}_2 = W\boldsymbol{e}_2$. Then (12) can be rewritten as

$$\min_{\boldsymbol{c}, \boldsymbol{e}_1, \tilde{\boldsymbol{e}}_2} \quad \alpha\|T_M(\boldsymbol{d})\|_{w,*} + (1-\alpha)\|\tilde{\boldsymbol{e}}_2\|_1 + \beta\|\boldsymbol{x}\|_1$$

$$\text{s.t.} \quad W(\boldsymbol{y} - D\boldsymbol{x} - \boldsymbol{e}_1) = \tilde{\boldsymbol{e}}_2, \ \boldsymbol{d} = \boldsymbol{e}_1 \tag{13}$$

As a constrained optimization problem, (13) can be solved iteratively through the ADMM method [5]. Then the augmented Lagrangian function is defined as

$$L_\mu(\boldsymbol{e}_1, \tilde{\boldsymbol{e}}_2, \boldsymbol{x}, \boldsymbol{d}, \boldsymbol{z}_1, \boldsymbol{z}_2) =$$
$$\alpha\|T_M(\boldsymbol{d})\|_{w,*} + (1-\alpha)\|\tilde{\boldsymbol{e}}_2\|_1 + \beta\|\boldsymbol{x}\|_1$$
$$+ \boldsymbol{z}_1^T(W(\boldsymbol{y} - D\boldsymbol{x} - \boldsymbol{e}_1) - \tilde{\boldsymbol{e}}_2) + \boldsymbol{z}_2^T(\boldsymbol{d} - \boldsymbol{e}_1)$$
$$+ \frac{\mu}{2}(\|W(\boldsymbol{y} - D\boldsymbol{x} - \boldsymbol{e}_1) - \tilde{\boldsymbol{e}}_2\|_2^2 + \|\boldsymbol{d} - \boldsymbol{e}_1\|_2^2) \tag{14}$$

We can decompose the optimization problem into several subproblems as follows, and solve the subproblems and update the corresponding variables alternately.

$$\boldsymbol{e}_1^{t+1} = \arg\min_{\boldsymbol{e}_1} L_\mu(\boldsymbol{e}_1, \tilde{\boldsymbol{e}}_2^t, \boldsymbol{x}^t, \boldsymbol{d}^t, \boldsymbol{z}_1^t, \boldsymbol{z}_2^t) \tag{15}$$

$$\tilde{\boldsymbol{e}}_2^{t+1} = \arg\min_{\tilde{\boldsymbol{e}}_2} L_\mu(\boldsymbol{e}_1^{t+1}, \tilde{\boldsymbol{e}}_2, \boldsymbol{x}^t, \boldsymbol{d}^t, \boldsymbol{z}_1^t, \boldsymbol{z}_2^t) \tag{16}$$

$$\boldsymbol{x}^{t+1} = \arg\min_{\boldsymbol{x}} L_\mu(\boldsymbol{e}_1^{t+1}, \tilde{\boldsymbol{e}}_2^{t+1}, \boldsymbol{x}, \boldsymbol{d}^t, \boldsymbol{z}_1^t, \boldsymbol{z}_2^t) \tag{17}$$

$$\boldsymbol{d}^{t+1} = \arg\min_{\boldsymbol{d}} L_\mu(\boldsymbol{e}_1^{t+1}, \tilde{\boldsymbol{e}}_2^{t+1}, \boldsymbol{x}^{t+1}, \boldsymbol{d}, \boldsymbol{z}_1^t, \boldsymbol{z}_2^t) \tag{18}$$

$$\boldsymbol{z}_1^{t+1} = \boldsymbol{z}_1^t + \mu(W(\boldsymbol{y} - D\boldsymbol{x}^{t+1} - \boldsymbol{e}_1^{t+1}) - \tilde{\boldsymbol{e}}_2^{t+1}) \tag{19}$$

$$\boldsymbol{z}_2^{t+1} = \boldsymbol{z}_2^t + \mu(\boldsymbol{d}^{t+1} - \boldsymbol{e}_1^{t+1}) \tag{20}$$

where t denotes the iteration index in the algorithm.

To update $\boldsymbol{e}_1$, we fix the other variables, and remove irrelevant terms in the subproblem. Then $\boldsymbol{e}_1$ can be updated by solving the following problem,

$$\boldsymbol{e}_1^{t+1} = \arg\min_{\boldsymbol{e}_1} \|W(\boldsymbol{y} - D\boldsymbol{x}^t) - \tilde{\boldsymbol{e}}_2^t + \boldsymbol{z}_1^t/\mu - W\boldsymbol{e}_1\|_2^2$$
$$+ \|\boldsymbol{d}^t + \boldsymbol{z}_2^t/\mu - \boldsymbol{e}_1\|_2^2 \tag{21}$$

Let $\boldsymbol{g}_1 = W(\boldsymbol{y} - D\boldsymbol{x}^t) - \tilde{\boldsymbol{e}}_2^t + \boldsymbol{z}_1^t/\mu$, and $\boldsymbol{g}_2 = \boldsymbol{d}^t + \boldsymbol{z}_2^t/\mu$. This problem has a closed form solution,

$$\boldsymbol{e}_1^{t+1} = (W^TW + I)^{-1}(W^T\boldsymbol{g}_1 + \boldsymbol{g}_2) \tag{22}$$

To update $\tilde{\boldsymbol{e}}_2$, we have the following equation,

$$\tilde{\boldsymbol{e}}_2^{t+1} = \arg\min_{\tilde{\boldsymbol{e}}_2} \frac{1}{2}\|W(\boldsymbol{y} - D\boldsymbol{x}^t - \boldsymbol{e}_1^{t+1}) + \boldsymbol{z}_1^t/\mu - \tilde{\boldsymbol{e}}_2\|_2^2$$
$$+ ((1-\alpha)/\mu)\|\tilde{\boldsymbol{e}}_2\|_1 \tag{23}$$

This ℓ_1-min problem has also a closed form solution. Let $\boldsymbol{g}_3 = W(\boldsymbol{y} - D\boldsymbol{x}^t - \boldsymbol{e}_1^{t+1}) + \boldsymbol{z}_1^t/\mu$. The optimal solution of $\tilde{\boldsymbol{e}}_2$ can be obtained by soft-thresholding [6] as follows,

$$\tilde{\boldsymbol{e}}_2^{t+1} = \text{soft}(\boldsymbol{g}_3, ((1-\alpha)/\mu)) \tag{24}$$

The soft-thresholding operator is defined element-wise as

$$\text{soft}(\boldsymbol{x}, \alpha)_i = \text{sign}(x_i) \cdot \max(|x_i| - \alpha, 0) \tag{25}$$

Let $\boldsymbol{g}_4 = W(\boldsymbol{y} - \boldsymbol{e}_1^{t+1}) + \boldsymbol{z}_1^t/\mu - \tilde{\boldsymbol{e}}_2^{t+1}$. The minimization problem for updating $\boldsymbol{x}$ is

$$\boldsymbol{x}^{t+1} = \arg\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{g}_4 - WD\boldsymbol{x}\|_2^2 + \frac{\beta}{\mu}\|\boldsymbol{x}\|_1 \tag{26}$$

This is a ℓ_1-ℓ_2 minimization problem, which can be solved via the fast iterative shrinkage-thresholding algorithm (FISTA) [2].

Let $\boldsymbol{g}_5 = \boldsymbol{e}_1^{t+1} - \boldsymbol{z}_2^t/\mu$. The optimization problem for $\boldsymbol{d}$ is

$$\boldsymbol{d}^{t+1} = \arg\min_{\boldsymbol{d}} \frac{1}{2}\|\boldsymbol{d} - \boldsymbol{g}_5\|_2^2 + \frac{\alpha}{\mu}\|T_M(\boldsymbol{d})\|_{w,*} \tag{27}$$

Algorithm 1 : The iterative optimization algorithm for the proposed LR-LUM model

---

Input: A test image $\boldsymbol{y} \in \mathbb{R}^m$, the dictionary $D \in \mathbb{R}^{m \times n}$

Output: The identity of $\boldsymbol{y}$

1: Initialize $t = 0$, $\boldsymbol{x}^t = [\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}]$, $\boldsymbol{e}_1^t = \boldsymbol{e}_2^t = \boldsymbol{d}^t = \frac{1}{2}(\boldsymbol{y} - D\boldsymbol{x}^t)$, $\boldsymbol{z}_1^t = \boldsymbol{z}_2^t = \boldsymbol{0}$

2: **repeat**

3:     Update the weight matrix W via (10)

4:     **repeat**

5:         Update $\boldsymbol{e}_1$ and $\tilde{\boldsymbol{e}}_2$ via (22) and (24)

6:         Update $\boldsymbol{x}$ via FISTA

7:         Update $\boldsymbol{d}$ via (28)

8:         Update $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ via (19) and (20)

9:         $t = t + 1$

10:     **until** convergence or maximum iterations.

11: **until** convergence or maximum iterations.

12: Output the identity with the converged W^* and $\boldsymbol{x}^*$

    $\text{identity}(\boldsymbol{y}) = \arg\min_i \|W^*(\boldsymbol{y} - D\delta_i(\boldsymbol{x}^*))\|_2^2$,

    where $\delta_i(\boldsymbol{x}^*)$ is a vector whose nonzero components are only those of $\boldsymbol{x}^*$ corresponding to the i-th class.

This problem can be solved by singular value thresholding (SVT) as in [36]. Let $G = T_M(\boldsymbol{g}_5)$, and $G = U\Sigma V^T$ be the SVD of G, where Σ is a semi-positive diagonal matrix. The optimal solution for $\boldsymbol{d}$ can be obtained as follows,

$$\boldsymbol{d}^{t+1} = \text{vec}(US_{\omega,\frac{\alpha}{\mu}}(\Sigma)V^T) \tag{28}$$

where $S_{\omega,\lambda}(\cdot)$ is the weighted SVT operator defined as

$$S_{\omega,\lambda}(\Sigma) = \text{diag}\left(\max(\Sigma_{i,i} - \omega_i\lambda, 0)\right) \tag{29}$$

Now that we have the solutions for all the subproblems. We only need to update the variables alternately and iteratively until convergence (relevant variables only show negligible changes). The weight matrix W for random errors is fixed during the above iteration, and updated using (10) in an outer loop. The overall algorithm can be summarized as in Algorithm 1.

4. Experiments

4.1. Databases & Parameter Settings

To verify the effectiveness of the LR-LUM model and Algorithm 1, we conduct extensive experiments on several face recognition benchmark databases, including Extended Yale-B (EYB) [10, 22], AR [24], and Labeled Face in the Wild (LFW) [15, 16]. Face images in EYB and AR are all collected in restricted condition with frontal view, while faces in LFW are captured in uncontrolled environments. Images from the EYB and AR databases are resized

to 96×84 and 42×30 pixels, respectively, to verify the performance of the proposed method in dealing with face images of different resolutions. EYB is divided into 5 subsets according to different illumination conditions. AR contains Session 1 and Session 2, which are collected on different dates. Faces from LFW are detected using the Viola-Jones face detector [32], and then cropped and aligned as in [11] and resized to 90×90 pixels.

The parameters b and c in (10) should fit the actual distribution of coding residuals. They are estimated in each iteration using the mean e_m of the absolute value of coding residuals. Specifically, b and c are set to e_m^t/k and $c_0 \exp\left(-e_m^t/b\right)$, respectively. $k = 10$ and $c_0 = 50$ in all experiments, except for testing on clean LFW images, for which we set $k = 2$ and $c_0 = 0.01$ to incline to small residuals caused by inaccurate alignment. To balance the penalties, we set $\alpha = 0.25$ and $\beta = 0.2$ in the objective function (11). μ in (14) is set to $2m/\|\boldsymbol{y}\|_1$ as in [37]. The maximum iteration number of the outer loop in Algorithm 1 is set to 25. We set the maximum iteration number of the inner loop to 8 in the first 20 iterations of the outer loop for higher efficiency, and to 200 in the last 5 iterations of the outer loop to guarantee convergence.

Comparison is carried out between the proposed model and a number of state-of-the-art and competitive robust methods, including CESR [12], S-SRC [35], R-SRC [35], RRC-L1 [40], RRC-L2 [40], HQ-M [13], HQ-A [13], F-IRNNLS [17], and F-LR-IRNNLS [17].

4.2. Evaluation on the EYB Database

Two types of evaluation settings are adopted on EYB, which are denoted as "single training sample" and "multi training samples" protocols, respectively. For the single training sample setting, we choose the first sample of each subject in Subset 1 for training, and all images in Subset 3 (more extreme lighting conditions) for testing. In the case of multi training samples, we use all images from Subset 1 and Subset 2 for training, and Subset 3 for testing, as in [35], [40], and [17].

4.2.1 Face recognition on clean images

The basic recognition performances of all methods on clean images of EYB, i.e., images without further occlusion or corruption, are shown in Fig. 3. For multi training samples, there are 17-19 training samples for each subject. Most of the compared methods achieved pretty good performances under this setting. However, when only one training sample is available, the performances of all compared methods degrade significantly, while the proposed method still demonstrates an accuracy of 87.2%, which outperforms the other methods by over 27%.
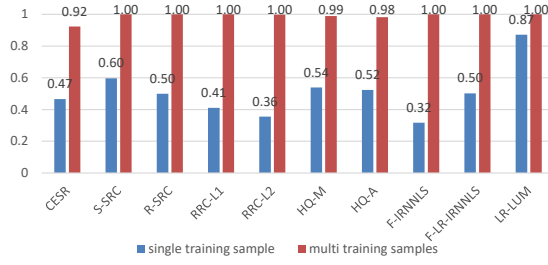
Figure 3. Face recognition accuracies on clean images of the EYB database.

4.2.2 Robust face recognition under various noises

We evaluate the robustness of the proposed method against various noises, including pixel corruption, block occlusion, and their mixture. The test set is further subject to different illumination conditions, which can be considered as another type of noise. Examples of face images under different noises are shown in Fig. 4.
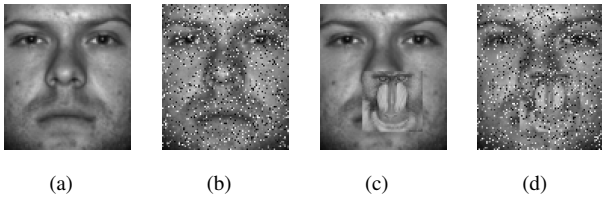

Figure 4. Face images of EYB under different noises. (a) A clean face image. (b) With 20% pixel corruption. (c) With 20% occlusion. (d) With 20% occlusion and 20% pixel corruption.

We evaluate all methods under 10 to 60 percent of three types of noises, i.e., pixel corruption, block occlusion, and their mixture. For example, 10% pixel corruption is introduced by setting 10% pixels at random locations of the face image to random values, and 10% occlusion is added by replacing a square block (10% size of the image) at a random position with a commonly used baboon image. 10% mixed noises means 10% occlusion plus 10% pixel corruption.

As shown in Fig. 5, our method demonstrates the best overall performance. In particular, with a single training sample, the proposed model significantly outperforms the other methods, which means that our method can be applied in difficult scenarios when only few training samples are available. For example, such an algorithm can potentially provide useful information for law enforcement department to identify faces captured in a crime scene, in which case extreme illumination condition, occlusion, and corruption often take place, and only limited gallery samples are available for verification. As for the setting with multi training samples, the proposed method achieves the best results against pixel corruption and mixed noises, and the second best against block occlusion. F-LR-IRNNLS [17] performs
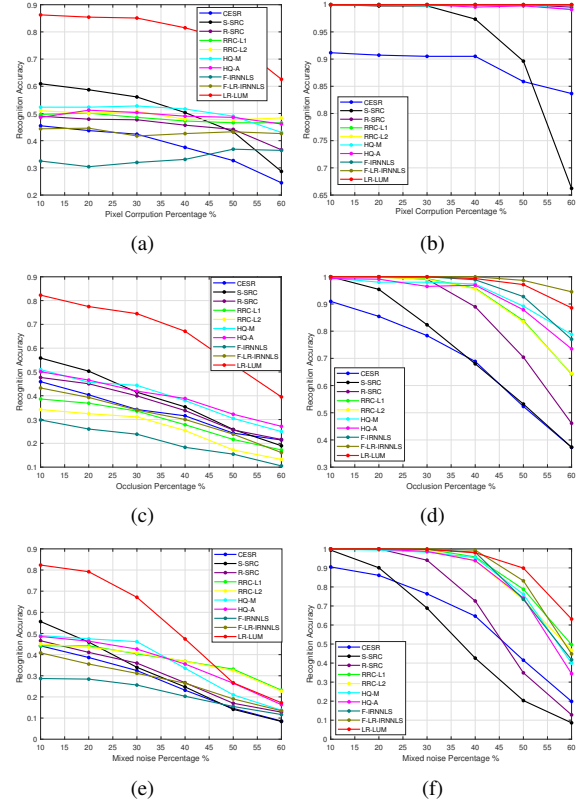

Figure 5. Recognition accuracies on EYB under different test protocols. (a), (c), and (e): With a single training sample under 10%–60% pixel corruption, occlusion, and mixed noises, respectively. (b), (d), and (f): With multi training samples under pixel corruption, occlusion, and mixed noises.

slightly better against occlusion, owing to its stronger priori assumption for contiguous occlusion. The proposed method is significantly more robust against various contaminations than F-LR-IRNNLS, when looking at the overall performance.

Contaminations such as pixel corruption and block occlusion introduced in the above experiments are artificially synthesized, which may not be the case in real scenarios. Noises occur in real scenes could be more complicated and in more unpredictable forms. However, the above experiments do provide us comprehensive evaluations of how the proposed model performs against various types of individual or mixed noises. The evaluation of the robustness of our method against noises in real scenarios will be further demonstrated in the following sections.

4.3. Evaluation on the AR Database

We also adopt two evaluation protocols, i.e., the single and multi training sample settings, to evaluate the proposed method on the AR database. For each subject in each session of the AR database, there are three face images with sunglass disguise and three with scarf, where several exam-

ples are shown in Fig. 6. We randomly selected 50 male and 50 female subjects for the experiments. For the single training sample setting, we conduct experiments on each session separately. The only natural clean face image of each subject in each session is used for training. The three samples with sunglasses and the other three with scarf are used for testing, respectively. As for the multi training sample setting, we use the two natural clean face images from both sessions for training, and six samples with sunglasses and six with scarf from both sessions for testing.
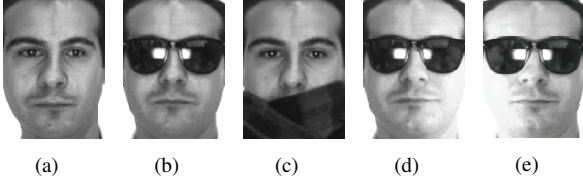

Figure 6. Face images in AR under real disguise. (a) A clean face image. (b) With sunglasses. (c) With scarf. (d) With sunglasses and left side light on. (e) With sunglasses and right side light on.

4.3.1   Evaluation under real disguise

The comparison of recognition accuracies under real disguise is shown in Fig. 7. The proposed method outperforms all the other methods under both single and multi training sample settings, against sunglass or scarf disguise. Under the single training sample setting, our method achieves 88% accuracy in Session 1 and 92% in Session 2 against sunglass occlusion, and 68.2% in Session 1 and 66.8% in Session 2 against scarf occlusion, which are over 3%, 5%, 10%, and 7% improvements, respectively, compared to all other methods. Under the multi training sample setting, our method achieves the highest accuracies of 92.5% against sunglasses and 66.9% against scarf, which are over 4% and 5% improvements compared to all other methods.

4.3.2   Recognition under mixed noises

To further evaluate the effectiveness of the proposed model under more extreme circumstances, we conduct experiments on the AR database with mixed noises of real disguise and pixel corruption. The test samples are further subject to 10 to 60 percent pixel corruption. The performance comparison is shown in Fig. 8.

From Fig. 8, we can observe that the proposed model outperforms other compared methods in all experiments, which indicates that our method can handle various types of noises and manages to maintain good and stable performance in complex circumstances. The experimental results verify that the proposed model successfully captures the characteristics of both non-structured random noise and structured occlusion, by using the Laplacian-uniform mixed

Figure 7. Face recognition under real disguise on the AR database. (a) With sunglasses. (b) With scarf.

distribution to model random errors and robust nuclear norm to constrain structured errors.

4.4. Evaluation on the LFW Database

In order to verify the robustness of our method against uncontrolled conditions in practice, we carry out further experiments on the LFW database. Face images in LFW contain many unpredictable variations, such as illumination changes, head pose variations, *etc.*, which pose serious challenges for recognition. We selected 100 experimental subjects, each of which contains no less than 14 face images. For each subject, 7 face images were randomly selected for training, and the other 7 for testing. Finally, there are 700 face images in each of the training and testing sets.

We evaluate all methods on the clean test images as well as those with 20% and 40% mixed noises of block occlusion and pixel corruption. As shown in Table 1, the proposed method demonstrates the best robustness among all compared methods. When the test samples are contaminated by 40% block occlusion and 40% pixel corruption, our method still achieves a recognition accuracy of 60.8%, which is over 15% higher than the second best. Note that the overall noises are very complicated and challenging by combining the mixed noises that we intentionally added to the test samples with the intrinsic noises that the dataset originally contains. The experimental results demonstrate the strong robustness of our method.
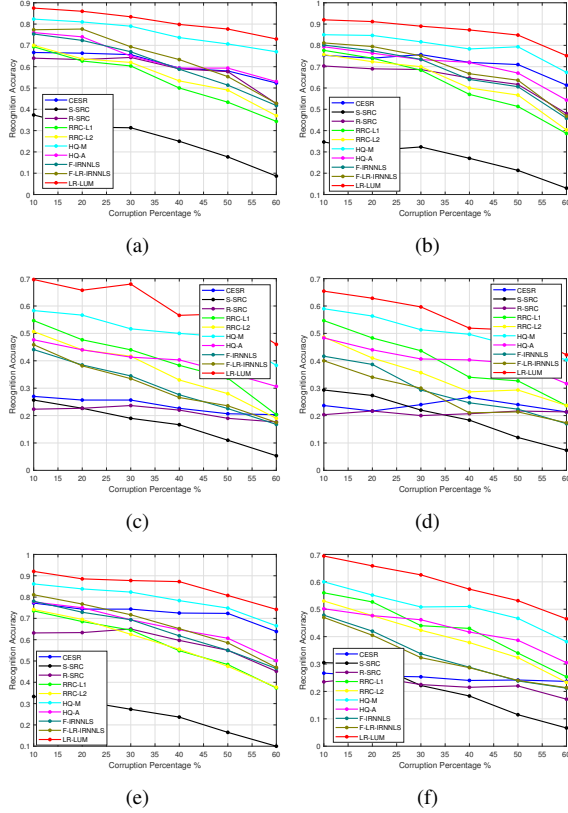
Figure 8. Recognition accuracies on AR under different test protocols. (a) and (b): Performance against sunglass disguise in Session 1 and Session 2, respectively, under the single training sample setting. (c) and (d): Performance against scarf disguise in Session 1 and Session 2, respectively, under the single training sample setting. (e) and (f): Performance against sunglasses and scarf, respectively, under the multi training sample setting.

Methods	Clean	20% OC	40% OC
CESR [12]	66.3	60.3	32.4
S-SRC [35]	77.1	47.6	14.1
R-SRC [35]	79.9	69.3	30.7
RRC-L1 [40]	76.0	74.7	32.6
RRC-L2 [40]	76.7	67.7	24.4
HQ-M [13]	81.4	65.4	23.6
HQ-A [13]	81.3	65.4	24.0
F-IRNNLS [17]	78.9	73.0	45.7
F-LR-IRNNLS [17]	81.0	71.9	45.0
LR-LUM	**83.5**	**77.6**	**60.8**

Table 1. Accuracies (in percentage) under mixed noises on LFW. 20% OC means 20% block occlusion plus 20% pixel corruption.

5. Discussion

One issue to be noticed is that the recognition accuracy of our model on clean test samples in LFW might seem unsatisfying, considering many deep learning approaches

could achieve accuracies of higher than 90%. However, this could be an unfair comparison, since deep learning needs massive extra training data to achieve good performance. Moreover, deep neural networks for regular face recognition achieve poor performance under contaminated condition. We have tested a pre-trained VGG-Face model [27] on the LFW dataset constructed in Section 4.4, and its performance drops to about 1% under 20% mixed noises.

Another reason why we did not consider experimental comparison to deep learning methods in the above experiments is, to the best of our knowledge, that there is lack of comprehensive works using deep neural networks for face recognition against dense noises. We noticed one specific work [34], which employed a MaskNet trained to be better activated on non-occluded facial regions to avoid errors caused by contiguous occlusion. However, as mentioned above, they used a large face dataset to pre-train their network. Besides, occluded samples are used for fine-tuning, which actually assumes that the type and the form of noises are known beforehand. Also, there is still lack of evaluation to demonstrate its robustness under more challenging situations, such as random corruption or mixed noises.

6. Conclusion

In this paper, we attempt to better model errors caused by complicated noises in robust face recognition, by introducing appropriate penalty terms. A low-rank Laplacian-uniform mixed (LR-LUM) model is proposed, which models complex errors as a combination of two types of noises, i.e., continuous structured noises and random noises. The joint model integrates the LUM function to better fit the empirical distribution of random noises, and robust nuclear norm to enforce the low-rank constraint on the structured errors. An iteratively reweighted algorithm is developed to effectively solve the proposed objective function. A series of experiments have been designed to verify the discriminative ability and robustness of the proposed algorithm under various challenging circumstances. Extensive experimental comparison with many classical and state-of-the-art methods demonstrates the superiority of our method in terms of robustness against various types of contaminations.

References

[1] S. Albelwi and A. Mahmood. A deep architecture for face recognition based on multiple feature extraction techniques. In *IEEE International Conference on Signal and Image Processing Applications*, pages 390–395, Sept. 2017. 1

[2] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009. 4

[3] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, Jul. 1997. 1

[4] M. Bicego, A. Lagorio, E. Grosso, and M. Tistarelli. On the use of SIFT features for face authentication. In *IEEE Conference on Computer Vision and Pattern Recognition Workshop*, pages 35–35, Jun. 2006. 1

[5] S. P. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011. 4

[6] P. Combettes and V. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling & Simulation*, 4(4):1168–1200, 2005. 4

[7] W. Deng, J. Hu, and J. Guo. Extended SRC: Undersampled face recognition via intraclass variant dictionary. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(9):1864–1870, Sept. 2012. 2

[8] W. Deng, J. Hu, and J. Guo. Face recognition via collaborative representation: Its discriminant nature and superposed representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(10):2513–2521, Oct. 2018. 2

[9] X. Ding, L. He, and L. Carin. Bayesian robust principal component analysis. *IEEE Transactions on Image Processing*, 20(12):3419–3430, Dec. 2011. 1

[10] 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. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):643–660, Jun. 2001. 5

[11] T. Hassner, S. Harel, E. Paz, and R. Enbar. Effective face frontalization in unconstrained images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4295–4304, Jun. 2015. 5

[12] R. He, W. Zheng, and B. Hu. Maximum correntropy criterion for robust face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1561–1576, Aug. 2011. 2, 5, 8

[13] R. He, W. Zheng, T. Tan, and Z. Sun. Half-quadratic-based iterative minimization for robust sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(2):261–275, Feb. 2014. 2, 5, 8

[14] G. E. Hinton, S. Osindero, and Y. W. Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7):1527–1554, 2006. 1

[15] G.B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, Oct. 2007. 5

[16] G. B. Huang and E. Learned-Miller. Labeled faces in the wild: Updates and new reporting procedures. Technical Report UM-CS-2014-003, University of Massachusetts, Amherst, 2014. 5

[17] M. Iliadis, H. Wang, R. Molina, and A. K. Katsaggelos. Robust and low-rank representation for fast face identification with occlusions. *IEEE Transactions on Image Processing*, 26(5):2203–2218, May 2017. 4, 5, 6, 8

[18] R. Jenatton, G. Obozinski, and F. R. Bach. Structured sparse principal component analysis. In *International Conference on Artificial Intelligence and Statistics*, pages 366–373, 2010. 1

[19] H. Jia and A. M. Martinez. Face recognition with occlusions in the training and testing sets. In *IEEE International Conference on Automatic Face Gesture Recognition*, pages 1–6, Sept. 2008. 2

[20] K. Jia, T-H. Chan, and Y. Ma. Robust and practical face recognition via structured sparsity. In *European Conference on Computer Vision*, pages 331–344, 2012. 2

[21] M. A. Kashem, N. Akhter, S. Ahmed, and M. Alam. Face recognition system based on principal component analysis (PCA) with back propagation neural networks (BPNN). *International Journal of Scientific & Engineering Research*, 2(6):1–10, 2011. 1

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

[23] S. Liao, A. K. Jain, and S. Z. Li. Partial face recognition: Alignment-free approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(5):1193–1205, May 2013. 2

[24] A.M. Martinez and R. Benavente. The AR face database. Technical Report 24, Computer Vision Center, Autonomous University of Barcelona, Bellaterra, Barcelona, Mar. 1998. 5

[25] A. M. Martinez. Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(6):748–763, Jun. 2002. 2

[26] I. Naseem, R. Togneri, and M. Bennamoun. Linear regression for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(11):2106–2112, Nov. 2010. 2

[27] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *British Machine Vision Conference*, pages 41.1–41.12, 2015. 1, 8

[28] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 815–823, Jun. 2015. 1

[29] Á. Serrano, I. M. de Diego, C. Conde, and E. Cabello. Recent advances in face biometrics with Gabor wavelets: A review. *Pattern Recognition Letters*, 31(5):372–381, 2010. 1

[30] Y. Sun, X. Wang, and X. Tang. Deeply learned face representations are sparse, selective, and robust. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2892–2900, Jun. 2015. 1

[31] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1701–1708, Jun. 2014. 1

[32] P. A. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004. 5

[33] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma. Toward a practical face recognition system: Robust alignment and illumination by sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(2):372–386, Feb. 2012. 2

[34] W. Wan and J. Chen. Occlusion robust face recognition based on mask learning. In *IEEE International Conference on Image Processing*, pages 3795–3799, Sept. 2017. 8

[35] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, Feb. 2009. 1, 2, 5, 8

[36] J. Xie, J. Yang, J. J. Qian, Y. Tai, and H. M. Zhang. Robust nuclear norm-based matrix regression with applications to robust face recognition. *IEEE Transactions on Image Processing*, 26(5):2286–2295, May 2017. 2, 3, 4, 5

[37] A. Y. Yang, Z. Zhou, A. G. Balasubramanian, S. S. Sastry, and Y. Ma. Fast ℓ_1-minimization algorithms for robust face recognition. *IEEE Transactions on Image Processing*, 22(8):3234–3246, Aug. 2013. 2, 5

[38] J. Yang, L. Luo, J. Qian, Y. Tai, F. Zhang, and Y. Xu. Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(1):156–171, Jan. 2017. 2

[39] M. Yang and L. Zhang. Gabor feature based sparse representation for face recognition with Gabor occlusion dictionary. In *European Conference on Computer Vision*, pages 448–461, 2010. 1

[40] M. Yang, L. Zhang, J. Yang, and D. Zhang. Regularized robust coding for face recognition. *IEEE Transactions on Image Processing*, 22(5):1753–1766, May 2013. 2, 5, 8

[41] X. Yuan and S. Yan. Visual classification with multi-task joint sparse representation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3493–3500, Jun. 2010. 1

[42] H. Zhang, J. Yang, J. Qian, and W. Luo. Nonconvex relaxation based matrix regression for face recognition with structural noise and mixed noise. *Neurocomputing*, 269:188–198, 2017. 2