

Robust Low-Rank Regularized Regression for Face Recognition with Occlusion

Jianjun Qian¹, Jian Yang¹, Fanlong Zhang¹ and Zhouchen Lin²

¹School of Computer Science and Engineering, Nanjing University of Science and Technology

²Key Lab. of Machine Perception (MOE), Peking University

csjqian@gmail.com, csjyang@njjust.edu.cn, zhangfanlong@gmail.com, zlin@pku.edu.cn

Abstract

Recently, regression analysis based classification methods are popular for robust face recognition. These methods use a pixel-based error model, which assumes that errors of pixels are independent. This assumption does not hold in the case of contiguous occlusion, where the errors are spatially correlated. Observing that occlusion in a face image generally leads to a low-rank error image, we propose a low-rank regularized regression model and use the alternating direction method of multipliers (ADMM) to solve it. We thus introduce a novel robust low-rank regularized regression (RLR³) method for face recognition with occlusion. Compared with the existing structured sparse error coding models, which perform error detection and error support separately, our method integrates error detection and error support into one regression model. Experiments on benchmark face databases demonstrate the effectiveness and robustness of our method, which outperforms state-of-the-art methods.

1. Introduction

Automatic face recognition has been a hot topic in the area of computer vision and pattern recognition due to the increasing need for real-world applications [1]. Recently regression analysis becomes a popular tool for face recognition. Naseem et al. presented a linear regression classifier (LRC) for face classification [13]. Wright et al. proposed a sparse representation based classification (SRC) method to identify human faces with varying illumination changes, occlusion and real disguise [2]. A test sample image is coded as a sparse linear combination of the training images, and then the classification is made by identifying which class yields the least reconstruction residual. Although SRC performs well in face recognition, it lacks theoretical justification. Yang et al. gave an insight into SRC and sought reasonable supports for its effectiveness [3]. They thought that the L_1 -optimizer has two properties, sparseness and closeness. Sparseness determines a small number of nonzero representation coefficients and

closeness makes the nonzero representation coefficients concentrate on the training samples having the same class label as the test sample. However, the L_0 -optimizer can only achieve sparseness. Yang et al. constructed a Gabor occlusion dictionary to improve the performance and efficiency of SRC [4]. Yang and Zhang proposed a robust sparse coding (RSC) model for face recognition [6]. RSC is robust to various kinds of outliers (e.g., occlusion and facial expression). Based on the maximum correntropy criterion, He et al. [7, 8] presented robust sparse representation for face recognition. Recently, some researchers have begun to question the role of sparseness in face recognition [9, 10]. In [11], Zhang et al. analyzed the work rule of SRC and believed that it is the collaborative representation that improves the classification performance, rather than the L_1 -norm sparseness. They further introduced the collaborative representation based classification (CRC) with the non-sparse L_2 -norm to regularize the representation coefficients. CRC can achieve similar results as SRC and significantly speed up the algorithm.

The regression methods mentioned above all use the pixel-based error model [6], which assumes that errors of pixels are independent. This assumption does not hold in the case of contiguous occlusion, where errors are spatially correlated [5]. In addition, characterizing the representation error pixel by pixel individually neglects the whole structure of the error image. To address these problems, Zhou et al. incorporated the Markov Random Field model into the sparse representation framework for spatial continuity of the occlusion [5]. Li et al. explored the intrinsic structure of contiguous occlusion and proposed a structured sparse error coding (SSEC) model [12]. These two works share the same two-step iteration strategy: (1) Detecting errors via sparse representation or coding, and (2) Estimating error supports (i.e. determining the real occluded part) using graph cuts. The difference is that SSEC uses more elaborate techniques, such as the iteratively reweighted sparse coding in the error detection step and a morphological graph model in the error support step for achieving better performance. However, SSEC does not numerically converge to the desired solution; it needs an additional quality assessment model to choose the desired solution from the iteration sequence.

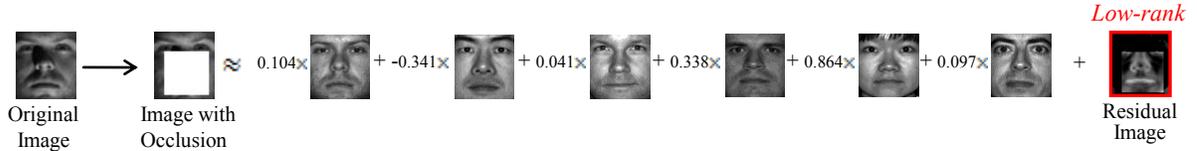


Figure 1: The image with block occlusion is linearly represented by six different images and the residual (noise) image.

Some recent works point out that the visual data has low-rank structure [17]. Most of the existing methods aim to find a low-rank approximation for matrix completion. However, the rank minimization problem is NP hard in general. In [14], Fazel et al. applied the nuclear norm heuristic to solve the rank minimization problem, where the nuclear norm of a matrix is the sum of its singular values. Based on these results, robust principle component analysis (RPCA) is proposed to decompose an image into two parts: data matrix (low-rank) and the noise (sparse) [15, 16]. Zhang et al. introduced a matrix completion algorithm based on the Truncated Nuclear Norm Regularization for estimating missing values [17]. Ma et al. integrated rank minimization into sparse representation for dictionary learning and applied the model for face recognition [18]. Chen et al. presented a novel low-rank matrix approximation algorithm with structural incoherence for robust face recognition [19].

This paper focuses on face recognition with occlusion. We observe that contiguous occlusion in a face image generally leads to a low-rank representation error image (i.e. error image or noise image), as shown in Fig. 1. Motivated by this observation, we add a rank function (which is replaced by a nuclear norm for easy optimization) of the representation error image into a regression model. The model can be solved via the alternating direction method of multipliers (ADMM) [22]. The proposed method has the following merits:

- (1) Compared with state-of-the-art regression methods such as SRC, RSC and CESR, which characterize the representation error individually and neglect the whole structure of the error image, our model views the error image as a whole and takes full use of its low-rank structure.
- (2) Compared with SSEC [12] and Zhou’s method [5], which perform the error detection step and the error support step iteratively but cannot guarantee the convergence of the whole algorithm, our method integrates error detection and error support into one regression model, and its ADMM algorithm theoretically converges well.

2. Low-Rank Regularized Regression

In this section, we present the low-rank regularized regression model to code the image and use the alternating direction method of multipliers [22] to solve the model.

2.1. Low-Rank Regularized Regression

Suppose that we are given a dataset of n matrices $\mathbf{A}_1, \dots, \mathbf{A}_n \in R^{d \times l}$ and a matrix $\mathbf{Y} \in R^{d \times l}$. Let us represent \mathbf{Y} linearly by taking the following form:

$$\mathbf{Y} = \mathbf{F}(\mathbf{x}) + \mathbf{E}, \quad (1)$$

where $\mathbf{F}(\mathbf{x}) = x_1 \mathbf{A}_1 + x_2 \mathbf{A}_2 + \dots + x_n \mathbf{A}_n$, $\mathbf{x} = (x_1, \dots, x_n)^T$ is the representation coefficient vector. \mathbf{E} is the noise (representation error).

Generally, the \mathbf{x} can be determined by solving the following optimization problem (linear regression):

$$\min_{\mathbf{x}} \|\mathbf{F}(\mathbf{x}) - \mathbf{Y}\|_F^2, \quad (2)$$

where $\|\cdot\|_F$ is the Frobenius norm of a matrix.

To avoid over fitting, we often solve the following regularized model (ridge regression) instead:

$$\min_{\mathbf{x}} \|\mathbf{F}(\mathbf{x}) - \mathbf{Y}\|_F^2 + \eta \|\mathbf{x}\|_2^2. \quad (3)$$

The above optimization problem can be solved in a closed form.

The rank of a matrix is a good tool to describe the structural characteristics of an error image. The error (representation error) image, as shown in Fig. 1, is typically low rank as opposed to the full rank original image. However, the existing linear regression models do not make use of this kind of structural information. To address this problem, we introduce the low-rank regularization to the ridge regression model. Specifically, the optimization problem is formulated as follows:

$$\min_{\mathbf{x}} \|\mathbf{F}(\mathbf{x}) - \mathbf{Y}\|_F^2 + \lambda \text{rank}(\mathbf{F}(\mathbf{x}) - \mathbf{Y}) + \eta \|\mathbf{x}\|_2^2, \quad (4)$$

where λ is the balance factor and η is the regularized parameter.

The optimization problem of Eq. (4) is extremely difficult to solve due to the discrete nature of the rank function. Fortunately, as suggested by the matrix completion method [14, 20, 21], the rank minimization problem can be replaced by the nuclear norm minimization problem:

$$\min_{\mathbf{x}} \|\mathbf{F}(\mathbf{x}) - \mathbf{Y}\|_F^2 + \lambda \|\mathbf{F}(\mathbf{x}) - \mathbf{Y}\|_* + \eta \|\mathbf{x}\|_2^2. \quad (5)$$

In the following section, we will develop the optimization algorithm to solve Eq. (5) by using the alternating direction method of multipliers [22].

2.2. Optimization via ADMM

In this section, we show how the alternating direction method of multipliers (ADMM) can be adopted to solve Eq. (5) efficiently. For more details of ADMM, we refer readers to [22]. To deal with our problem, the model in Eq. (5) is rewritten as:

$$\min_{\mathbf{x}, \mathbf{E}} \|\mathbf{E}\|_F^2 + \lambda \|\mathbf{E}\|_* + \frac{1}{2} \eta \mathbf{x}^T \mathbf{x}, \quad (6)$$

s.t. $\mathbf{F}(\mathbf{x}) - \mathbf{Y} = \mathbf{E}.$

The augmented Lagrange function is given by:

$$\begin{aligned} L_\mu(\mathbf{x}, \mathbf{E}, \mathbf{Z}) = & \|\mathbf{E}\|_F^2 + \lambda \|\mathbf{E}\|_* + \frac{1}{2} \eta \mathbf{x}^T \mathbf{x} + \\ & \text{Tr}(\mathbf{Z}^T (\mathbf{F}(\mathbf{x}) - \mathbf{E} - \mathbf{Y})) + \\ & \frac{\mu}{2} \|\mathbf{F}(\mathbf{x}) - \mathbf{E} - \mathbf{Y}\|_F^2, \end{aligned} \quad (7)$$

where $\mu > 0$ is a penalty parameter, \mathbf{Z} is the Lagrange multiplier, and $\text{Tr}(\cdot)$ is the trace operator.

ADMM consists of the following iterations:

$$\mathbf{x}^{k+1} = \arg \min_{\mathbf{x}} L_\mu(\mathbf{x}) \quad (8)$$

$$\mathbf{E}^{k+1} = \arg \min_{\mathbf{E}} L_\mu(\mathbf{E}) \quad (9)$$

$$\mathbf{Z}^{k+1} = \mathbf{Z}^k + \mu(\mathbf{F}(\mathbf{x}^{k+1}) - \mathbf{E}^{k+1} - \mathbf{Y}) \quad (10)$$

Updating \mathbf{x}

Denote $\mathbf{H} = [\text{Vec}(\mathbf{A}_1), \dots, \text{Vec}(\mathbf{A}_n)]$, $\mathbf{g} = \text{Vec}(\mathbf{E} + \mathbf{Y} - \frac{1}{\mu} \mathbf{Z})$, where $\text{Vec}(\cdot)$ convert matrix into a vector, then the objective function $L_\mu(\mathbf{x})$ in Eq. (8) is equivalent to

$$\begin{aligned} L_\mu(\mathbf{x}) = & \eta \frac{1}{2} \mathbf{x}^T \mathbf{x} + \text{Tr}(\mathbf{Z}^T \mathbf{F}(\mathbf{x})) + \frac{\mu}{2} \|\mathbf{F}(\mathbf{x}) - \mathbf{E} - \mathbf{Y}\|_F^2 \\ = & \frac{1}{2} \eta \mathbf{x}^T \mathbf{x} + \frac{\mu}{2} \|\mathbf{F}(\mathbf{x}) - (\mathbf{E} + \mathbf{Y} - \frac{1}{\mu} \mathbf{Z})\|_F^2 + \\ & \text{Tr}((\mathbf{E} + \mathbf{Y})\mathbf{Z}^T - \frac{1}{2\mu} \mathbf{Z}\mathbf{Z}^T). \end{aligned} \quad (11)$$

The problem of Eq. (8) can be reformulated as:

$$\mathbf{x}^{k+1} = \arg \min_{\mathbf{x}} \left(\frac{\mu}{2} \|\mathbf{H}\mathbf{x} - \mathbf{g}\|_2^2 + \frac{1}{2} \eta \mathbf{x}^T \mathbf{x} \right). \quad (12)$$

Eq. (12) is actually a ridge regression model. So we can obtain the solution of Eq. (12) by:

$$\mathbf{x}^{k+1} = (\mathbf{H}^T \mathbf{H} + \frac{\eta}{\mu} \mathbf{I})^{-1} \mathbf{H}^T \mathbf{g} \quad (13)$$

Updating \mathbf{E}

The objective function $L_\mu(\mathbf{E})$ in Eq. (9) can be rewritten by

$$\begin{aligned} L_\mu(\mathbf{E}) = & \|\mathbf{E}\|_F^2 + \lambda \|\mathbf{E}\|_* - \text{Tr}(\mathbf{Z}^T \mathbf{E}) + \frac{\mu}{2} \|\mathbf{F}(\mathbf{x}) - \mathbf{E} - \mathbf{Y}\|_F^2 \\ = & \|\mathbf{E}\|_F^2 + \lambda \|\mathbf{E}\|_* - \text{Tr}(\mathbf{Z}^T \mathbf{E}) \\ & + \frac{\mu}{2} \text{Tr}(((\mathbf{F}(\mathbf{x}) - \mathbf{Y})^T - \mathbf{E}^T)(\mathbf{F}(\mathbf{x}) - \mathbf{E} - \mathbf{Y})) \\ = & \lambda \|\mathbf{E}\|_* + \frac{\mu}{2} \frac{\mu+2}{\mu} \text{Tr}(\mathbf{E}^T \mathbf{E} \\ & - 2 \frac{\mu}{2+\mu} (\mathbf{F}(\mathbf{x})^T - \mathbf{Y}^T + \frac{1}{\mu} \mathbf{Z}^T) \mathbf{E}) + \text{const1} \end{aligned} \quad (14)$$

Algorithm 1 Solving LR³ via ADMM

Input: A set of matrices $\mathbf{A}_1, \dots, \mathbf{A}_n$ and a matrix $\mathbf{Y} \in R^{p \times q}$, the model parameter λ , the termination condition parameter ε .

Initialize $\mathbf{E}^0, \mathbf{Z}^0, \mu$

while $\|\mathbf{F}(\mathbf{x}^{k+1}) - \mathbf{E}^{k+1} - \mathbf{Y}\|_F > \varepsilon$ or
 $\max(\|\mathbf{x}^{k+1} - \mathbf{x}^k\|_F, \|\mathbf{E}^{k+1} - \mathbf{E}^k\|_F) > \varepsilon$

do

$$\mathbf{x}^{k+1} = (\mathbf{H}^T \mathbf{H} + \frac{\eta}{\mu} \mathbf{I})^{-1} \mathbf{H}^T \mathbf{g},$$

$$\begin{aligned} \mathbf{E}^{k+1} = & \arg \min_{\mathbf{E}} \frac{\lambda}{\mu+2} \|\mathbf{E}\|_* \\ & + \frac{1}{2} \|\mathbf{E} - \frac{\mu}{2+\mu} (\mathbf{F}(\mathbf{x}) - \mathbf{Y} + \frac{1}{\mu} \mathbf{Z})\|_F^2, \end{aligned}$$

$$\mathbf{Z}^{k+1} = \mathbf{Z}^k + \mu(\mathbf{F}(\mathbf{x}^k) - \mathbf{E}^k - \mathbf{Y}).$$

end while

Output: Optimal representation coefficient \mathbf{x} .

$$\begin{aligned} = & \lambda \|\mathbf{E}\|_* + \frac{\mu+2}{2} \left\| \mathbf{E} - \frac{\mu}{2+\mu} (\mathbf{F}(\mathbf{x})^T - \mathbf{Y}^T + \frac{1}{\mu} \mathbf{Z}^T) \right\|_F^2 \\ & + \text{const2} \end{aligned}$$

where *const1* and *const2* are constant terms, which are independent of the variable \mathbf{E} . The optimization problem Eq. (9) can be reformulated as

$$\begin{aligned} \mathbf{E}^{k+1} = & \arg \min_{\mathbf{E}} \frac{\lambda}{\mu+2} \|\mathbf{E}\|_* \\ & + \frac{1}{2} \left\| \mathbf{E} - \frac{\mu}{2+\mu} (\mathbf{F}(\mathbf{x}) - \mathbf{Y} + \frac{1}{\mu} \mathbf{Z}) \right\|_F^2 \end{aligned} \quad (15)$$

As suggested by [23], the above optimization problem is solved by

$$\mathbf{E}^{k+1} = \mathbf{U} T_{\frac{\lambda}{\mu+2}}[\mathbf{S}] \mathbf{V}, \quad (16)$$

where $(\mathbf{U}, \mathbf{S}, \mathbf{V}^T) = \text{svd}\left(\frac{\mu}{2+\mu} (\mathbf{F}(\mathbf{x}) - \mathbf{Y} + \frac{1}{\mu} \mathbf{Z})\right)$.

The singular value shrinkage operator $T_{\frac{\lambda}{\mu+2}}[\mathbf{S}]$ is defined as

$$T_{\frac{\lambda}{\mu+2}}[\mathbf{S}] = \text{diag}\left(\left\{\max\left(0, S_{j,j} - \frac{\lambda}{\mu+2}\right)\right\}_{1 \leq j \leq n}\right). \quad (17)$$

Stopping criterion

As suggested in [22], the stopping criterion of the algorithm is: the primal residual must be small: $\|\mathbf{F}(\mathbf{x}^{k+1}) - \mathbf{E}^{k+1} - \mathbf{Y}\|_F \leq \varepsilon$, and the difference between successive iterations should also be small: $\max(\|\mathbf{x}^{k+1} - \mathbf{x}^k\|_F, \|\mathbf{E}^{k+1} - \mathbf{E}^k\|_F) \leq \varepsilon$, where ε is a given tolerance.

In summary, the pseudo code of our method to solve Eq. (6) is shown in Algorithm 1.

Algorithm 1 can be interpreted as using the two-step iteration strategy for robust face recognition as those used in [5, 12]. The step of updating \mathbf{x} is actually an error detection

step for determining the representation coefficients and representation errors, and the step of updating \mathbf{E} is actually an error support step for determining the real occluded part. So we can say that LR³ provides a unified framework to integrate error detection and error support detection into one simple model.

3. Classification based on Robust Low-Rank Regularized Regression

We notice that SSEC [12] adopts a robust sparse representation model, i.e. iteratively reweighted sparse coding in the error detection step, but our low-rank regularized regression (LR³) only uses a simple ridge regression model for updating \mathbf{x} . To further improve the robustness of the proposed method, in this section we borrow the idea of robust regression to our model and introduce a robust low-rank regularized regression model (RLR³):

$$\min_{\mathbf{x}} \|\mathbf{W} \circ (\mathbf{F}(\mathbf{x}) - \mathbf{Y})\|_F + \lambda \|\mathbf{W} \circ (\mathbf{F}(\mathbf{x}) - \mathbf{Y})\|_* + \eta \|\mathbf{x}\|_2^2. \quad (18)$$

where \mathbf{W} is a weight matrix of the representation error, and \circ denotes the Hadamard product of two matrices.

The robust low-rank regularized regression model can be solved by using the iteratively reweighted process [6]. Each iteration step is to solve a low-rank regularized regression problem. Specifically, given a test sample \mathbf{Y} , we compute the representation coefficient \mathbf{x} via Algorithm 1 and the representation error \mathbf{E} of \mathbf{Y} in order to initialize the weight. The representation error \mathbf{E} is initialized as $\mathbf{E} = \mathbf{Y} - \mathbf{Y}_{ini}$, where \mathbf{Y}_{ini} is the initial estimation of the images from the gallery set. In this study, we simply set \mathbf{Y}_{ini} as the mean image of all samples in the coding dictionary since we do not know which class the test image \mathbf{Y} belongs to. With the initialized \mathbf{Y}_{ini} , our method can estimate the weight matrix \mathbf{W} iteratively. $\mathbf{W}_{i,i}$ is the weight assigned to each pixel of the test image. The weight function [6] is:

$$\mathbf{W}_{i,j} = \frac{\exp(\alpha\beta - \alpha(\mathbf{E}_{i,j})^2)}{1 + \exp(\alpha\beta - \alpha(\mathbf{E}_{i,j})^2)}, \quad (19)$$

where α and β are positive scalars.

The convergence is achieved when the difference between the weights in successive iterations satisfies the following condition:

$$\|\mathbf{W}^{(t)} - \mathbf{W}^{(t-1)}\|_2 / \|\mathbf{W}^{(t-1)}\|_2 < \gamma. \quad (20)$$

Based on the optimization solution \mathbf{x} via the iterative process, we obtain a weighted dictionary $\mathbf{B} = [\mathbf{B}_1 \cdots \mathbf{B}_n]$, where $\mathbf{B}_i = \mathbf{W} \circ \mathbf{D}_i$, $i = 1, \dots, n$ and \mathbf{D} is the coding dictionary which is composed of the training samples. The test sample \mathbf{Y} is reconstructed as $\hat{\mathbf{Y}}_i = \sum_{j \in \delta_i(\mathbf{x})} x_j \mathbf{B}_{i,j}$, where

$\delta_i(\mathbf{x})$ is the function that selects the indices of the coefficients associated with the i -th class. The corresponding reconstruction error of the i -th class is defined as:

$$r_i(\mathbf{y}) = \|\mathbf{W} \circ (\mathbf{Y} - \hat{\mathbf{Y}}_i)\|_*. \quad (21)$$

The decision rule is: if $r_l(\mathbf{y}) = \min_i r_i(\mathbf{y})$, then \mathbf{y} is assigned to Class l .

Table 1 The recognition rates (%) of each classifier for face recognition on the AR database with disguise occlusion.

Methods	Sunglasses	Scarves
CRC	65.5	88.5
LR ³	75.0	90.0
SRC [2]	87.0	59.5
CESR [8]	99.0	42.0
SSEC	96.5	94.0
RSC [6]	99.0	97.0
RLR ³	99.0	100

4. Experiments

In this section, we compare the proposed methods LR³ and RLR³ with CRC, SRC, CESR, SSEC and RSC. In our experiments, there are five parameters of the proposed RLR³. The parameters α and β in Eq. (19) follows the suggestion in [6]. The default value for penalty parameter μ is 1. Both the balance factor λ and the regularized parameter η are introduced in the following experiments.

4.1. Face Recognition with Real Disguise

In our experiments, we only use a subset of AR face image database [24]. The subset contains 100 individuals, 50 males and 50 females. All the individuals have two session images and each session contains 13 images. The face portion of each image is manually cropped and then normalized to 42×30 pixels.

The first experiment chooses the first four images (with various facial expressions) from session 1 and session 2 of each individual to form the training set. The total training images is 800. There are two test sets: the images with sunglasses and the images with scarves. Each set contains 200 images (one image per session of each individual with neutral expression). The balance factor λ is 10^2 and 10^{-1} for the test images with sunglasses and scarves, respectively. The regularized parameter η is 4×10^4 . Table 1 lists the recognition rates of CRC, SRC, CESR, SSEC, RSC, LR³ and RLR³. From Table 1, we can see that RLR³ achieves the best performance among all the methods. LR³ also gives better results than CRC. Both RSC and CESR obtain the same results as RLR³ when the test images are with sunglasses. However, the results of SRC and CESR are

significantly lower than those of RLR³ when the test images are with scarves.

In the second experiment, four neutral images with different illumination from the first session of each individual are used for training. The disguise images with various illumination and glasses or scarves per individual in session 1 and session 2 for testing. The balance factor λ is 10^{-2} and the regularization parameter η is 4×10^4 . The recognition rates of each method are listed in Table 2. From Table 2, we can see that RLR³ significantly outperforms CRC, LR³, SRC, CESR, SSEC and RSC on different test subsets. SRC and CESR perform well on images with sunglasses and poorly on images with scarves. SSEC gives similar results as RSC in different cases. Compared to RSC, 3.0%, 2.0%, 4.0% and 4.6% improvement are achieved by RLR³ on four different testing sets.

Table 2 The recognition rates (%) of each classifier for face recognition on the AR database with disguise occlusion.

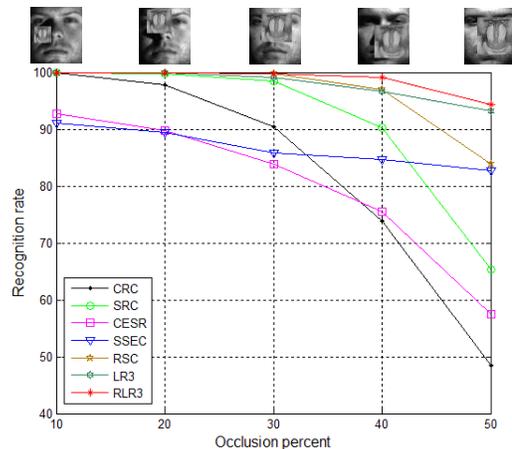
Methods	Sunglasses		Scarves	
	Session1	Session2	Session1	Session2
CRC	61.3	26.3	56.3	37.0
LR ³	75.7	38.3	72.0	45.3
SRC [2]	89.3	57.3	32.3	12.7
CESR [8]	95.3	79.0	38.0	20.7
SSEC	95.3	72.0	89.7	75.3
RSC [6]	94.7	80.3	91.0	72.7
RLR ³	97.7	82.3	95.0	77.3

4.2. Face Recognition with Random Block Occlusion

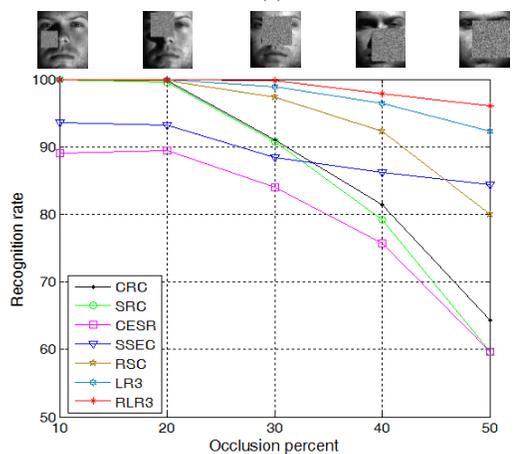
The extended Yale B face image database [25] contains 38 human subjects under 9 poses and 64 illumination conditions. All frontal-face images marked with P00 are used in our experiment, and each is resized to 96×84 pixels.

In the first experiment, we use the same experiment setting as in [2] to test the robustness of RLR³. Subsets 1 and 2 of Extended Yale B are used for training and subset 3 with the unrelated block images is used for testing. Both λ and η are set to 10. Fig. 2 (a) plots recognition rates of CRC, SRC, CESR, SSEC, RSC, LR³ and RLR³ under different levels of occlusions (from 10% to 50%). With the increment of the level of occlusion, RLR³ begins to significantly outperform the other methods. When the occlusion percentage is 50%, the recognition rate of RLR³ is 10.4%, 11.6%, 36.9% and 29% higher than RSC, SSEC, CESR and SRC, respectively.

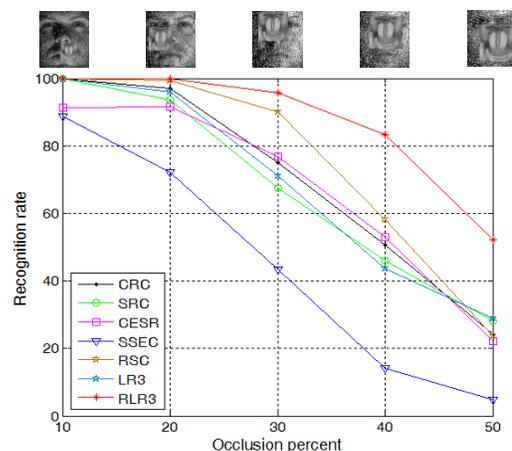
The second experiment setting is similar to that of the first experiment. The only difference is that subset 3 with noise block images is used for testing. λ is 0.1 and the regularization parameter η is set to 10. The recognition rates of each method versus the various levels of occlusion (from



(a)



(b)



(c)

Figure 2: The recognition rates (%) of CRC, SRC, CESR, SSEC, RSC, LR³ and RLR³ with the occlusion percentage being from 0% to 50%. (a) the test images are with unrelated block images; (b) the test images are with noise block images; (c) the test images are with mixture noise.

10% to 50%) are shown in Fig. 2 (b). From Fig. 2 (b), we observe that the proposed RLR³ significantly outperforms CRC, SRC, CESR, SSEC and RSC. The performances of SRC and CESR are not good in this case. SSEC gives good performance when the occlusion level is higher. However, SSEC cannot perform well when the occlusion level is lower. RSC achieves comparable results when the occlusion percentage is lower than 40%. However, the recognition rate of RLR³ is 16.1% higher than that of RSC when the occlusion percentage is 50%.

In the third experiment, subsets 1 and 2 of Extended Yale B are used for training and subset 3 with the mixture noise (pixel corruption and block occlusion) is used for testing. λ is 1 and the regularization parameter η is set to 10. The recognition rates of each method with different level of pixel corruption (and occlusion) are shown in Fig. 2 (c). Although the performance of each method degrades with the increment of the mixture noise level, RLR³ still achieves the best results among all the methods. The recognition rates of SSEC are poor when facing with the mixture noises (pixel corruption and image occlusion). A probable reason is that SSEC mainly addresses the contiguous occlusion problem.

5. Conclusions

In this paper, we present a novel low-rank regularized regression model and apply the alternating direction method of multipliers to solve it. The robust low-rank regularized regression based classification (RLR³) method is introduced for face recognition. RLR³ takes advantage of the structural characteristics of noise and provides a unified framework for integrating error detection and error support into one regression model. Extensive experiments demonstrate that the proposed RLR³ is robust to corruptions: real disguise and random block occlusion, and yields better performances as compared to state-of-the-art methods.

References

- [1] W. Zhao, R. Chellappa, P. J. Phillips *et al.*, Face recognition: A literature survey, *ACM Computing Surveys*, vol. 35, no. 4, pp. 399-459, Dec, 2003.
- [2] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE T. PAMI*, 31(2):210–227, 2009.
- [3] J. Yang, L. Zhang, Y. Xu, and J.Y. Yang, Beyond sparsity: the role of L1-optimizer in pattern classification, *Pattern Recognition*, 45 (2012) 1104–1118.
- [4] M. Yang and L. Zhang. Gabor feature based sparse representation for face recognition with Gabor occlusion dictionary. In *ECCV*, 2010.
- [5] Z. Zhou, A. Wagner, H. Mobahi, J. Wright, and Y. Ma. Face recognition with contiguous occlusion using Markov random fields. In *ICCV*, 2009.
- [6] M. Yang, L. Zhang, J. Yang, and D. Zhang, Robust sparse coding for face recognition, In *CVPR*, 2011.
- [7] R. He, W.S. Zheng, B.G. Hu, and X.W. Kong, A regularized correntropy framework for robust pattern recognition, *Neural Computation*, vol. 23, pp. 2074-2100, 2011.
- [8] R. He, W.S. Zheng, and B.G. Hu, Maximum correntropy criterion for robust face recognition, *IEEE T. PAMI*, vol. 33, no. 8, pp. 1561-1576, 2011.
- [9] R. Rigamonti, M. Brown, and V. Lepetit. Are sparse representations really relevant for image classification? In *CVPR*, 2011.
- [10] Q. Shi, A. Eriksson, A. Hengel, and C. Shen. Is face recognition really a compressive sensing problem? In *CVPR*, 2011.
- [11] L. Zhang, M. Yang, and X. C. Feng. Sparse representation or collaborative representation which helps face recognition? In *ICCV*, 2011.
- [12] X.-X. Li, D.-Q. Dai, X.-F. Zhang, and C.-X. Ren, Structured sparse error coding for face recognition with occlusion, *IEEE T. IP*, 2013, 22 (5), 1889-1999.
- [13] I. Naseem, R. Togneri, and M. Bennamoun. Linear regression for face recognition. *IEEE T. PAMI*, 32(11):2106-2112, 2010.
- [14] M. Fazel, H. Hindi, and S. Boyd, A rank minimization heuristic with application to minimum order system approximation, in *Proceedings of the American Control Conference*, vol. 6, 2001, pp. 4734–4739.
- [15] E. Candès, X.D. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM*, 58(3): Article 11, 2011.
- [16] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization, In *NIPS*, 2009.
- [17] D. Zhang, Y. Hu, J. P. Ye, X. L. Li, and X. F. He. Matrix completion by truncated nuclear norm regularization. In *CVPR*, 2012.
- [18] L. Ma, C. Wang, B. Xiao, and W. Zhou. Sparse representation for face recognition based on discriminative low-rank dictionary learning. In *CVPR*, 2012.
- [19] C.-F. Chen, C.-P. Wei, and Y.-C. Frank Wang. Low-rank matrix recovery with structural incoherence for robust face recognition. In *CVPR*, 2012.
- [20] E. J. Candès and Y. Plan. Matrix completion with noise. *Proceedings of the IEEE*, 98(6):925–936, 2009.
- [21] E. J. Candès and B. Recht, Exact matrix completion via convex optimization, *Commun. ACM* 55(6): 111-119 (2012).
- [22] S. 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.
- [23] J.F. Cai, E.J. Candès, Z. Shen, A singular value thresholding algorithm for matrix completion, *SIAM Journal on Optimization*, 2010.
- [24] A. Martinez and R. benavente. The AR face database. Technical Report 24, CVC, 1998.
- [25] K.C. Lee and J. Ho, and D. Driegman, Acquiring linear subspaces for face recognition under variable lighting, *IEEE T. PAMI*, 2005, 27(5), pp 684-698.