Pattern Recognition 48 (2015) 3145–3159

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Robust nuclear norm regularized regression for face recognition with occlusion

Jianjun Qian [a], Lei Luo [a], Jian Yang [a,*], Fanlong Zhang [a], Zhouchen Lin [b]

[a] School of Computer Science and Engineering, Nanjing University of Science and Technology, China
[b] Key Laboratory of Machine Perception (MOE), Peking University, China

ARTICLE INFO

Article history:
Received 23 September 2014
Received in revised form
15 April 2015
Accepted 15 April 2015
Available online 28 April 2015

Keywords:
Nuclear norm
Robust regression
Regularization
Face recognition

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. Furthermore, these methods ignore the whole structure of the error image. Nuclear norm as a matrix norm can describe the structural information well. Based on this point, we propose a nuclear-norm regularized regression model and use the alternating direction method of multipliers (ADMM) to solve it. We thus introduce a novel robust nuclear norm regularized regression (RNR) 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 areas of computer vision and pattern recognition due to the increasing need from 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 [16]. Wright et al. proposed a sparse representation based classification (SRC) method to identify human faces with varying illumination changes, occlusion and real disguise [2]. In SRC, 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-regularizer has two properties, sparseness and closeness. Sparseness determines a small number of nonzero representation coefficients and closeness makes the nonzero representation coefficients concentrating on the training samples have the same class label as the test sample. However, the L_0-regularizer can only achieve sparseness. So Yang et al. constructed a Gabor occlusion dictionary to improve the performance and efficiency of SRC [4,5]. Yang and Zhang proposed a robust sparse coding (RSC) model for face recognition [7]. RSC is robust to various kinds of outliers (e.g. occlusion and facial expression). Based on the maximum correntropy criterion, He et al. [8,9] presented robust sparse representation for face recognition. To unify the existing robust sparse regression models: the additive model represented by SRC for error correction and multiplicative model represented by CESR and RSC for error detection, He et al. [10] built a half-quadratic framework by defining different half-quadratic functions. The framework enables to perform both error correction and error detection. Recently, some researchers have begun to question the role of sparseness in face recognition [11,12]. In addition, Naseem et al. further extended their LRC to the robust linear regression classification (RLRC) using the Huber estimator to deal with severe random pixel noise and illumination changes [13]. In [14], Zhang et al. analyzed the work rule of SRC and believed that it is the collaborative representation, rather than the L_1-norm sparseness, that improves the classification performance. Zhang et al. 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 [7], which assumes that errors of pixels are independent. This assumption does not hold in the case of contiguous occlusion, where errors are spatially correlated [6]. 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

* Corresponding author.
E-mail addresses: csjqian@njust.edu.cn (J. Qian), zzdxpyy3001@163.com (L. Luo), csjyang@njust.edu.cn (J. Yang), zhangfanlong@gmail.com (F. Zhang), zlin@pku.edu.cn (Z. Lin).

http://dx.doi.org/10.1016/j.patcog.2015.04.017
0031-3203/

representation framework for spatial continuity of the occlusion [6]. Li et al. explored the intrinsic structure of contiguous occlusion and proposed a structured sparse error coding (SSEC) model [15]. These two works share the same two-step iteration strategy: (1) Meanwhile detecting errors via sparse representation or coding, and (2) That 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.

Some recent works point out that the visual data has low rank structure. Most of the exiting methods aim to find a low-rank approximation for matrix completion. However, the rank minimization problem is NP hard in general. In [17,18], 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 part) and the noise (sparse part) [19,20]. Zhang et al. introduced a matrix completion algorithm based on the Truncated Nuclear Norm Regularization for estimating missing values [21]. Ma et al. integrated rank minimization into sparse representation for dictionary learning and applied the model for face recognition [22]. Chen et al. presented a novel low-rank matrix approximation algorithm with structural incoherence for robust face recognition [23]. Zhang et al. proposed a novel image classification model to learn structured low-rank representation [37]. He et al. investigated the recovery of corrupted low-rank matrix via non-convex minimization and introduced a novel algorithm to solve this problem [38].

This paper focuses on face recognition with occlusion. We observe that contiguous occlusion in a face image generally leads to the error image with strong structure information, as shown in Fig. 1. And the error image is not sparse when there exist occlusions in test image [40]. Additionally, sparse based methods also 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. Meanwhile, characterizing the representation error pixel by pixel individually neglects the whole structure of the error image. Fortunately, nuclear norm not only can alleviate these correlations via the involved singular value decomposition (SVD) [41], but also directly characterizes the holistically structure of error image. Based on this, we add a nuclear norm of the representation residual image into a regression model. The model can be solved via the alternating direction method of multipliers (ADMM) [27]. 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 structure information.

(2) Compared with SSEC [15] and Then the Zhou's method [6], 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 the ADMM algorithm converges well with theoretical guarantee. In addition, our method can be used as a general face recognition algorithm. Our experiments will show that when there is no occlusion, our method still performs well, but SSEC cannot.

This paper is an extended version of our conference paper [32]. In this paper, we provide more in-depth analysis and more extensive experiments on the proposed model. The rest of the paper is organized as follows. Section 2 presents the nuclear norm regularized regression model and uses the ADMM to solve the model. Additionally, we also provide the complexity analysis and convergence analysis in this section. Section 3 introduces the robust nuclear norm regularized regression model for classification. Section 4 gives further analysis on the proposed method. Section 5 evaluates the performance of the proposed methods on several commonly used face recognition databases. Section 6 concludes our paper.

2. Nuclear norm regularized regression

In this section, we present the nuclear norm regularized regression model to code the image and use the alternating direction method of multipliers [27] to solve the model. Subsequently, we also provide the complexity analysis and convergence analysis of the proposed model.

2.1. Nuclear norm regularized regression

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

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

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

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

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

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

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

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

where η is a positive parameter. The above optimization problem can be solved in a closed form. For more details, please refer to [24].

Nuclear norm of a matrix is a good tool to describe the structural characteristics of an error image. However, the existing linear regression models do not make use of this kind of structural information. To address this problem, we introduce the nuclear norm regularization to the ridge regression model. Specifically, the

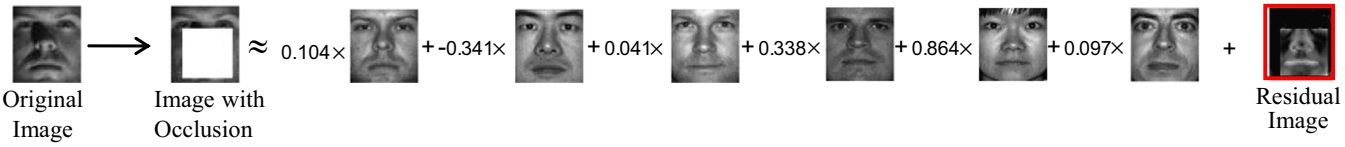

Fig. 1. The image with block occlusion is linearly represented by six different images and the residual (noise) image.

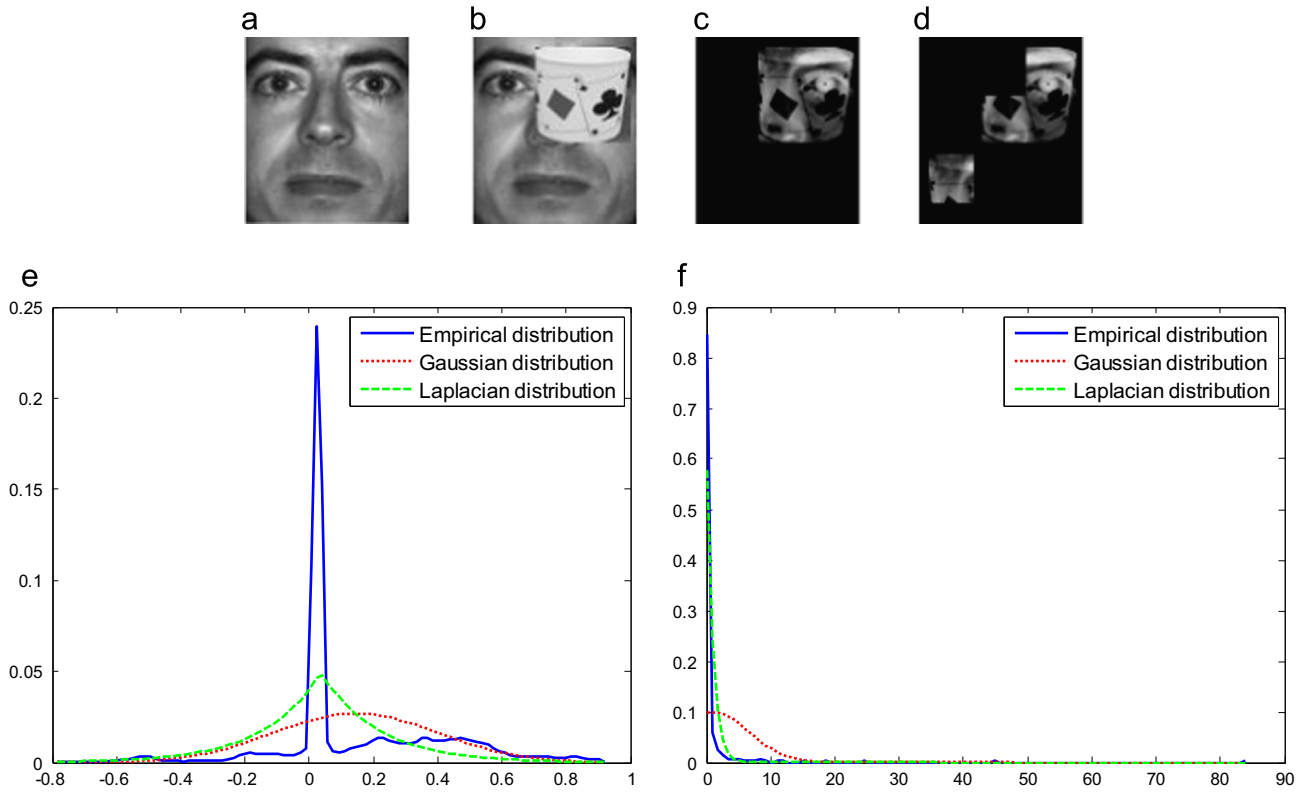

Fig. 2. (a) Original image; (b) observed image; (c) error image; (d) rearranged error image; (e) distributions of error image; and (f) distributions of singular values of error image (c).

optimization problem is formulated as follows:

$$\min_{\mathbf{x}} \|F(\mathbf{x})-\mathbf{Y}\|_F^2+\lambda\|F(\mathbf{x})-\mathbf{Y}\|_*+\frac{\eta}{2}\|\mathbf{x}\|_2^2, \tag{4}$$

where λ is a positive balance factor.

Meanwhile There are mainly two merits of using nuclear norm to describe the error image:

(1) Compared with L_2 norm and L_1 norm, nuclear norm can characterize structural error effectively. We give an example to support our point of view. In Fig. 2, (a) is a face image from Ext Yale B dataset. The image (a) is occluded by an unrelated block image as shown in (b). The error image between (a) and (b) is shown in (c). We rearrange pixels of image (c) as shown in (d). In previous work, L_2 norm and L_1 norm are usually used to measure the error image. However, these schemes ignore the structural information of error image. L_2 norm (or L_1 norm) of image (c) is same with image (d). It is difficult to distinguish the differences between (c) and (d). Fortunately, nuclear norm can characterize the structural information of error image well. For example, nuclear norm of images (c) and (d) are 82.04 and 96.56, respectively. So we believe that nuclear norm can achieve better performance than L_1 norm and L_2 norm in dealing with structural error information.

(2) From the distribution point of view, we can see that the distribution of an error image does not follow the Gaussian or Laplacian distribution in Fig. 2(e). In general, L_1 norm is the best option to describe the error image when the error image follows the Laplacian distribution, while L_2 norm is the best one when the error image follows the Gaussian distribution. So L_1 and L_2 norm cannot characterize this kind of occlusion effectively. In Fig. 2(f), the singular values of error image (c) fit the Laplacian distribution. In other words, nuclear norm can be considered as L_1 norm of singular value vector since nuclear norm is sum of singular values of error image matrix. Additionally, we also give another two examples to support our view. From Figs. 3 and 4, we can see that error images Fig. 3(c) and Fig. 4(c) are not sparse and the singular values of them still follow Laplacian distribution well. Based on this point, we believe that nuclear norm can perform better performance than L_1 (or L_2) norm to describe general structural error. Motivated by above observations, we use nuclear norm to characterize the error image.

In the following section, we will develop the optimization algorithm to solve Eq. (4) by using the alternating direction method of multipliers.

2.2. Optimization via ADMM

In this section, we adopt the alternating direction method of multipliers (ADMM) to solve Eq. (4) efficiently. For more details of ADMM, we refer readers to [27,34]. To deal with our problem, we rewrite the model in Eq. (4) as

$$\begin{aligned}&\min_{\mathbf{x},\mathbf{E}}\|\mathbf{E}\|_F^2+\lambda\|\mathbf{E}\|_*+\tfrac{1}{2}\eta\mathbf{x}^T\mathbf{x},\\ &s.t.\ \mathrm{F}(\mathbf{x})-\mathbf{Y}=\mathbf{E}.\end{aligned} \tag{5}$$

The augmented Lagrange function is given by

$$\begin{aligned}L_\mu(\mathbf{x},\mathbf{E},\mathbf{Z}) = &\|\mathbf{E}\|_F^2+\lambda\|\mathbf{E}\|_*+\tfrac{1}{2}\eta\mathbf{x}^T\mathbf{x}+\mathrm{Tr}\left(\mathbf{Z}^T(\mathrm{F}(\mathbf{x})-\mathbf{E}-\mathbf{Y})\right)\\ &+\tfrac{\mu}{2}\|\mathrm{F}(\mathbf{x})-\mathbf{E}-\mathbf{Y}\|_F^2,\end{aligned} \tag{6}$$

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

ADMM consists of the following iterations:

$$\mathbf{x}^{k+1}=\arg\min_{\mathbf{x}} L_\mu(\mathbf{x}), \tag{7}$$

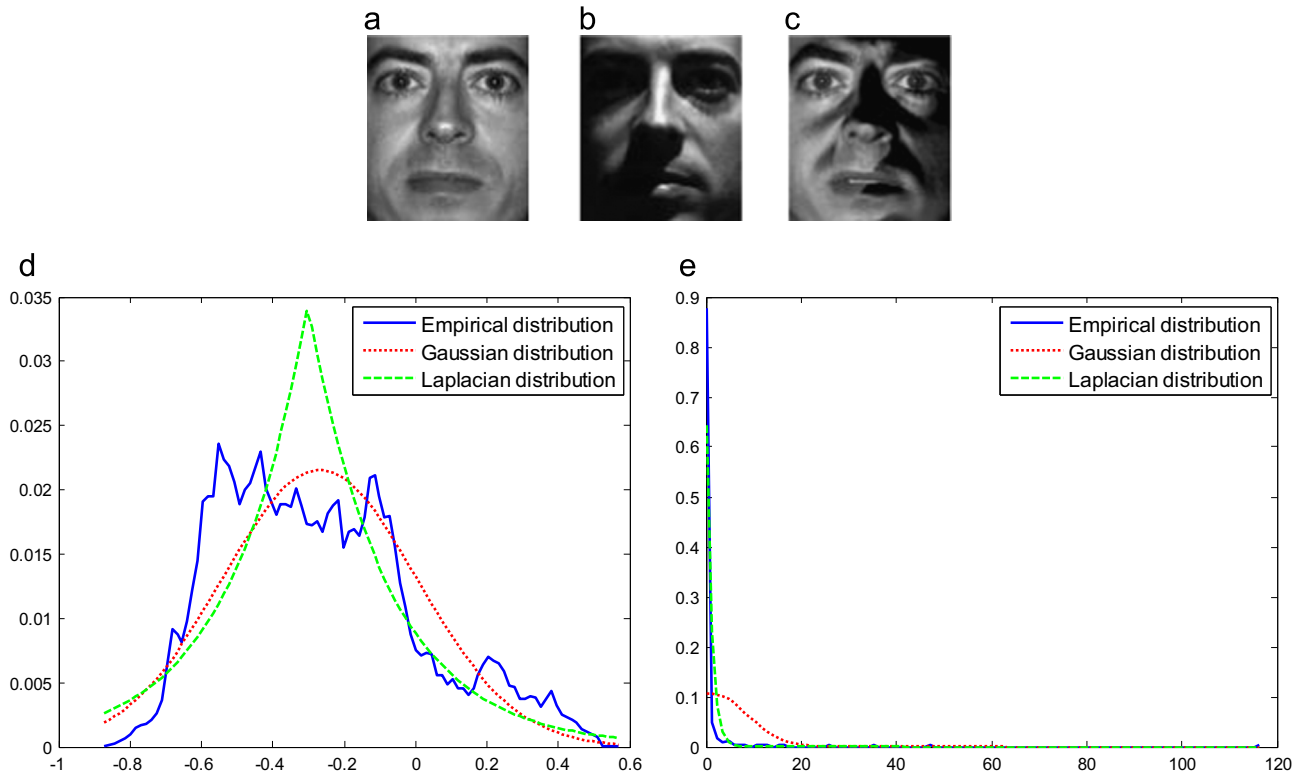

Fig. 3. (a) Original image; (b) observed image; (c) error image; (d) distributions of error image; and (e) distributions of singular values of error image (c).

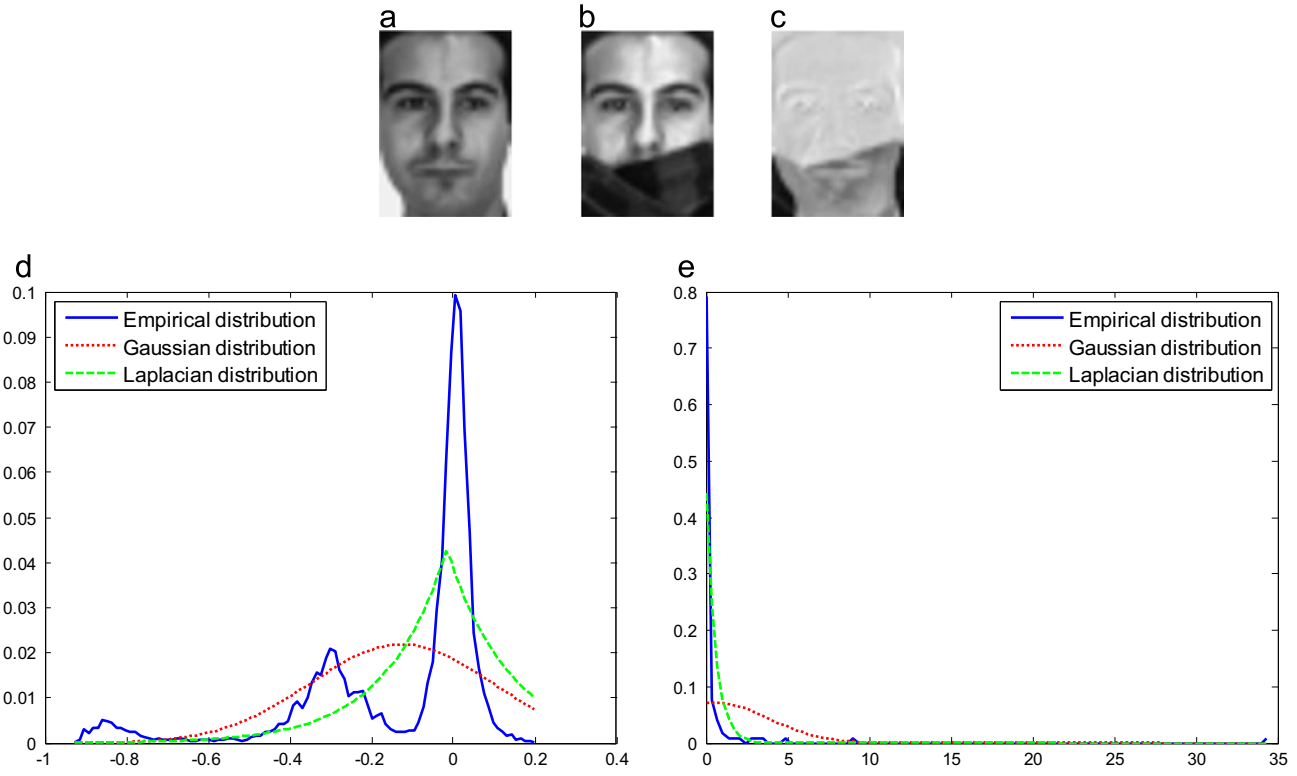

Fig. 4. (a) Original image; (b) observed image; (c) error image; (d) distributions of error image; and (e) distributions of singular values of error image (c).

$$\mathbf{E}^{k+1} = \arg\min_{\mathbf{E}} L_{\mu}(\mathbf{E}), \tag{8}$$

$$\mathbf{Z}^{k+1} = \mathbf{Z}^{k} + \mu(F(\mathbf{x}^{k+1}) - \mathbf{E}^{k+1} - \mathbf{Y}). \tag{9}$$

Updating x

Denote $\mathbf{H} = [Vec(\mathbf{A}_1), \cdots, Vec(\mathbf{A}_n)]$, $\mathbf{g} = Vec(E + Y - \frac{1}{\mu}\mathbf{Z})$, where Vec(·)convert matrix into a vector, then the objective function

$L_\mu(\mathbf{x})$ in Eq. (7) is equivalent to

$$\begin{aligned}L_\mu(\mathbf{x}) &= \eta\tfrac{1}{2}\mathbf{x}^T\mathbf{x}+\mathrm{Tr}\left(\mathbf{Z}^T F(\mathbf{x})\right)+\tfrac{\mu}{2}\|F(\mathbf{x})-\mathbf{E}-\mathbf{Y}\|_F^2\\ &= \tfrac{1}{2}\eta\mathbf{x}^T\mathbf{x}+\tfrac{\mu}{2}\mathrm{Tr}\Big(\mathrm{F}(\mathbf{x})^T\mathrm{F}(\mathbf{x})+(\mathbf{E}^T+\mathbf{Y}^T)(\mathbf{E}+\mathbf{Y})\\ &\quad -2(\mathbf{E}^T+\mathbf{Y}^T-\tfrac{1}{\mu}\mathbf{Z}^T)\mathrm{F}(\mathbf{x})\Big)\\ &= \tfrac{1}{2}\eta\mathbf{x}^T\mathbf{x}+\tfrac{\mu}{2}\|F(\mathbf{x})-(\mathbf{E}+\mathbf{Y}-\tfrac{1}{\mu}\mathbf{Z})\|_F^2\\ &\quad +\mathrm{Tr}\left((\mathbf{E}+\mathbf{Y})\mathbf{Z}^T-\tfrac{1}{2\mu}\mathbf{Z}\mathbf{Z}^T\right).\end{aligned} \tag{10}$$

Then the problem (7) can be reformulated as

$$\mathbf{x}^{k+1}=\arg\min_{\mathbf{x}}\left(\tfrac{\mu}{2}\|Hx-\mathbf{g}\|_2^2+\tfrac{1}{2}\eta\mathbf{x}^T\mathbf{x}\right). \tag{11}$$

Eq. (11) is actually a ridge regression model. So we can obtain the solution of Eq. (11) by

$$\mathbf{x}^{k+1}=(\mathbf{H}^T\mathbf{H}+\tfrac{\eta}{\mu}\mathbf{I})^{-1}\mathbf{H}^T\mathbf{g}. \tag{12}$$

Updating E

The objective function $L_\mu(\mathbf{E})$ in Eq. (8) can be rewritten as

$$\begin{aligned}L_\mu(\mathbf{E}) &= \|\mathbf{E}\|_F^2+\lambda\|\mathbf{E}\|_*-\mathrm{Tr}\left(\mathbf{Z}^T\mathbf{E}\right)+\tfrac{\mu}{2}\|F(\mathbf{x})-\mathbf{E}-\mathbf{Y}\|_F^2\\ &= \|\mathbf{E}\|_F^2+\lambda\|\mathbf{E}\|_*-\mathrm{Tr}\left(\mathbf{Z}^T\mathbf{E}\right)+\tfrac{\mu}{2}\mathrm{Tr}\left(\left((\mathrm{F}(\mathbf{x})-\mathbf{Y})^T-\mathbf{E}^T\right)(F(\mathbf{x})-\mathbf{E}-\mathbf{Y})\right)\\ &= \lambda\|\mathbf{E}\|_*+\tfrac{\mu}{2}\mathrm{Tr}\left((\tfrac{2}{\mu}+1)\mathbf{E}^T\mathbf{E}-2\left(\mathrm{F}(\mathbf{x})^T-\mathbf{Y}^T+\tfrac{1}{\mu}\mathbf{Z}^T\right)\mathbf{E}\right)+const_1\\ &= \lambda\|\mathbf{E}\|_*+\tfrac{\mu}{2}\tfrac{\mu+2}{\mu}\mathrm{Tr}\left(\mathbf{E}^T\mathbf{E}-2\tfrac{\mu}{2+\mu}\left(\mathbf{F}(\mathbf{x})^T-\mathbf{Y}^T+\tfrac{1}{\mu}\mathbf{Z}^T\right)\mathbf{E}\right)+const_1\\ &= \lambda\|\mathbf{E}\|_*+\tfrac{\mu+2}{2}\|\mathbf{E}-\tfrac{\mu}{2+\mu}\left(\mathrm{F}(\mathbf{x})^T-\mathbf{Y}^T+\tfrac{1}{\mu}\mathbf{Z}^T\right)^T\|_F^2+const_2,\end{aligned} \tag{13}$$

where $const_1$ and $const_2$ are constant terms, which are independent of the variable $\mathbf{E}$. The optimization problem Eq. (8) can be reformulated as

$$\mathbf{E}^{k+1}=\arg\min_{\mathbf{E}}\left(\tfrac{\lambda}{\mu+2}\|\mathbf{E}\|_*+\tfrac{1}{2}\|\mathbf{E}-\tfrac{\mu}{2+\mu}\left(\mathrm{F}(\mathbf{x})-\mathbf{Y}+\tfrac{1}{\mu}\mathbf{Z}\right)\|_F^2\right). \tag{14}$$

Its solution is [28]

$$\begin{aligned}&\mathbf{E}^{k+1}=\mathbf{U}T_{\frac{\lambda}{\mu+2}}[\mathbf{S}]\mathbf{V},\\ &where\ (\mathbf{U},\mathbf{S},\mathbf{V}^T)=svd\left(\tfrac{\mu}{2+\mu}\left(\mathrm{F}(\mathbf{x})-\mathbf{Y}+\tfrac{1}{\mu}\mathbf{Z}\right)\right).\end{aligned} \tag{15}$$

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

$$T_{\frac{\lambda}{\mu+2}}[\mathbf{S}]=\mathrm{diag}\left(\left\{\max(0,s_{jj}-\tfrac{\lambda}{\mu+2})\right\}_{1\le j\le r}\right), \tag{16}$$

where r is the rank of $\mathbf{S}$.

Stopping criterion

As suggested in [27], the stopping criterion of the algorithm is: the primal residual $r=\|F(\mathbf{x}^{k+1})-\mathbf{E}^{k+1}-\mathbf{Y}\|_F$ must be small: $\|F(\mathbf{x}^{k+1})-\mathbf{E}^{k+1}-\mathbf{Y}\|_F\le\varepsilon$, and the difference between successive iterations should also be small: $\max(\|\mathbf{x}^{k+1}-\mathbf{x}^k\|_F,\ \|\mathbf{E}^{k+1}-\mathbf{E}^k\|_F)\le\varepsilon$, where ε is a given tolerance.

Algorithm 1. Solving NR via ADMM

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

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

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

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

$$\mathbf{E}^{k+1}=\arg\min_{\mathbf{Y}}\tfrac{\lambda}{\mu+2}\|\mathbf{E}\|_*+\tfrac{1}{2}\|\mathbf{E}-\tfrac{\mu}{2+\mu}\left(\mathrm{F}(\mathbf{x})-\mathbf{Y}+\tfrac{1}{\mu}\mathbf{Z}\right)\|_F^2,$$

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

end while

Output: Optimal representation coefficient $\mathbf{x}$.

In summary, the pseudo code of our method to solve Eq. (5) is shown in Algorithm 1.

Algorithm 1 can be interpreted as using two-step iteration strategy for robust face recognition as those used in [6,15]. 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 detection step for determining the real occluded part. So we can say that NR provides a unified framework to integrate error detection and error support detection into one simple model.

2.3. Complexity analysis

Suppose that the training sample size is n and the image size is $p\times q$. The computational complexity of NR is mainly determined by the singular value decomposition and the matrix multiplications. For convenience, we assume that $q\le p$. Then the computational complexity for performing SVD on the $p\times q$ matrix $\mu/(2+\mu)(\mathrm{F}(\mathbf{x})-\mathbf{Y}+(1/\mu)\mathbf{Z})$ is $O(pq^2)$. The computational complexity of matrix multiplications is $O(npq+n^2)$. So the computational complexity of NR is $O(k(pq^2+npq+n^2))$, where k is the number of iterations.

2.4. Convergence analysis

In this subsection, we mainly investigate the convergence of the proposed Algorithm 1. Indeed, Algorithm 1 is a special case of augmented Lagrange multiplier algorithms (known as alternating directions methods) [27,34]. The convergence of these algorithms has been studied extensively. Stephen Boyd et al. investigated convergence of ADMM in [34] by using the properties of the saddle points, and gave three important results: residual convergence, objective convergence and dual variable convergence. He et al. [35,36] presented some significant convergence results by virtue of variational inequalities. Motivated by these techniques, we will present a convergence theorem which can point out the accumulation points of the iterative variables in Algorithm 1.

In the following, any solution of problem (5) is denoted by $(\mathbf{x}^*,\mathbf{E}^*)$. From standard theory of convex programming, there exists a $\mathbf{Z}^*$ such that the following conditions are satisfied:

$$\eta\mathbf{x}^*+\mathbf{H}^T vec(\mathbf{Z}^*)=0,\quad \mathbf{Z}^*\in\partial(\|\mathbf{E}\|_F+\lambda\|\mathbf{E}\|_*),\quad \mathrm{F}(\mathbf{x}^*)-\mathbf{Y}=\mathbf{E}^*. \tag{17}$$

Theorem 4.1. *Let $(\mathbf{E}^0,\mathbf{Z}^0)$ be an arbitrary initial point. Then for any fixed $\mu>0$, the sequence $\{(\mathbf{x}^k,\mathbf{E}^k,\mathbf{Z}^k)\}$ generated by Algorithm 1 converges to $(\mathbf{x}^*,\mathbf{E}^*,\mathbf{Z}^*)$.*

Proof. First, it is noted that the solutions of sub-problem (7) and (8) satisfy

$$\tfrac{1}{2}\eta\|\mathbf{x}^*\|_2^2-\tfrac{1}{2}\eta\|\mathbf{x}^{k+1}\|_2^2+(\mathbf{x}^*-\mathbf{x}^{k+1})^T(\mathbf{H}^T Vec(Z^k)+\mu\mathbf{H}^T Vec(F(\mathbf{x}^{k+1})-\mathbf{E}^k-\mathbf{Y}))\ge 0, \tag{18}$$

$$\|\mathbf{E}^*\|_F^2+\lambda\|\mathbf{E}^*\|_*-\|\mathbf{E}^{k+1}\|_F^2-\lambda\|\mathbf{E}^{k+1}\|_*+tr\left(\left(\mathbf{E}^*-\mathbf{E}^{k+1}\right)^T\right.$$

$$\left(-\mathbf{Z}^k-\mu\left(\mathrm{F}(\mathbf{x}^{k+1})-\mathbf{E}^{k+1}-\mathbf{Y}\right)\right)\geq 0, \tag{19}$$

respectively.

Meanwhile, sub-problem (9) can be written as

$$(\mathbf{Z}^*-\mathbf{Z}^{k+1})^T(-(F(\mathbf{x}^{k+1})-\mathbf{E}^{k+1}-\mathbf{Y})+\frac{1}{\mu}(\mathbf{Z}^{k+1}-\mathbf{Z}^k)) \tag{20}$$

Substituting $\mathbf{Z}^{k+1}=\mathbf{Z}^k+\mu(F(\mathbf{x}^{k+1})-\mathbf{E}^{k+1}-\mathbf{Y})$ into (18) and (19), we obtain

$$\frac{1}{2}\eta\|\mathbf{x}^*\|_2^2-\frac{1}{2}\eta\|\mathbf{x}^{k+1}\|_2^2+(\mathbf{x}^*-\mathbf{x}^{k+1})^T(\mathbf{H}^TZ^{k+1}+\mu\mathbf{H}^T(Vec(\mathbf{E}^{k+1})-Vec(\mathbf{E}^k)))\geq 0. \tag{21}$$

$$\|\mathbf{E}^*\|_F^2+\lambda\|\mathbf{E}^*\|_*-\|\mathbf{E}^{k+1}\|_F^2-\lambda\|\mathbf{E}^{k+1}\|_*+tr((\mathbf{E}^*-\mathbf{E}^{k+1})^T(-\mathbf{Z}^{k+1}))\geq 0. \tag{22}$$

For the sake of convenience, we introduce some notations $\mathbf{e}=Vec(\mathbf{E})$, $\mathbf{z}=Vec(\mathbf{Z})$, $\mathbf{y}=Vec(\mathbf{Y})$, $\mathbf{r}=(\mathbf{x};\mathbf{e})$, $\mathbf{s}=(\mathbf{x};\mathbf{e};\mathbf{z})$, $\mathbf{t}=(\mathbf{e};\mathbf{z})$, $u=p\times q$, $f(\mathbf{r}^*)=\frac{1}{2}\eta\|\mathbf{x}^*\|_2^2+\|\mathbf{E}^*\|_F^2+\lambda\|\mathbf{E}^*\|_*$, $f(\mathbf{r}^{k+1})=\frac{1}{2}\eta\|\mathbf{x}^{k+1}\|_2^2+\|\mathbf{E}^{k+1}\|_F^2+\lambda\|\mathbf{E}^{k+1}\|_*$.

Thus, adding (20)–(22) and considering $tr((\mathbf{E}^*-\mathbf{E}^{k+1})^T(\mathbf{Z}^{k+1}))=(\mathbf{e}^*-\mathbf{e}^{k+1})^T(\mathbf{z}^{k+1})$, we have

$$f(\mathbf{r}^*)-f\left(\mathbf{r}^{k+1}\right)+\left(\mathbf{s}^*-\mathbf{s}^{k+1}\right)^TV\left(\mathbf{s}^{k+1}\right)+\left(\mathbf{s}^*-\mathbf{s}^{k+1}\right)^T\kappa\left(\mathbf{e}^k,\mathbf{e}^{k+1}\right)\geq\left(\mathbf{t}^*-\mathbf{t}^{k+1}\right)^T\mathbf{M}\left(\mathbf{t}^k-\mathbf{t}^{k+1}\right) \tag{23}$$

where

$$V(\mathbf{s})=\begin{pmatrix}\mathbf{H}^T\mathbf{z}\\-\mathbf{z}\\-(\mathbf{H}\mathbf{x}-\mathbf{e}-\mathbf{y})\end{pmatrix},$$

$$\kappa\left(\mathbf{e}^k,\mathbf{e}^{k+1}\right)=\mu\begin{pmatrix}\mathbf{H}^T\\-\mathbf{I}_{u\times u}\\0\end{pmatrix}\left(\mathbf{e}^{k+1}-\mathbf{e}^k\right),$$

$$\mathbf{M}=\begin{pmatrix}\mu\mathbf{I}_{u\times u}&0\\0&\frac{1}{\mu}\mathbf{I}_{u\times u}\end{pmatrix}.$$

Then, we prove that

$$\left(\mathbf{t}^{k+1}-\mathbf{t}^*\right)^T\mathbf{M}\left(\mathbf{t}^k-\mathbf{t}^{k+1}\right)\geq 0. \tag{24}$$

Using (23), we have

$$\left(\mathbf{t}^{k+1}-\mathbf{t}^*\right)^T\mathbf{M}\left(\mathbf{t}^k-\mathbf{t}^{k+1}\right)\geq\left(\mathbf{s}^{k+1}-\mathbf{s}^*\right)^T\kappa\left(\mathbf{e}^k,\mathbf{e}^{k+1}\right)+f\left(\mathbf{r}^{k+1}\right)-f(\mathbf{r}^*)+\left(\mathbf{s}^{k+1}-\mathbf{s}^*\right)^TV\left(\mathbf{s}^{k+1}\right). \tag{25}$$

Since $V(\mathbf{s})$ is monotone, it follows that

$$f\left(\mathbf{r}^{k+1}\right)-f(\mathbf{r}^*)+\left(\mathbf{s}^{k+1}-\mathbf{s}^*\right)^TV\left(\mathbf{s}^{k+1}\right)\geq f\left(\mathbf{r}^{k+1}\right)-f(\mathbf{r}^*)+\left(\mathbf{s}^{k+1}-\mathbf{s}^*\right)^TV(\mathbf{s}^*)\geq 0, \tag{26}$$

The last inequality is due to the property of the optimal solution.

Combining (25) with (26), we have

$$\left(\mathbf{t}^{k+1}-\mathbf{t}^*\right)^T\mathbf{M}\left(\mathbf{t}^k-\mathbf{t}^{k+1}\right)\geq\left(\mathbf{s}^{k+1}-\mathbf{s}^*\right)^T\kappa\left(\mathbf{e}^k,\mathbf{e}^{k+1}\right). \tag{27}$$

Furthermore, we have

$$\begin{aligned}\left(\mathbf{s}^{k+1}-\mathbf{s}^*\right)^T\kappa\left(\mathbf{e}^k,\mathbf{e}^{k+1}\right)&=\left(\mathbf{e}^{k+1}-\mathbf{e}^k\right)^T\mu\left[\left(\mathbf{H}\mathbf{x}^{k+1}-\mathbf{e}^{k+1}\right)-(\mathbf{H}\mathbf{x}^*-\mathbf{e}^*)\right]\\&=\left(\mathbf{e}^{k+1}-\mathbf{e}^k\right)^T\mu\left\{\left(\mathbf{H}\mathbf{x}^{k+1}-\mathbf{e}^{k+1}\right)-\mathbf{y}\right\}\\&=\left(\mathbf{z}^{k+1}-\mathbf{z}^k\right)^T\left(\mathbf{e}^{k+1}-\mathbf{e}^k\right).\end{aligned} \tag{28}$$

In (22), by replacing $\mathbf{E}^*$ with any $\mathbf{E}$, and considering the previous iteration, we obtain that

$$\|\mathbf{E}\|_F^2+\lambda\|\mathbf{E}\|_*-\|\mathbf{E}^{k+1}\|_F^2-\lambda\|\mathbf{E}^{k+1}\|_*+tr\left(\left(\mathbf{E}-\mathbf{E}^{k+1}\right)^T\left(-\mathbf{Z}^{k+1}\right)\right)\geq 0. \tag{29}$$

$$\|\mathbf{E}\|_F^2+\lambda\|\mathbf{E}\|_*-\|\mathbf{E}^k\|_F^2-\lambda\|\mathbf{E}^k\|_*+tr\left(\left(\mathbf{E}-\mathbf{E}^k\right)^T\left(-\mathbf{Z}^k\right)\right)\geq 0 \tag{30}$$

Set $\mathbf{E}=\mathbf{E}^k$ in (29) and $\mathbf{E}=\mathbf{E}^{k+1}$ in (30), respectively, and then adding the them, we have

$$\left(\mathbf{z}^k-\mathbf{z}^{k+1}\right)^T\left(\mathbf{e}^k-\mathbf{e}^{k+1}\right)\geq 0. \tag{31}$$

Therefore, (28) follows (31), (32), (35).

Let $\|\mathbf{t}-\mathbf{t}'\|_M^2=(\mathbf{t}-\mathbf{t}')^T\mathbf{M}(\mathbf{t}-\mathbf{t}')=\mu\|\mathbf{e}-\mathbf{e}'\|^2+\frac{1}{\mu}\|\mathbf{z}-\mathbf{z}'\|^2$. Then we have

$$\begin{aligned}\|\mathbf{t}^k-\mathbf{t}^*\|_M^2&=\|\left(\mathbf{t}^{k+1}-\mathbf{t}^*\right)+\left(\mathbf{t}^k-\mathbf{t}^{k+1}\right)\|_M^2\\&=\|\mathbf{t}^{k+1}-\mathbf{t}^*\|_M^2+2\left(\mathbf{t}^{k+1}-\mathbf{t}^*\right)^TM\left(\mathbf{t}^k-\mathbf{t}^{k+1}\right)+\|\mathbf{t}^k-\mathbf{t}^{k+1}\|_M^2\\&\geq\|\mathbf{t}^{k+1}-\mathbf{t}^*\|_M^2+\|\mathbf{t}^k-\mathbf{t}^{k+1}\|_M^2.\end{aligned}$$

That is

$$\|\mathbf{t}^k-\mathbf{t}^*\|_M^2-\|\mathbf{t}^{k+1}-\mathbf{t}^*\|_M^2\geq\|\mathbf{t}^k-\mathbf{t}^{k+1}\|_M^2.$$

Thus, $\|\mathbf{t}^k-\mathbf{t}^{k+1}\|_M^2\to 0$, i.e., (32)

$\mathbf{e}^{k+1}-\mathbf{e}^k\to 0$ and $\mathbf{z}^{k+1}-\mathbf{z}^k\to 0$.

Further, considering $\mathbf{Z}^k=\mathbf{Z}^{k-1}+\mu(F(\mathbf{x}^k)-\mathbf{E}^k-\mathbf{Y})$, we obtain

$$F(\mathbf{x}^k)-\mathbf{E}^k-\mathbf{Y}\to\mathbf{0}. \tag{33}$$

Meanwhile, (32) also implies $\{\mathbf{t}^k\}$ lies in a compact region. Thus, it has a subsequence $\{\mathbf{t}^{k_j}\}$ converging to $\mathbf{t}^\bullet=(\mathbf{e}^\bullet;\mathbf{z}^\bullet)$, i.e., $\mathbf{e}^{k_j}\to\mathbf{e}^\bullet$ and $\mathbf{z}^{k_j}\to\mathbf{z}^\bullet$. In addition, from (12) we have

$$\mathbf{x}^k=(\mathbf{H}^T\mathbf{H}+\tfrac{\eta}{\mu}\mathbf{I})^{-1}\mathbf{H}^TVec(\mathbf{E}^{k-1}+Y-\tfrac{1}{\mu}\mathbf{Z}^{k-1}). \tag{34}$$

Thus $\mathbf{x}^{k_j}\to\mathbf{x}^\bullet=(\mathbf{H}^T\mathbf{H}+\frac{\eta}{\mu}\mathbf{I})^{-1}\mathbf{H}^T(\mathbf{e}^\bullet+\mathbf{b}-\frac{1}{\mu}\mathbf{z}^\bullet)$, as $j\to\infty$.

We transform $(\mathbf{x}^\bullet;\mathbf{e}^\bullet;\mathbf{z}^\bullet)$ back into its original form $(\mathbf{x}^\bullet,\mathbf{E}^\bullet,\mathbf{Z}^\bullet)$. Then by (32), $(\mathbf{x}^\bullet,\mathbf{E}^\bullet,\mathbf{Z}^\bullet)$is a limit point of $\{(\mathbf{x}^k,\mathbf{E}^k,\mathbf{Z}^k)\}$.

Next, we show that $(\mathbf{x}^\bullet,\mathbf{E}^\bullet,\mathbf{Z}^\bullet)$ satisfies the optimality conditions in (17). First, we take the equivalent form of (12)

$$\eta\mathbf{x}^{k+1}=-\mu\mathbf{H}^T\left(\tfrac{1}{\mu}\mathbf{z}^{k+1}+\mathbf{e}^{k+1}-\mathbf{e}^k\right). \tag{35}$$

By taking the limit of the above equality over k_j, it follows that:

$$\eta\mathbf{x}^\bullet=-\mu\mathbf{H}^T\mathbf{z}^\bullet. \tag{36}$$

Second, from (33), it easy to see that

$$F(\mathbf{x}^\bullet)-\mathbf{E}^\bullet-Y=0. \tag{37}$$

By (14), we know that $\mu(\mathbf{E}^{k+1}-\mathbf{E}^k)+\mathbf{Z}^{k+1}\in\partial\left(\|\mathbf{E}\|_F^2+\lambda\|\mathbf{E}\|_*\right)$. Since $\mathbf{E}^{k_j+1}-\mathbf{E}^{k_j}\to 0$ and $\mathbf{Z}^{k_j+1}\to\mathbf{Z}^\bullet$, we have

$$\mathbf{Z}^\bullet\in\partial\left(\|\mathbf{E}\|_F^2+\lambda\|\mathbf{E}\|_*\right), \tag{38}$$

which in conjunction with (36), (37) imply that $(\mathbf{x}^\bullet,\mathbf{E}^\bullet,\mathbf{Z}^\bullet)$ satisfies the optimality conditions (17). We now have shown that any limit point of $\left\{\left(\mathbf{x}^k,\mathbf{E}^k,\mathbf{Z}^k\right)\right\}$ is an optimal solution of problem (5).

Since (32) holds for any optimal solution of problem (5), by letting $(\mathbf{x}^*,\mathbf{E}^*)=(\mathbf{x}^\bullet,\mathbf{E}^\bullet)$ at the beginning and considering (32), we obtain the convergence of $\left\{\left(\mathbf{x}^k,\mathbf{E}^k,\mathbf{Z}^k\right)\right\}$.

3. Classification based on robust nuclear norm regularized regression

3.1. Robust sparse coding

In CRC and SRC, the representation residual is measured by the L_2-norm or L_1-norm of the error image. Such models inherently assume that the error image follows Gaussian or Laplacian distribution. However, the distribution of error image is more complicated in real-world applications. To this end, Yang et al. borrowed the idea of robust regression and proposed a robust sparse coding based classification (RSC) method [7]. RSC is more robust to outliers (occlusion and corruption, etc.) than SRC since it introduces the weight matrix for image pixels motivated by the robust regression theory. The RSC model is

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{W}^{1/2}(\mathbf{y} - A\mathbf{x})\|_2^2 + \eta\|\mathbf{x}\|_1, \tag{39}$$

where $\mathbf{W}$ is a weight matrix, $\mathbf{y}$ is the test sample and $\mathbf{A}$ is the dictionary. RSC is solved by using the iteratively reweighted sparse coding algorithm. The remaining steps of RSC are the same as SRC.

3.2. Robust nuclear norm regularized regression

We notice that SSEC [15] adopts a robust sparse representation model, i.e. iteratively reweighted sparse coding in the error detection step, but our nuclear norm regularized regression (NR) only uses a simple ridge regression model for updating $\mathbf{x}$. In real-world applications, however, it is difficult to preserve the good performance for most methods when facing with complicated distributions of error image (e.g. the test image with mixture of different types of noises). In Fig. 5(a) and (b), we give a simple example to demonstrate the performance of NR and RSC. In this example, we just select 76 face images of four persons from Subsets 1 and 2 of Extended Yale B database as training samples. The face image with mixture noises (pixel corruption and image occlusion) is used for testing. We represent the test image using the training samples via NR and RSC, respectively. The resulting reconstructed image and error image are shown in Fig. 5(a) and (b). From Fig. 5(a) and (b), we can see that NR and RSC have their own virtues and disadvantages. NR is good at recovering the structural noise (e.g. illumination) but loses the detail features. However, RSC is adaptive to recover the facial features but loses some structural information. The reconstructed image of RSC is more similar to the original image than that of NR. Additionally, it is insufficient if only use nuclear-norm and L_2-norm to constrain the error image. So we want to combine the advantages of nuclear norm regularization and robust regression to handle the complicated distribution of error image and further improve the classification performance of our model.

Based on above intuitions, in this section we borrow the idea of robust sparse coding to our model and give the MLE (maximum likely estimation) solution of representation coefficients to construct a more robust model to handle face recognition with occlusion. Motivated by the work [7], the robust regularized regression model can be formulated as

$$\min_{\mathbf{x}} \|\mathbf{W} \circ (F(\mathbf{x}) - Y)\|_F^2 + \frac{\eta}{2}\|\mathbf{x}\|_2^2 \tag{40}$$

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

However, in many cases of occlusion, the performance of the above model is limited. For example, in the black scarf caused occlusion part, pixel values are zeros. So, the ideal representation errors in the occluded part are correlated, because pixels in a local area in a real-world image are generally highly-correlated. Moreover, pixels in a local area are still correlated after the weight is assigned on each pixel of error image. In other words, the above model ignores the structural information of error image. Based on above analysis, we introduce the nuclear norm constraint term:

$$\min_{\mathbf{x}} \|\mathbf{W} \circ (F(\mathbf{x}) - Y)\|_F^2 + \frac{\eta}{2}\|\mathbf{x}\|_2^2 \quad \text{s.t.} \|\mathbf{W} \circ (F(\mathbf{x}) - Y)\|_* \le \tau \tag{41}$$

where τ is a parameter. However, we prefer to solve problem (42) in Lagrangian form, i.e.

$$\min_{\mathbf{x}} \|\mathbf{W} \circ (F(\mathbf{x}) - Y)\|_F^2 + \lambda\|\mathbf{W} \circ (F(\mathbf{x}) - Y)\|_* + \frac{\eta}{2}\|\mathbf{x}\|_2^2, \tag{42}$$

where $\lambda > 0$ is a parameter. From optimization theory, it is well known that problems (41) and (42) are equivalent in the sense that solving one will determine a parameter value in the other so that the two share the same solution.

The robust nuclear norm regularized regression model can be solved by using the iteratively reweighted algorithm. Each iteration step is to solve a nuclear norm 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 residual $\mathbf{E}$ 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

$$\mathbf{W}_{i,j} = \frac{\exp(\alpha\beta - \alpha(\mathbf{E}_{i,j})^2)}{1 + \exp(\alpha\beta - \alpha(\mathbf{E}_{i,j})^2)}, \tag{43}$$

where α and β are positive scalars.

Based on the optimization solution $\mathbf{x}$ via the iterative process, we obtain a weighted dictionary $\mathbf{B} = [\mathbf{B}_1, \ldots, \mathbf{B}_n]$, where $\mathbf{B}_i = \mathbf{W} \circ \mathbf{D}_i$, $i = 1, \ldots, n$ 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.

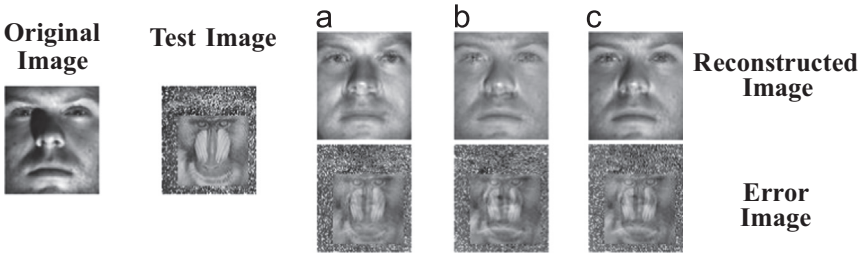

Fig. 5. Example for dealing with complex noise: (a) NR, (b) RSC, and (c) RNR.

The corresponding reconstruction error of i-th class is defined as

$$r_i(\mathbf{y}) = \|\mathbf{Y} - \hat{\mathbf{Y}}_i\|_* \tag{44}$$

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

Algorithm 2. RNR for classification

Input: Dictionary $\mathbf{D}$, test sample $\mathbf{Y}$. Initial values $\mathbf{Y}_{ini}$.

1. $\mathbf{Y}^t$ is initialized as $\mathbf{Y}_{ini}$. $\mathbf{Y}$ is initialized as $\mathbf{Y}_0$.
2. The test sample $\mathbf{Y}$ is coded by the dictionary $\mathbf{D}$.
   a) Compute residual $\mathbf{E}^{(t)} = \mathbf{Y} - \mathbf{Y}^{(t)}$.
   b) Estimate weights
   $$\mathbf{W}_{i,j} = \frac{\exp(\alpha\beta - \alpha(\mathbf{E}_{i,j})^2)}{1+\exp(\alpha\beta - \alpha(\mathbf{E}_{i,j})^2)}.$$
   c) $\mathbf{B}_i = \mathbf{W} \circ \mathbf{D}_i, i = 1, \cdots, n$, $\mathbf{Y} = \mathbf{W} \circ \mathbf{Y}_0$.
   d) Code using Algorithm 1
   $$\mathbf{x}^* = \arg\min_{\mathbf{x}} \|G(\mathbf{x}) - \mathbf{Y}\|_F^2 + \lambda\|G(\mathbf{x}) - \mathbf{Y}\|_* + \frac{\eta}{2}\|\mathbf{x}\|_2^2,$$
   where $G(\mathbf{x}) = \sum_{i=1}^n x_i \mathbf{B}_i$.
   e) Compute the reconstructed test sample $\mathbf{Y}^{(t)} = \sum_{i=1}^n x_i^{(t)} \mathbf{B}_i$, and let $t = t+1$.
   f) Go back to step (a) until the maximal number of iterations is reached, or convergence criterion shown in Eq. (45) is met.
3. Compute the residual of each class.

Output: $\mathbf{Y}$ is assigned to the class which yields the minimum residual.

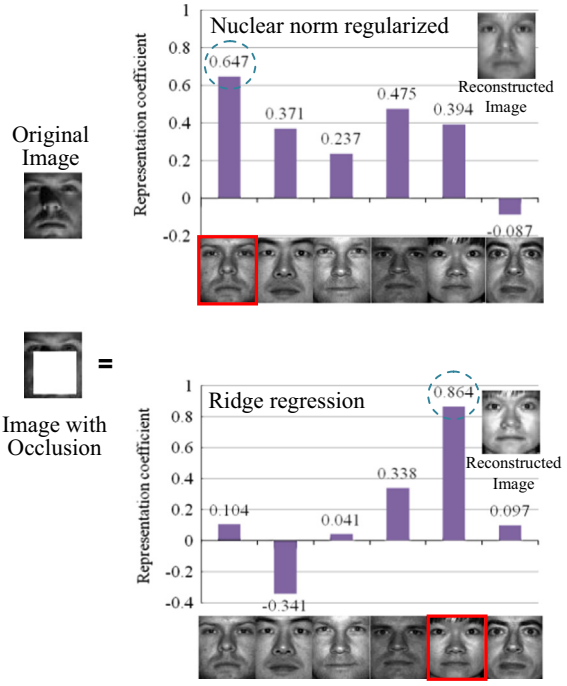

Fig. 6. The example shows the comparison between NR and ridge regression.

The RNR algorithm for classification is summarized in Algorithm 2. Finally, we perform the same test with NR. Fig. 2 (c) shows the resulted reconstructed images and error images. From Fig. 2(c), we can see that RNR not only preserves the advantage of NR, but also take into account the merits of robust regression. The reconstructed image of RNR is significantly better than NR and RSC as we desired.

To guarantee the convergence of Algorithm 2, we employ the standard line-search process [39] to choose a proper $v^{(t)}$ for updating representation coefficients in each step, where t is the iterative number. If $t = 1$, the $\mathbf{x}^{(1)} = \mathbf{x}^*$; if $t \geq 1$, $\mathbf{x}^{(t)} = \mathbf{x}^{(t-1)} + v^{(t)}(\mathbf{x}^* - \mathbf{x}^{(t-1)})$, where $v^{(t)} \in (0, 1]$ is suitable step size that makes $\|G(\mathbf{x}^t) - \mathbf{Y}\|_F^2 + \lambda\|G(\mathbf{x}^t) - \mathbf{Y}\|_* + \frac{\eta}{2}\|\mathbf{x}^t\|_2^2 < \|G(\mathbf{x}^{t-1}) - \mathbf{Y}\|_F^2 + \lambda\|G(\mathbf{x}^{t-1}) - \mathbf{Y}\|_* + \frac{\eta}{2}\|\mathbf{x}^{t-1}\|_2^2$. In each iteration, the objective function value of Eq. (42) decreases by Algorithm 2. Since the original cost function Eq. (42) is lower-bounded (Eq. (42) ≥ 0), the iterative minimization procedure in Algorithm 2 will converge.

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

$$\|\mathbf{W}^{(t)} - \mathbf{W}^{(t-1)}\|_2 / \|\mathbf{W}^{(t-1)}\|_2 < \gamma. \tag{45}$$

4. Further analysis on RNR

Compared with the existing regularized coding methods [2,4,7,9,14], the proposed method RNR can make use of the structural characteristics of the noise image well via nuclear norm. So in the case of contiguous occlusion, it can yield better reconstruction results.

To further analyze the proposed model, we give two examples here. In the first example, we select six different face images from the Extended Yale B database to linearly represent the face image with block occlusion via ridge regression and nuclear norm regularized regression, respectively. Fig. 6 shows the comparative results between ridge regression and nuclear norm regularized regression. From Fig. 6, we can see that NR can achieve the right results for classification while ridge regression fails. Additionally, the reconstructed image of NR is still similar to the target image. However, the reconstructed image of ridge regression is more similar to that of another person.

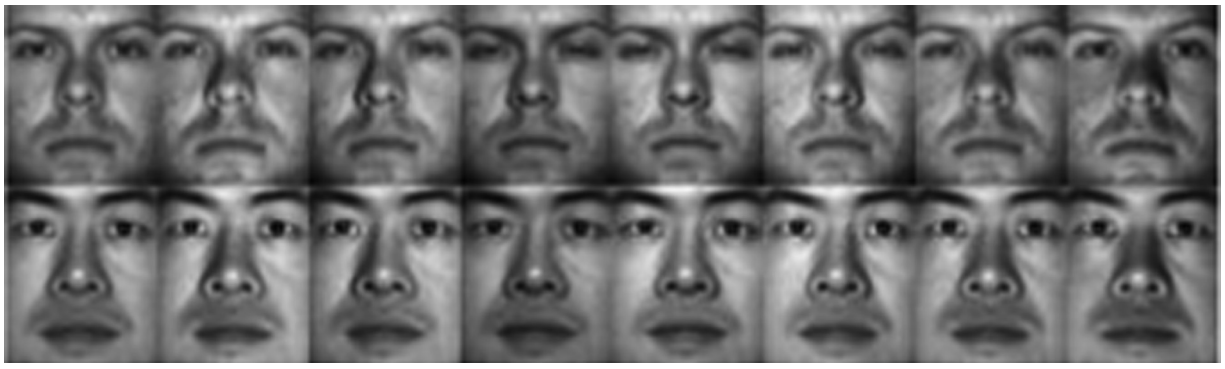

Fig. 7. Two classes of samples from the Extended Yale B database.

In the second example, we select two classes of face images from the Extended Yale B database as shown in Fig. 7. In our test, there are two cases of block occlusion: the images with white image and the images with an unrelated image. In our test, RNR, RSC, SRC and CRC are employed to deal with the occlusion. For each occluded image, the reconstructed images (recovered clean image) and the representation error image (recovered occlusion) are shown in Fig. 8. From Fig. 8, we can observe that the reconstruction performance of RSC is unsatisfactory when the test image has the white block occlusion. However, RNR still gives better results than other methods.

5. Experiments

In this section, we compare the proposed methods NR and RNR with CRC, SRC, CESR, SSEC and RSC. In our experiments, there are five parameters of the proposed RNR. The parameters α and β in Eq. (43) follows the suggestion in [7]. The default value of the penalty parameter μ is 1. Both the balance factor λ and the regularized parameter η are introduced in the respective experiments.

5.1. Face recognition with real disguise

The AR face database [29] contains over 4000 color face images of 126 persons, including frontal views of faces with different facial expressions, lighting conditions and occlusions. The images of 120 individuals were taken in two sessions (separated by two weeks) and each session contains 13 color images.

In our experiments, we only use a subset of AR face image database. The subset contains 100 individuals, 50 males and 50 females. All the individuals have two session images and each

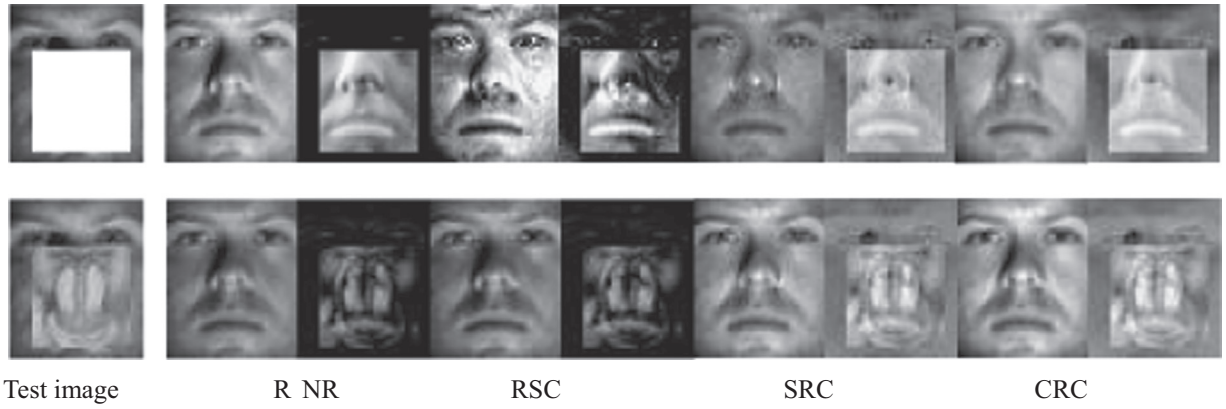

Fig. 8. Recovered clean images and occluded parts via four methods for images with white block images or unrelated block images.

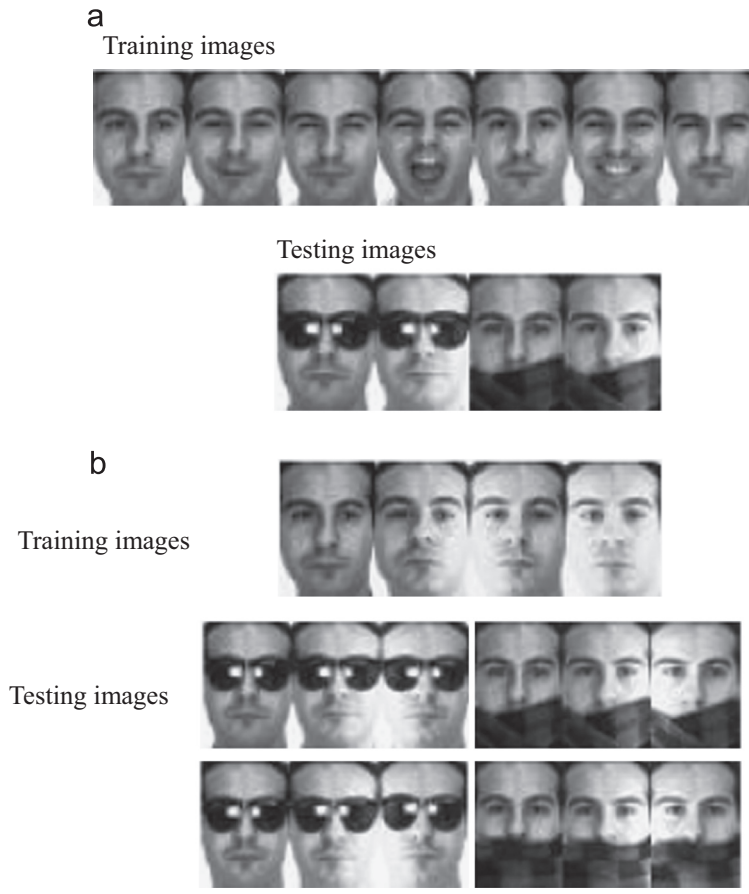

Fig. 9. Sample images for one person in the AR database. (a) Sample images in the first experiment. (b) Sample images in the second experiment.

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
NR	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
RNR	***99.0***	***100***

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

Methods	Sunglasses		Scarves	
	Session 1	Session 2	Session 1	Session 2
CRC	61.3	26.3	56.3	37.0
NR	75.7	38.3	72.0	45.3
SRC [2]	89.3	57.3	32.3	12.7
CESR [9]	95.3	79.0	38.0	20.7
SSEC	95.3	72.0	89.7	75.3
RSC [7]	94.7	80.3	91.0	72.7
RNR	***97.7***	***82.3***	***95.0***	***77.3***

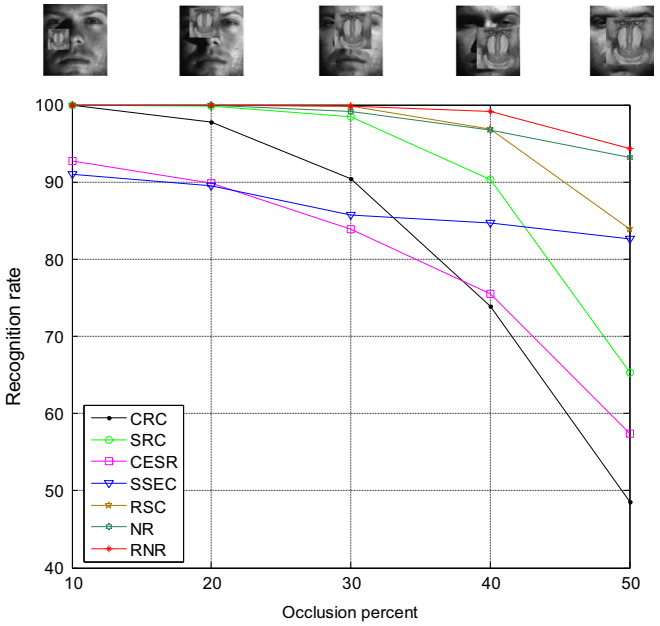

Fig. 10. The recognition rates (%) of CRC, SRC, CESR, SSEC, RSC, NR and RNR with the occlusion (with unrelated block image) percentage ranging from 0 to 50.

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 sessions 1 and 2 of each individual to form the training set. The total number of 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 sample images of one person are shown in Fig. 9(a). The balance factor λ is 10^2 and

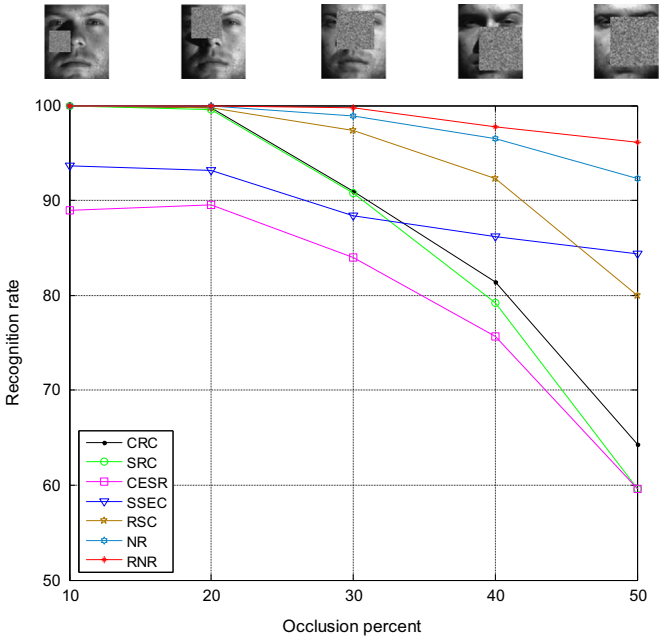

Fig. 11. The recognition rates (%) of CRC, SRC, CESR, SSEC, RSC, NR and RNR with the occlusion (with noise block image) percentage ranging from 0 to 50.

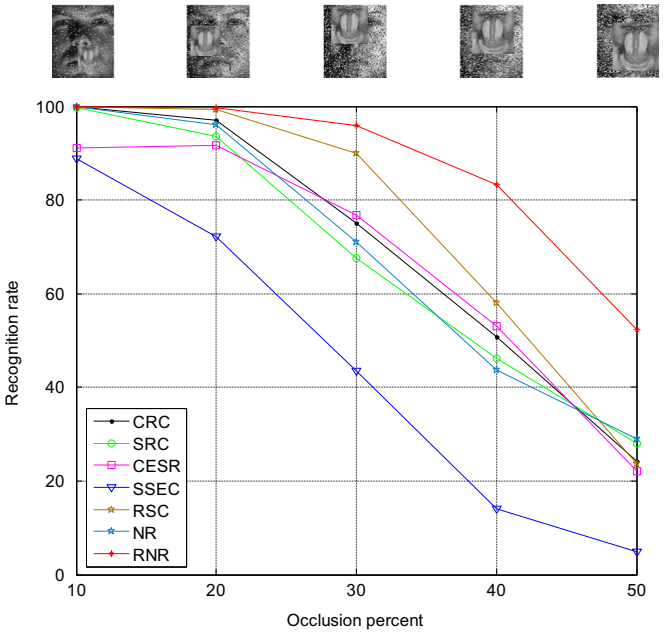

Fig. 12. The recognition rates (%) of CRC, SRC, CESR, SSEC, RSC, NR and RNR with the occlusion (with mixture noise) percentage ranging from 0 to 50.

Fig. 13. Sample images from the FRGC database.

10^{-1} for the test images with sunglasses and scarves, respectively. The regularization parameter η is 4×10^4. Table 1 lists the recognition rates of CRC, SRC, CESR, SSEC, RSC, NR and RNR. From Table 1, we can see that RNR achieves the best performance among all the methods. NR also gives better results than CRC. Both RSC and CESR obtain the same results as RNR when the test images are with sunglasses. However, the results of SRC and CESR are significantly lower than those of RNR 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 sessions 1 and 2 are for testing. The sample images of one person are shown in Fig. 9(b). 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 RNR significantly outperforms CRC, NR, 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 RNR on four different testing sets.

Table 3
The recognition rates (%) of each classifier for face recognition on the FRGC database.

Image sizes	16×16	32×32
CRC	90.2	92.2
SRC	88.6	89.2
CESR	79.1	81.9
SSEC	60.0	70.5
RSC	89.9	92.0
NR	91.3	93.5
RNR	***91.4***	***94.1***

5.2. Face recognition with random block occlusion

The extended Yale B face image database [31] contains 38 human subjects under 9 poses and 64 illumination conditions. The 64 images of a subject in a particular pose are acquired at a camera frame rate of 30 frames per second. So there are only small changes in head poses and facial expressions for those 64 images. 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 RNR. 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. 10 plots recognition rates of CRC, SRC, CESR, SSEC, RSC, NR and RNR under different levels of occlusions (from 10% to 50%). With the increment of the level of occlusion, RNR begins to significantly

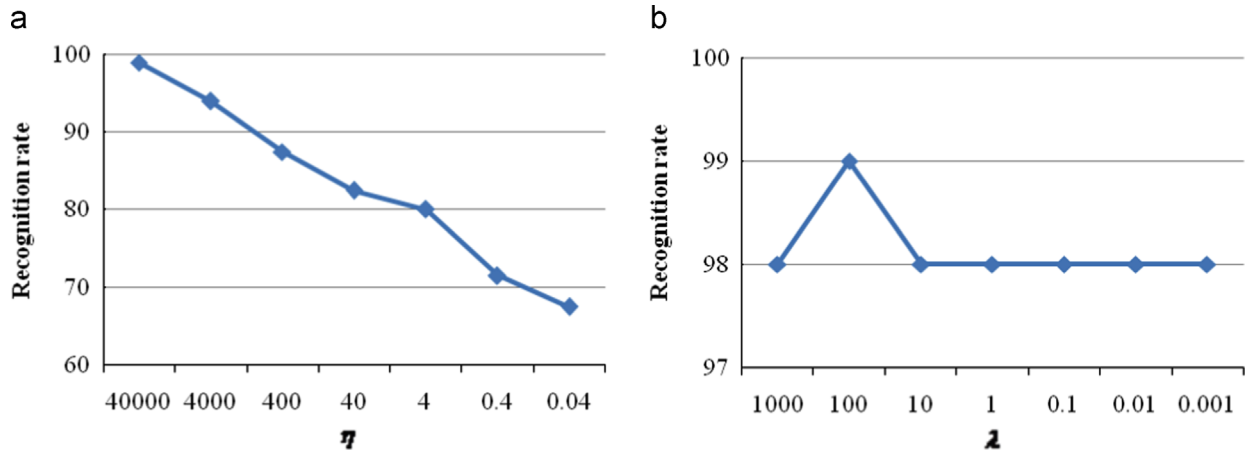

Fig. 14. The recognition rates (%) of RNR with different parameters on the AR database with sunglasses. (a) Regularization parameter and (b) balance factor.

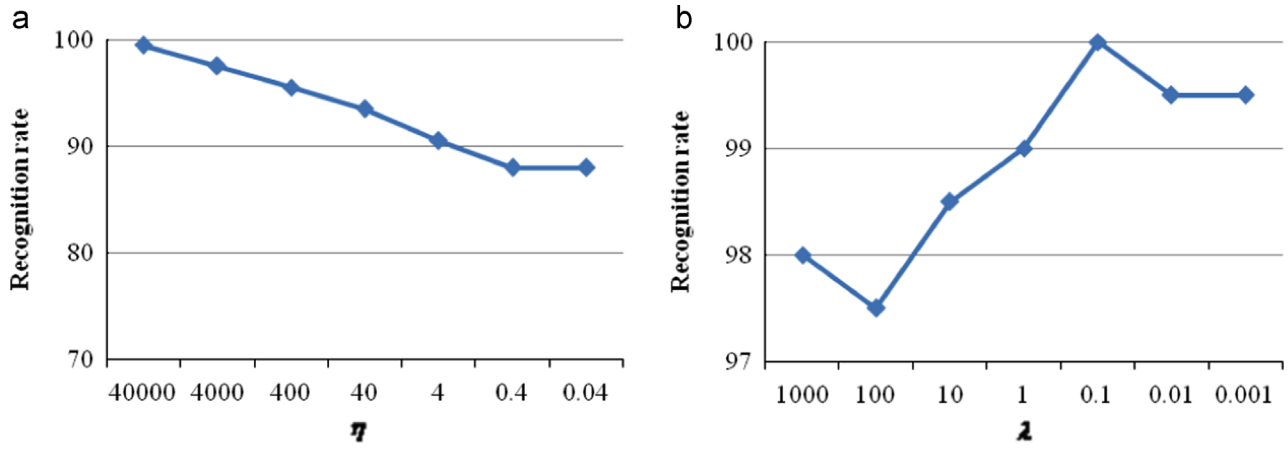

Fig. 15. The recognition rates (%) of RNR with different parameters on the AR database with scarf. (a) Regularization parameter and (b) balance factor.

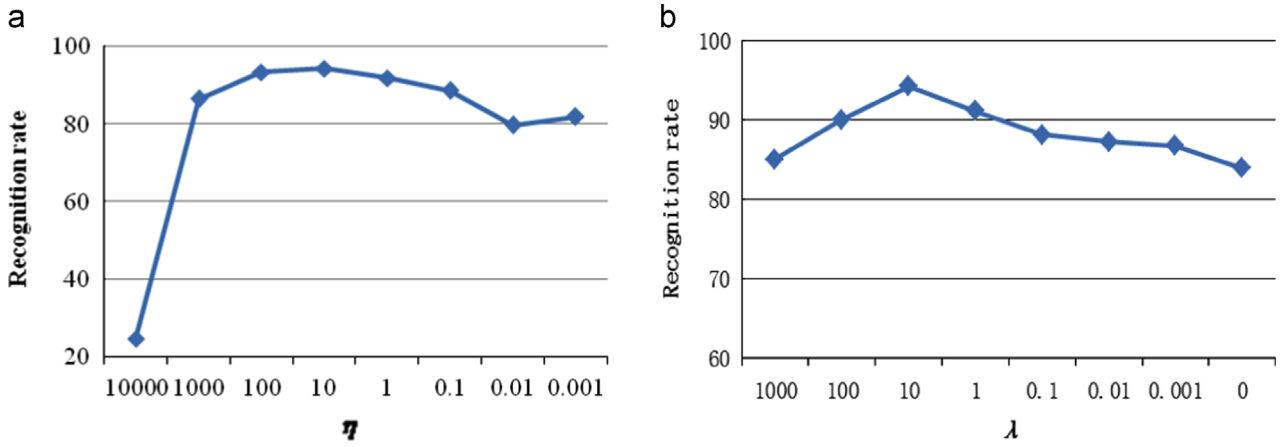

Fig. 16. The recognition rates (%) of RNR with different parameters on the Extended Yale B database with unrelated image occlusion. (a) Regularization parameter and (b) balance factor.

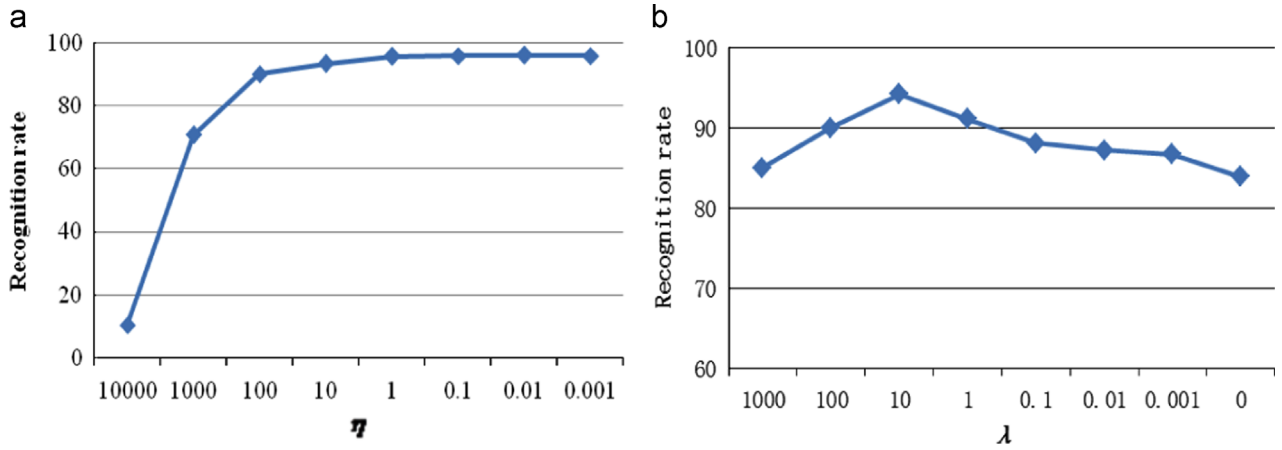

Fig. 17. The recognition rates (%) of RNR with different parameters on the Extended Yale B database with noise block occlusion. (a) Regularization parameter and (b) balance factor.

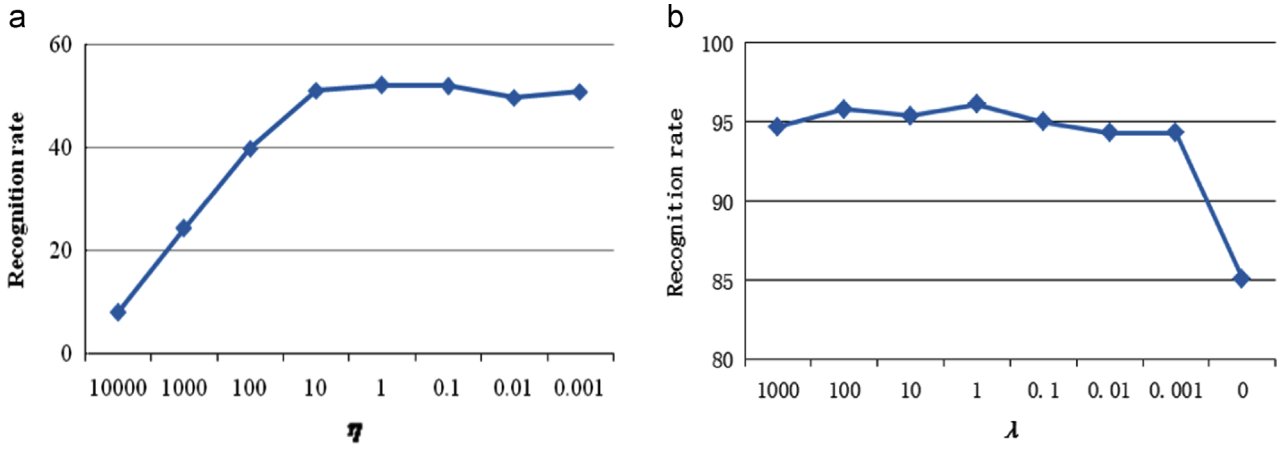

Fig. 18. The recognition rates (%) of RNR with different parameters on the Extended Yale B database with mixture noise (pixel corruption and block occlusion). (a) Regularization parameter and (b) balance factor.

outperform the other methods. When the occlusion percentage is 50%, the recognition rate of RNR is 10.4%, 11.6%, 36.9% and 29% higher than RSC, SSEC, CESR and SRC, respectively.

The setting of the second experiment is similar to that of the first one. The only difference is that subset 3 with noise block images is used for testing. λ is 0.1 and the regularization

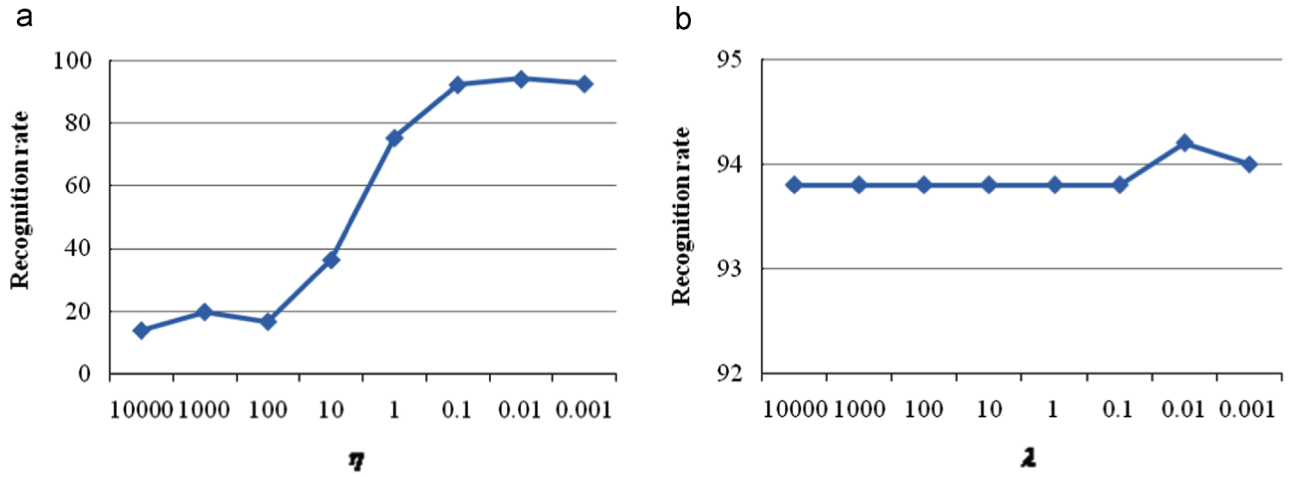

Fig. 19. The recognition rates (%) of RNR with different parameters on the FRGC database without occlusion. (a) Regularization parameter and (b) balance factor.

parameter η is set to 10. The recognition rates of each method versus the various levels of occlusion (from 10% to 50%) are shown in Fig. 11. From Fig. 11, we observe that the proposed RNR 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 RNR 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. 12. Although the performance of each method degrades with the increment of the mixture noise level, RNR 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 continuous occlusion problem.

5.3. Experiments on the FRGC database (without occlusion)

Although our motivation is to design robust methods for face recognition with occlusion, the proposed method can be used as a general face recognition algorithm. In this section, we evaluate the performance of the proposed method on the FRGC database.

The FRGC version 2.0 is a large scale face image database, including controlled and uncontrolled images [30]. This database contains 12,776 training images (6360 controlled images and 6416 uncontrolled ones) from 222 individuals, 16,028 controlled target images and 8014 uncontrolled query images from 466 persons. We use a subset (220 persons, each person having 20 images) of FRGC. The face region of each image is first cropped from the original high-resolution images and resized to a resolution of 16×16 and 32×32 pixels, respectively. Fig. 13 shows some images used in our experiments.

In our experiments, the first 10 images per class are used for training, and the remaining images are used for testing. So there are totally 2200 training images and 2200 testing images, respectively. Both the balance factor λ and the regularization parameter η of RNR are set to 10^{-2} here. Table 3 shows the experimental results of CRC, SRC, CESR, SSEC, RSC, NR and RNR. From Table 3, we can see that the proposed RNR achieves the best results in both image sizes for face recognition. RNR gives 2.1%, 1.9% and 4.9% improvement over RSC, SRC and CRC, respectively, when the image size is 32×32. SSEC was designed exclusively for contiguous occlusion, but its performance is not good for face recognition without occlusion.

Table 4
The recognition rates (%) of RNR using different weight functions on AR dataset.

	Sunglasses	Scarves	Session 1		Session 2	
			Sunglasses	Scarves	Sunglasses	Scarves
Welsch	99.0	98.5	94.7	95.3	83.0	75.3
Cauchy	99.0	99.0	95.3	94.7	81.7	74.3
logistic	99.0	100	97.7	95.0	82.3	77.3

5.4. Effects of parameters

In this section, we mainly introduce how the parameters (balance factor$\lambda$and regularization parameter η) affect the performance of our method RNR. We perform experiments on three public face image databases (AR, Extended Yale B and FRGC) and the experimental setting is the same as the above experiments. In our experiments, we just change one parameter while fixing the other one.

For face recognition with real disguise, the recognition rates of RNR on the AR database with sunglasses (or scarf) are shown in Fig. 14 (or Fig. 15). From Figs. 14 and 15, we observe that the performance of RNR degrades with decreasing the regularization parameter η. In addition, RNR achieves the best results when the λ is 100 for the test images with sunglasses and 0.1 for the test images with scarf.

For face recognition with block occlusion, we plot the recognition rates of RNR versus different parameters on the Extended Yale B database with unrelated image (or noise block image) occlusion as shown in Fig. 16 (or Fig. 17). From Fig. 16, we can see that RNR gives the best results when both η and λ are 10. However, RNR achieves the best performance when η is 0.01 and λ is 1 for the test images with noise block occlusion. In addition, Fig. 18 shows the recognition rates of RNR versus different parameters on the Extended Yale B database with mixture noise (pixel corruption and block occlusion).

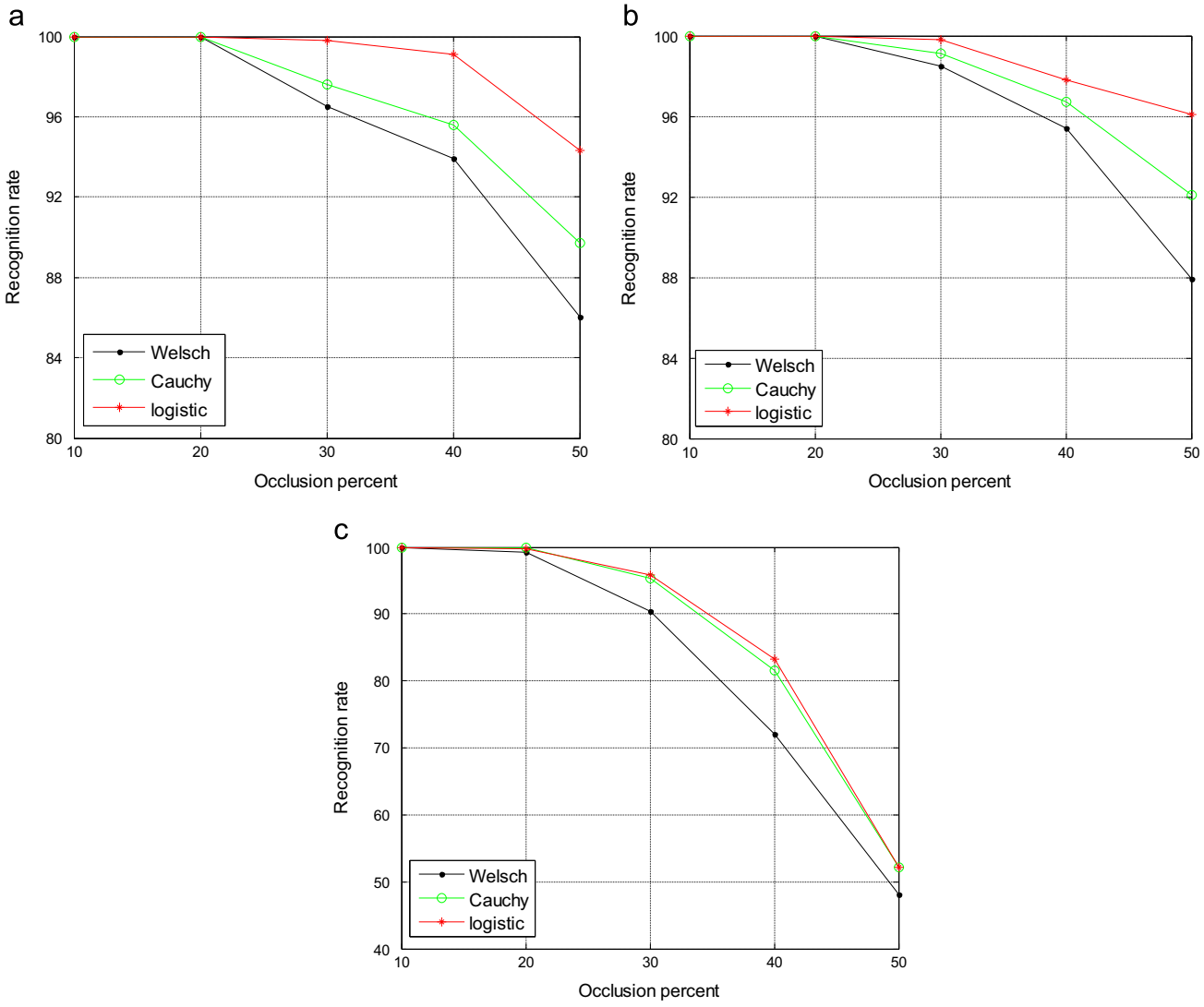

Fig. 20. The recognition rates (%) of RNR using different weight functions on Extended Yale B dataset. (a) Test image with unrelated image occlusion; (b) test image with noise image occlusion; and (c) test image with mixture noises.

For face recognition without occlusion, Fig. 19 shows recognition rates of RNR on the FRGC database with different balance factor λ and regularization parameter η, respectively. From Fig. 19, we can see that RNR achieves better results when η is lower than 1 and worse results when η is larger than 10. However, RNR is not sensitive to the balance factor λ.

Finally, we also show the performance of the proposed model with different weight functions (logistic, Welsch and Cauchy) to handle face recognition with occlusion. The experiments setting are same with Sections 5.1 and 5.2. Table 4 lists the results of RNR using different weight function on AR dataset. Fig. 20 shows the recognition rates of RNR using different weight function on Extended Yale B dataset. From Table 4 and Fig. 20, we can see that the proposed model using logistic function performs better than Welsch and Cauchy functions in most cases. So we choose logistic function as weight function in our model.

6. Conclusions

In this paper, we present a novel nuclear norm regularized regression model and apply the alternating direction method of multipliers to solve it. The robust nuclear norm regularized regression based classification (RNR) method is introduced for face recognition. RNR 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 RNR is robust to corruptions: real disguise and random block occlusion, and yields better performances as compared to state-of-the-art methods.

Conflict of interest

None declare.

Acknowledgment

This work was partially supported by the National Science Fund for Distinguished Young Scholars under Grant nos. 61125305, 61472187, 61233011 and 61373063, the Key Project of Chinese Ministry of Education under Grant no. 313030, the 973 Program No. 2014CB349303, Fundamental Research Funds for the Central Universities No. 30920140121005, and Program for Changjiang Scholars and Innovative Research Team in University No.

IRT13072, the Open Research Fund of Jiangsu Key Laboratory of Spectral Imaging & Intelligent Sense. Zhouchen Lin is supported by 973 Program of China (Grant no. 2015CB352502), NSF China (Grant nos. 61272341 and 61231002), and MSRA.

References

[1] W. Zhao, R. Chellappa, P.J. Phillips, et al., Face recognition: a literature survey, ACM Comput. Surv. 35 (4) (2003) 399–459.
[2] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. PAMI 31 (2) (2009) 210–227.
[3] J. Yang, L. Zhang, Y. Xu, J.Y. Yang, Beyond sparsity: the role of L1-optimizer in pattern classification, Pattern Recognit. 45 (2012) 1104–1118.
[4] M. Yang, L. Zhang, Gabor feature based sparse representation for face recognition with Gabor occlusion dictionary, in: ECCV, 2010.
[5] M. Yang, L. Zhang, Simon C.K. Shiu, David Zhang, Gabor feature based robust representation and classification for face recognition with Gabor occlusion dictionary, Pattern Recognit. 46 (2013) 1865–1878.
[6] Z. Zhou, A. Wagner, H. Mobahi, J. Wright, Y. Ma, Face recognition with contiguous occlusion using Markov random fields, in: ICCV, 2009.
[7] M. Yang, L. Zhang, J. Yang, D. Zhang, Robust sparse coding for face recognition, in: CVPR, 2011.
[8] R. He, W.S. Zheng, B.G. Hu, X.W. Kong, A regularized correntropy framework for robust pattern recognition, Neural Comput. 23 (2011) 2074–2100.
[9] R. He, W.S. Zheng, B.G. Hu, Maximum correntropy criterion for robust face recognition, IEEE PAMI 33 (8) (2011) 1561–1576.
[10] R. He, W.S. Zheng, T. Tan, Z. Sun, Half-quadratic based iterative minimization for robust sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI) 36 (2) (2014) 261–275.
[11] R. Rigamonti, M. Brown, V. Lepetit. Are sparse representations really relevant for image classification, in: CVPR, 2011.
[12] Q. Shi, A. Eriksson, A. Hengel, C. Shen, Is face recognition really a compressive sensing problem, in: CVPR, 2011.
[13] I. Naseem, R. Togneri, M. Bennamoun, Robust regression for face recognition, Pattern Recognit. 45 (2012) 104–118.
[14] L. Zhang, M. Yang, X.C. Feng, Sparse representation or collaborative representation which helps face recognition, in: ICCV, 2011.
[15] X.-X. Li, D.-Q. Dai, X.-F. Zhang, C.-X. Ren, Structured sparse error coding for face recognition with occlusion, IEEE Trans. Image Process. 22 (5) (2013) 1889–1999.
[16] I. Naseem, R. Togneri, M. Bennamoun, Linear regression for face recognition, IEEE PAMI 32 (11) (2010) 2106–2112.
[17] M. Fazel, Matrix Rank Minimization with Applications (Ph.D. thesis), Stanford University, 2002.
[18] M. Fazel, H. Hindi, S. Boyd, A rank minimization heuristic with application to minimum order system approximation, Proc. Am. Control Conf. 6 (2001) 4734–4739.
[19] E. Candès, X.D. Li, Y. Ma, J. Wright, Robust principal component analysis, J. ACM 58 (3) (2011) (Article 11).
[20] J. Wright, A. Ganesh, S. Rao, Y. Peng, Y. Ma, Robust principal component analysis: exact recovery of corrupted low-rank matrices via convex optimization, in: Proceedings of Neural Information Processing Systems (NIPS), December 2009.
[21] D. Zhang, Y. Hu, J.P. Ye, X.L. Li, X.F. He. Matrix completion by truncated nuclear norm regularization, in: CVPR, 2012.
[22] Long Ma, Chunheng Wang, Baihua Xiao, Wen Zhou, Sparse representation for face recognition based on discriminative low-rank dictionary learning, in: CVPR, 2012.
[23] Chih-Fan Chen, Chia-Po Wei and Yu-Chiang Frank Wang. Low-rank matrix recovery with structural incoherence for robust face recognition, in: CVPR, 2012.
[24] T. Poggio, S. Smale., The mathematics of learning: dealing with data, Not. AMS 50 (5) (2003) 537–544.
[27] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Found. Trends Mach. Learn. 3 (1) (2011) 1122.
[28] J.F. Cai, E.J. Candès, Z. Shen, A singular value thresholding algorithm for matrix completion, SIAM J. Optim., 2010.
[29] A. Martinez, R. Benavente, The AR face database. Technical Report 24, CVC, 1998.
[30] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the face recognition grand challenge, in: CVPR, 2005.
[31] K.C. Lee, J. Ho, D. Driegman, Acquiring linear subspaces for face recognition under variable lighting, IEEE Trans. PAMI 27 (5) (2005) 684–698.
[32] J. Qian, J. Yang, F. Zhang, Z. Lin, Robust low-rank regularized regression for face recognition with occlusion, in: Proceedings of the Biometrics Workshop in conjunction with IEEE Conference on Computer Vision and Pattern Recognition (CVPRW), Columbus, Ohio, June 23, 2014.
[34] Z. Lin, R. Liu, Z. Su, Linearized alternating direction method with adaptive penalty for low rank representation, in: NIPS, 2011.
[35] B.S. He, M.H. Xu, X.M. Yuan, Solving large-scale least squares covariance matrix problems by alternating direction methods, SIAM J. Matrix Anal. Appl. 32 (2011) 136–152.
[36] B.S. He, H. Yang, Some convergence properties of a method of multipliers for linearly constrained monotone variational inequalities, Oper. Res. Lett. 23 (1998) 151–161.
[37] Y. Zhang, Z. Jiang, Larry S. Davis, Learning structured low-rank representations for image classification, in: CVPR, 2013.
[38] R. He, Z. Sun, T. Tan, W.S. Zheng, Recovery of corrupted low-rank matrices via half-quadratic based nonconvex minimization, in: CVPR, 2011.
[39] J. Hiriart-Urruty, C. Lemarechal, Convex Analysis and Minimization Algorithms, Springer-Verlag, New York, 1996.
[40] K. Jia, T. Chan, Y. Ma, Robust and practical face recognition via structured sparsity, in: ECCV, 2012.
[41] J. Yang, C. Liu, Horizontal and vertical 2DPCA-based discriminant analysis for face verification on a large-scale database, IEEE Trans. Inf. Forensics Secur. 2 (4) (2007) 781–792.

Jianjun Qian received the B.S. and M.S. degrees in 2007 and 2010, respectively, and the Ph. D. degree in pattern recognition and intelligence systems from Nanjing University of Science and Technology (NUST), in 2014. Now, he is an assistant professor in the School of Computer Science and Engineering of NUST. His research interests include pattern recognition, computer vision and face recognition in particular.

Lei Luo received the B.S. degree from Xinyang Normal University, Xinyang, China in 2008, the M.S. degree from Nanchang University, Nanchang, China in 2011. He is currently pursuing the Ph.D. degree in pattern recognition and intelligence system from School of Computer Science and engineering, Nanjing University of Science and Technology, Nanjing, China. His current research interests include pattern recognition and optimization algorithm.

Jian Yang received the B.S. degree in mathematics from the Xuzhou Normal University in 1995. He received the MS degree in applied mathematics from the Changsha Railway University in 1998 and the Ph.D. degree from the Nanjing University of Science and Technology(NUST), on the subject of pattern recognition and intelligence systems in 2002. In 2003, he was a postdoctoral researcher at the University of Zaragoza. From 2004 to 2006, he was a Postdoctoral Fellow at Biometrics Centre of HongKong Polytechnic University. From 2006 to 2007, he was a Postdoctoral Fellow at Department of Computer Science of New Jersey Institute of Technology. Now, he is a professor in the School of Computer Science and Technology of NUST. He is the author of more than 80 scientific papers in pattern recognition and computer vision. His journal papers have been cited more than 3000 times in the ISI Web of Science, and 7000 times in the Web of Scholar Google. His research interests include pattern recognition, computer vision and machine learning. Currently, he is an associate editor of Pattern Recognition Letters and IEEE Transactions on Neural Networks and Learning Systems, respectively.

Fanlong Zhang received the B.S. and M.S. degrees in 2007 and 2010, respectively. Currently, he is pursuing the Ph.D. degree with the School of Computer Science and Engineering, Nanjing University of Science and Technology (NUST), Nanjing, China. His current research interests include pattern recognition and optimization.

Zhouchen Lin received the Ph.D. degree in applied mathematics from Peking University, Beijing, China, in 2000. He is currently a Professor with the Key Laboratory of Machine Perception, school of Electronics Engineering and Computer Science, Peking University. He is a Chair Professor with Northeast Normal University, Changchun, China. In 2012, he was a Lead Researcher with the Visual Computing Group, Microsoft Research Asia. He was a Guest Professor with Shanghai Jiaotong University, Shanghai, China, Beijing Jiao Tong University, Beijing, and Southeast University, Nanjing, China. He was a Guest Researcher with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing. His current research interests include computer vision, image processing, computer graphics, machine learning, pattern recognition, and numerical computation and optimization. Dr. Lin is an Associate Editor of the International Journal of Computer Vision.