Signal Processing 222 (2024) 109523

Contents lists available at ScienceDirect

Signal Processing

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

Short communication

Robust sparse representation based on fitting error decomposition

Xiang-Yu Wang [a], Xiao-Peng Li [b,*], Hing Cheung So [a,1]

[a] Department of Electrical Engineering, City University of Hong Kong, Hong Kong Special Administrative Region, China
[b] College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060, China

ARTICLE INFO

Keywords:
Robust algorithm
Sparse representation
Outlier
ℓ_0-norm
Logarithmic norm

ABSTRACT

Sparse representation (SR) of a signal aims at finding the minimum number of atoms for its representation. In several practical scenarios, the signal is vulnerable to outliers and thus robustness is required for SR based algorithms. However, most existing robust schemes are designed on the assumption that the anomalies are independently distributed, which may not perform well encountering complex interferences, especially the correlated ones. To deal with this drawback, a robust SR model is proposed, where both independent and correlated outliers are considered. Specifically, the fitting error is decomposed as the combination of a low-rank component and a sparse part, corresponding to the correlated gross error and independent outlier, respectively. Then, the group sparsity of the representation coefficient is utilized. Moreover, ℓ_0-norm and $\ell_{2,0}$-norm are adopted as the sparsity regularization for the sparse outlier and representation coefficients, respectively. The solutions to $\ell_0/\ell_{2,0}$-norm minimization are generated by the hard-thresholding strategy, where the decision threshold is adaptively determined using median absolute deviation operator. As for the low-rank regularization, due to the NP-hardness of rank minimization, we employ matrix logarithmic norm as the rank surrogate to lessen the approximation gap. Finally, we apply the proposed model to face recognition task, and the excellent performance demonstrates its effectiveness.

1. Introduction

Sparse representation (SR) has found versatile applications in signal processing [1,2], computer vision [3], and pattern recognition [4]. SR is closely related to compressive sensing which reconstructs a wide classes of signals, like audio and images, using much less measured values than those required by Shannon sampling theorem [5]. With the aid of a basis or dictionary, e.g., Fourier or wavelet basis, the compressible signals have a linear representation of the atoms in the dictionary, where the combination coefficient is sparse. Generally, SR is formulated as [6]

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{s.t.} \|\mathbf{y}-\mathbf{D}\mathbf{x}\|_2^2 \le \bar{r}, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^m$ is the observed signal, and $\mathbf{D} \in \mathbb{R}^{m\times N}$ with $m < N$ is an over-complete dictionary whose columns are atoms. Herein, $\bar{r}$ constrains the fitting error, which is measured by ℓ_2-norm. As for $\mathbf{D}$, it can be generated from a predefined transform, like discrete Fourier transform, which is easy and suitable for generic signals. Besides, we can also learn $\mathbf{D}$ from training data [7–10]. Although the latter involves high computational complexity, it offers improved performance.

Unfortunately, minimizing ℓ_0-norm is NP-hard [11]. Thereby, ℓ_1-norm is widely adopted as a convex surrogate of ℓ_0-norm for SR [12], and ℓ_1-norm optimization problem has an analytical solution. Notably, the approximation gap between ℓ_0-norm and ℓ_1-norm cannot be not negligible, which may damage the solution sparsity. Therefore, to mitigate this gap, ℓ_p-norm ($0 < p < 1$) is developed as the sparsity regularization [13] and performs well. However, how to choose p is an open problem. Besides, ℓ_p-norm optimization is not computationally efficient.

Without loss of generality, the signal is more likely to be represented by a group of closely related atoms, and the components of the solution are naturally grouped [14]. For the ideal solution to SR, its coefficients corresponding to each group are either all nonzero or zero. Employing this property, $\ell_{p,q}$-norm ($p > 1$, $0 \le q \le 1$) regularization for $\mathbf{x}$ is studied for group SR [14–16]. In addition,the work [17] proposes adaptive class preserving representation for classification (ACPRC), which balances lasso regression and group lasso regression.

In several practical situations, $\mathbf{y}$ suffers from various noises and interferences. For examples, electrocardiogram signals are typically contaminated by noise [18] during recording and transmission, while face images are subject to illuminations and occlusions [19,20]. Thus, robustness is a major concern. Utilizing ℓ_2-norm for data fidelity as in (1) only works well for Gaussian noise, which is not appropriate for

* Corresponding author.
E-mail addresses: xwang2286-c@my.cityu.edu.hk (X.-Y. Wang), x.p.li@szu.edu.cn (X.-P. Li), hcso@ee.cityu.edu.hk (H.C. So).
[1] EURASIP Member.

https://doi.org/10.1016/j.sigpro.2024.109523
Received 16 January 2024; Received in revised form 19 April 2024; Accepted 29 April 2024
Available online 9 May 2024

outliers. To handle this issue, various robust estimators are developed, like Huber estimator [21], correntropy induced estimator [22], and Lorentzian norm [23].

The above-mentioned robust algorithms deal with the anomalies in an element-wise manner, which is essentially based on the assumption that outliers are independently distributed. However, for certain applications, the interferences can be very complicated, like disguises in image classification. Such kind of interferences is locally continuously changed and bears structural information [24]. Therefore, low-rankness is suitable to characterize the correlated anomaly. Due to the NP-hardness of rank minimization [25], various rank surrogates are exploited for robust SR [15,19,20].

In this communication, a new robust SR model is proposed. We assume the atoms in the dictionary are grouped. Then, group sparsity of $\mathbf{x}$ is adopted, which is optimized by $\ell_{2,0}$-norm minimization. To deal with the complicated interferences, both independent and correlated outliers are considered. Specifically, the fitting error is correspondingly decomposed into a sparse part and a low-rank one, where the sparse outlier is regularized by ℓ_0-norm. For $\ell_{2,0}$-norm and ℓ_0-norm optimization, the hard-thresholding operator is adopted as the solver, where the decision threshold is determined by median absolute deviation (MAD) [26]. As for the low-rank part, owing to the intractable rank function, we employ the surrogate logarithmic norm as its substitute, which is then converted as a regularization. The logarithmic norm outperforms other rank substitutions in low-rank matrix completion task [27] and is first applied for SR to the best of our knowledge. The proposed robust SR model is then optimized using alternating direction method of multipliers (ADMM) [28]. Its effectiveness is verified with application to face recognition (FR).

2. Robust SR model

2.1. Problem formulation

Scalars, vectors, and matrices are represented by lowercase letters, bold lowercase letters, and bold uppercase letters, respectively. For a vector $\mathbf{v}$, its i-th element is denoted as v_i.

Generally, the sample $\mathbf{y}$ may be contaminated by outliers. Based on (1), robust SR is formulated as [19]

$$\min_{\mathbf{x}} \ \lambda \|\mathbf{x}\|_0 + \phi(\mathbf{e}) \quad \text{s.t. } \mathbf{y} - \mathbf{D}\mathbf{x} = \mathbf{e}, \tag{2}$$

where $\mathbf{e}$ denotes the interference or fitting error, $\phi(\cdot)$ is the regularization operator, and λ is a penalty parameter.

Since real-world interferences may be rather complicated, like illuminations and occlusions, single error term is insufficient to handle such situations. For example, Huber estimator [21] calculates the fitting error in an element-wise manner, which actually assumes that the errors are independently distributed. However, illuminations and occlusions are continuous changed in local areas, and errors are correlated. This single error term may not depict such interferences precisely.

In contrast, we consider both the independent and correlated outliers, and decompose the anomalies into a sparse part and a low-rank one, viz.

$$\mathbf{y} - \mathbf{D}\mathbf{x} = \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2, \tag{3}$$

where $\boldsymbol{\eta}_1$ denotes the vectorized low-rank part, and $\boldsymbol{\eta}_2$ is the sparse component. By doing so, our model can better deal with complicated anomalies than those using only a single error term. Then, (2) is further written as

$$\min_{\mathbf{x},\boldsymbol{\eta}_1,\boldsymbol{\eta}_2} \ \operatorname{rank}(\mathcal{M}(\boldsymbol{\eta}_1)) + \lambda_1 \|\boldsymbol{\eta}_2\|_0 + \lambda_2 \|\mathbf{x}\|_0 \quad \text{s.t. } \mathbf{y} - \mathbf{D}\mathbf{x} = \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2, \tag{4}$$

where $\mathcal{M}(\cdot)$ denotes the operator converting a vector to a matrix, and $\lambda_1 > 0$, $\lambda_2 > 0$ are penalty parameters. Here, $\mathcal{M}(\boldsymbol{\eta}_1) \in \mathbb{R}^{m_1 \times m_2}$ with $m = m_1 m_2$. As an illustration, for an image, $\mathbf{y}$ is generated from its vectorization with dimensions $m_1 \times m_2$.

Since rank-minimization is NP-hard, different rank surrogates are proposed. As a convex envelop of rank function, the nuclear norm is widely used. However, the nuclear norm regularization over-penalizes large singular values [27], resulting in loose approximation to the rank. Hence, non-convex rank substitutes are developed. Recently, logarithmic norm has shown better performance than the nuclear norm and other non-convex surrogates on matrix/tensor completion task [27,29]. Therefore, in this work, we adopt the logarithmic norm as the rank substitute. Given a matrix $\mathbf{M} \in \mathbb{R}^{m_1 \times m_2}$, the logarithmic norm is defined as [27]

$$\|\mathbf{M}\|_{\text{Log}} = \sum_{j=1}^{\min(m_1,m_2)} \log\left(\sigma_j(\mathbf{M}) + \varepsilon\right), \tag{5}$$

where $\sigma_j(\mathbf{M})$ is the j-th singular value of $\mathbf{M}$. Hyper-parameter ε is related to the approximation accuracy to the rank and can be determined according to the strategy in [29]. Substituting rank function in (4) by logarithmic norm, we get

$$\min_{\mathbf{x},\boldsymbol{\eta}_1,\boldsymbol{\eta}_2} \ \left\|\mathcal{M}(\boldsymbol{\eta}_1)\right\|_{\text{Log}} + \lambda_1 \|\boldsymbol{\eta}_2\|_0 + \lambda_2 \|\mathbf{x}\|_0 \quad \text{s.t. } \mathbf{y} - \mathbf{D}\mathbf{x} = \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2. \tag{6}$$

Next, we adopt group sparsity for (6). We assume there are c groups of the atoms. Then, $\mathbf{y}$ tends to be represented by a closely related group of atoms. Thus, the non-zero elements of solution $\mathbf{x}$ are naturally grouped. In this case, we reorganize $\mathbf{D}$ as $\mathbf{D} = [\mathbf{D}_1, \ldots, \mathbf{D}_i, \ldots, \mathbf{D}_c]$, where columns of $\mathbf{D}_i \in \mathbb{R}^{m \times N_i}$ are the i-th group of atoms and $N = \sum_{i=1}^{c} N_i$. Correspondingly, $\mathbf{x}$ is divided as $\mathbf{x} = [\mathbf{x}_1^T, \ldots, \mathbf{x}_i^T, \ldots, \mathbf{x}_c^T]^T$ with $\mathbf{x}_i \in \mathbb{R}^{N_i}$. The group sparsity of $\mathbf{x}$ is measured by $\ell_{2,0}$-norm [14,15], which is $\|\mathbf{x}\|_{2,0} = \|\boldsymbol{\theta}\|_0, \boldsymbol{\theta} = [\|\mathbf{x}_1\|_2, \ldots, \|\mathbf{x}_c\|_2]^T$. Then, applying the group sparsity to (6) yields

$$\min_{\mathbf{x},\boldsymbol{\eta}_1,\boldsymbol{\eta}_2} \ \left\|\mathcal{M}(\boldsymbol{\eta}_1)\right\|_{\text{Log}} + \lambda_1 \|\boldsymbol{\eta}_2\|_0 + \lambda_2 \|\mathbf{x}\|_{2,0} \quad \text{s.t. } \mathbf{y} - \mathbf{D}\mathbf{x} = \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2. \tag{7}$$

To further enhance the group sparsity of $\mathbf{x}$, a weight vector $\mathbf{w} \in \mathbb{R}^c$ is introduced in $\ell_{2,0}$-norm regularization [15]. In so doing, the optimization problem becomes:

$$\begin{aligned} &\min_{\mathbf{x},\boldsymbol{\theta},\boldsymbol{\eta}_1,\boldsymbol{\eta}_2} \ \left\|\mathcal{M}(\boldsymbol{\eta}_1)\right\|_{\text{Log}} + \lambda_1 \|\boldsymbol{\eta}_2\|_0 + \lambda_2 \|\mathbf{w} \odot \boldsymbol{\theta}\|_0 \\ &\text{s.t. } \mathbf{y} - \mathbf{D}\mathbf{x} = \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2, \ \boldsymbol{\theta} = [\|\mathbf{x}_1\|_2, \ldots, \|\mathbf{x}_c\|_2]^T, \end{aligned} \tag{8}$$

where hyper-parameter $\mathbf{w}$ penalizes $\boldsymbol{\theta}$ in an element-wise manner. As for how to choose $\mathbf{w}$, intuitively, if the ground-truth $\theta_i / \|\mathbf{x}_i\|_2$ is small, viz. $\mathbf{y}$ does not belong to group i, w_i should be set a large value. That is to say, w_i is positively correlated to the distance between $\mathbf{y}$ and the column space of $\mathbf{D}_i$, where the distance is represented as $\|\mathbf{y} - \mathbf{D}_i \mathbf{x}_i\|_2$ with $\mathbf{x}_i$ being unknown. To obtain a coarsely estimated $\mathbf{x}_i$, we solve the following least squares problem

$$\bar{\mathbf{x}}_i = \arg\min_{\mathbf{x}_i} \|\mathbf{y} - \mathbf{D}\mathbf{x}_i\|_2^2, \tag{9}$$

whose solution is $\bar{\mathbf{x}}_i = (\mathbf{D}_i^T \mathbf{D}_i)^{-1} \mathbf{D}_i^T \mathbf{y}$. With $\bar{\mathbf{x}}_i$, we calculate the distance (or representative residual) as $r_i = \|\mathbf{y} - \mathbf{D}_i \bar{\mathbf{x}}_i\|_2$. Then, we adopt the min–max normalized r_i as w_i, viz.

$$w_i = \frac{r_i - r_{\min}}{r_{\max} - r_{\min}}, \ r_{\max} = \max(r_1, \ldots, r_c), \ r_{\min} = \min(r_1, \ldots, r_c). \tag{10}$$

2.2. Optimization

We solve (8) via ADMM that is widely used to handle constrained optimization problems and is able to converge in just a few tens of iterations [28]. Introducing auxiliary variables $\boldsymbol{\upsilon} \in \mathbb{R}^c$ and $\mathbf{g} \in \mathbb{R}^N$, (8) is reformulated as

$$\begin{aligned} &\min_{\mathbf{x},\boldsymbol{\theta},\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\boldsymbol{\upsilon},\mathbf{g}} \left\|\mathcal{M}(\boldsymbol{\eta}_1)\right\|_{\text{Log}} + \lambda_1 \|\boldsymbol{\eta}_2\|_0 + \lambda_2 \|\boldsymbol{\upsilon}\|_0 \\ &\text{s.t. } \mathbf{y} - \mathbf{D}\mathbf{x} = \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2, \boldsymbol{\theta} = \tilde{\mathbf{g}} = [\|\mathbf{g}_1\|_2, \ldots, \|\mathbf{g}_c\|_2]^T, \boldsymbol{\upsilon} = \mathbf{w} \odot \boldsymbol{\theta}, \mathbf{x} = \mathbf{g}. \end{aligned} \tag{11}$$

Then, the augmented Lagrangian of (11) is

$$\begin{aligned}\mathcal{L}_\beta = &\left\|\mathcal{M}\left(\boldsymbol{\eta}_1\right)\right\|_{\mathrm{Log}} + \lambda_1\left\|\boldsymbol{\eta}_2\right\|_0 + \lambda_2\left\|\boldsymbol{\upsilon}\right\|_0 + \boldsymbol{\alpha}_1^T\left(\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 - \mathbf{y} + \mathbf{D}\mathbf{x}\right)\\ &+ \boldsymbol{\alpha}_2^T\left(\boldsymbol{\upsilon} - \mathbf{w}\odot\boldsymbol{\theta}\right) + \boldsymbol{\alpha}_3^T\left(\mathbf{x} - \mathbf{g}\right)\\ &+ \boldsymbol{\alpha}_4^T\left(\tilde{\mathbf{g}} - \boldsymbol{\theta}\right) + \frac{\beta}{2}\left(\left\|\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 - \mathbf{y} + \mathbf{D}\mathbf{x}\right\|_2^2\right.\\ &\left.+ \left\|\boldsymbol{\upsilon} - \mathbf{w}\odot\boldsymbol{\theta}\right\|_2^2 + \left\|\mathbf{x} - \mathbf{g}\right\|_2^2 + \left\|\tilde{\mathbf{g}} - \boldsymbol{\theta}\right\|_2^2\right),\end{aligned} \tag{12}$$

where $\boldsymbol{\alpha}_1 \in \mathbb{R}^m$, $\boldsymbol{\alpha}_2 \in \mathbb{R}^c$, $\boldsymbol{\alpha}_3 \in \mathbb{R}^m$, and $\boldsymbol{\alpha}_4 \in \mathbb{R}^c$ are the Lagrange multipliers, while $\beta > 0$ is a penalty parameter. In the k-th iteration, the minimization of $\mathcal{L}_\beta$ is decomposed into six subproblems.

The $\boldsymbol{\eta}_1$-subproblem is

$$\begin{aligned}\boldsymbol{\eta}_1^k = &\arg\min_{\boldsymbol{\eta}_1}\ \left\|\mathcal{M}\left(\boldsymbol{\eta}_1\right)\right\|_{\mathrm{Log}} + \boldsymbol{\alpha}_1^{k-1^T}\left(\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2^{k-1} - \mathbf{y} + \mathbf{D}\mathbf{x}^{k-1}\right)\\ &+ \frac{\beta^{k-1}}{2}\left\|\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2^{k-1} - \mathbf{y} + \mathbf{D}\mathbf{x}^{k-1}\right\|_2^2\\ \Rightarrow \boldsymbol{\eta}_1^k = &\arg\min_{\boldsymbol{\eta}_1}\ \left\|\mathcal{M}\left(\boldsymbol{\eta}_1\right)\right\|_{\mathrm{Log}} + \frac{\beta^{k-1}}{2}\left\|\mathcal{M}\left(\boldsymbol{\eta}_1\right) - \mathcal{M}\left(\hat{\boldsymbol{\eta}}_1^{k-1}\right)\right\|_F^2,\end{aligned} \tag{13}$$

where $\hat{\boldsymbol{\eta}}_1^{k-1} = \mathbf{y} - \mathbf{D}\mathbf{x}^{k-1} - \boldsymbol{\eta}_2^{k-1} - \frac{\boldsymbol{\alpha}_1^{k-1}}{\beta^{k-1}}$.

Setting the SVD of $\mathcal{M}\left(\hat{\boldsymbol{\eta}}_1^{k-1}\right)$ as $\mathbf{Q}_{\mathcal{M}\left(\hat{\boldsymbol{\eta}}_1^{k-1}\right)}\boldsymbol{\Lambda}_{\mathcal{M}\left(\hat{\boldsymbol{\eta}}_1^{k-1}\right)}\mathbf{R}^T_{\mathcal{M}\left(\hat{\boldsymbol{\eta}}_1^{k-1}\right)}$, the solution to (13) is [27]

$$\boldsymbol{\eta}_1^k = \mathcal{V}\left(\mathbf{Q}_{\mathcal{M}\left(\hat{\boldsymbol{\eta}}_1^{k-1}\right)}\mathcal{T}_{1/\beta^{k-1},\varepsilon}\left(\boldsymbol{\Lambda}_{\mathcal{M}\left(\hat{\boldsymbol{\eta}}_1^{k-1}\right)}\right)\mathbf{R}^T_{\mathcal{M}\left(\hat{\boldsymbol{\eta}}_1^{k-1}\right)}\right), \tag{14}$$

where $\mathcal{V}(\cdot)$ implements the inverse operation of $\mathcal{M}(\cdot)$ and the element-wise operator $\mathcal{T}_{\tau,\varepsilon}(\cdot)$ is defined as

$$\mathcal{T}_{\tau,\varepsilon}(\sigma) = \begin{cases} 0, & \Delta \le 0,\\ \arg\min\limits_{c\in\left\{0,\frac{1}{2}\left(\sigma-\varepsilon+\sqrt{\Delta}\right)\right\}} f(c), & \Delta > 0,\end{cases} \tag{15}$$

where $\Delta = (\sigma-\varepsilon)^2 - 4(\tau-\sigma\varepsilon)$ and $f(c) = \frac{1}{2}(c-\sigma)^2 + \tau log(c+\varepsilon)$.

The subproblem of $\boldsymbol{\eta}_2$ is

$$\begin{aligned}\boldsymbol{\eta}_2^k = &\arg\min_{\boldsymbol{\eta}_2}\ \lambda_1^k\left\|\boldsymbol{\eta}_2\right\|_0 + \boldsymbol{\alpha}_1^{k-1^T}\left(\boldsymbol{\eta}_1^k + \boldsymbol{\eta}_2 - \mathbf{y} + \mathbf{D}\mathbf{x}^{k-1}\right)\\ &+ \frac{\beta^{k-1}}{2}\left\|\boldsymbol{\eta}_1^k + \boldsymbol{\eta}_2 - \mathbf{y} + \mathbf{D}\mathbf{x}^{k-1}\right\|_2^2\\ \Rightarrow \boldsymbol{\eta}_2^k = &\arg\min_{\boldsymbol{\eta}_2}\ \left(\mu_{\boldsymbol{\eta}_2}^k\right)^2\left\|\boldsymbol{\eta}_2\right\|_0 + \left\|\boldsymbol{\eta}_2 - \hat{\boldsymbol{\eta}}_2^{k-1}\right\|_2^2,\end{aligned} \tag{16}$$

where $\hat{\boldsymbol{\eta}}_2^{k-1} = \mathbf{y} - \mathbf{D}\mathbf{x}^{k-1} - \boldsymbol{\eta}_1^k - \frac{\boldsymbol{\alpha}_1^{k-1}}{\beta^{k-1}}$. To better control the sparsity of $\boldsymbol{\eta}_2$, λ_1 is set to be adaptively adjusted during iterations. After simplification, $\left(\mu_{\boldsymbol{\eta}_2}^k\right)^2 = \frac{2\lambda_1^k}{\beta^{k-1}}$ is introduced for sparsity control. The solution to (16) is computed as [30]

$$\eta_{2,i}^k = \psi_{\mu_{\boldsymbol{\eta}_2}^k}\left(\hat{\eta}_{2,i}^{k-1}\right) = \begin{cases}\hat{\eta}_{2,i}^{k-1}, & \left|\hat{\eta}_{2,i}^{k-1}\right| \ge \mu_{\boldsymbol{\eta}_2}^k,\\ 0, & \text{otherwise.}\end{cases} \tag{17}$$

Here, $\psi_{\mu_{\boldsymbol{\eta}_2}^k}(\cdot)$ is a hard-thresholding operator which sets $\hat{\eta}_{2,i}^{k-1}$ whose magnitude smaller than $\mu_{\boldsymbol{\eta}_2}^k$ to zero. To adaptively determine $\mu_{\boldsymbol{\eta}_2}^k$, we suggest a strategy based on normalized MAD:

$$\begin{aligned}&\mu_{\boldsymbol{\eta}_2}^k = \min\left(\hat{\mu}_{\boldsymbol{\eta}_2}^k, \mu_{\boldsymbol{\eta}_2}^{k-1}\right),\\ &\hat{\mu}_{\boldsymbol{\eta}_2}^k = \epsilon_{\boldsymbol{\eta}_2}\times 1.4826\times \mathrm{Med}\left(\left|\hat{\boldsymbol{\eta}}_2^{k-1} - \mathrm{Med}\left(\hat{\boldsymbol{\eta}}_2^{k-1}\right)\right|\right),\end{aligned} \tag{18}$$

where $\epsilon_{\boldsymbol{\eta}_2} > 0$ is a parameter determining the confidence interval range, and Med$(\cdot)$ is the median operator. Generally, MAD measures the dispersion of a set of data in a robust manner. That is to say, for a vector, to ensure its sparsity, a threshold is determined based on the dispersion of its entry values. The large entries above the threshold are remained, and those below that threshold are set to zero. Furthermore, for the update of $\mu_{\boldsymbol{\eta}_2}^k$, we set the sequence $\left\{\mu_{\boldsymbol{\eta}_2}^k\right\}$ non-increasing to avoid the increase of objective function value.

The $\mathbf{x}$-subproblem is

$$\begin{aligned}\mathbf{x}^k = &\arg\min_{\mathbf{x}}\ \boldsymbol{\alpha}_1^{k-1^T}\left(\boldsymbol{\eta}_1^k + \boldsymbol{\eta}_2^k - \mathbf{y} + \mathbf{D}\mathbf{x}\right) + \boldsymbol{\alpha}_3^{k-1^T}\left(\mathbf{x} - \mathbf{g}^{k-1}\right)\\ &+ \frac{\beta^{k-1}}{2}\left(\left\|\boldsymbol{\eta}_1^k + \boldsymbol{\eta}_2^k - \mathbf{y} + \mathbf{D}\mathbf{x}\right\|_2^2 + \left\|\mathbf{x} - \mathbf{g}^{k-1}\right\|_2^2\right)\\ \Rightarrow \mathbf{x}^k = &\arg\min_{\mathbf{x}}\ \frac{\beta^{k-1}}{2}\left(\left\|\boldsymbol{\eta}_2^k - \left(\mathbf{y} - \mathbf{D}\mathbf{x} - \boldsymbol{\eta}_1^k - \frac{\boldsymbol{\alpha}_1^{k-1}}{\beta^{k-1}}\right)\right\|_2^2\right.\\ &\left.+ \left\|\mathbf{x} - \left(\mathbf{g}^{k-1} - \frac{\boldsymbol{\alpha}_3^{k-1}}{\beta^{k-1}}\right)\right\|_2^2\right),\end{aligned} \tag{19}$$

which is a least-squares problem and its closed-form solution is given by

$$\mathbf{x}^k = \left(\mathbf{D}^T\mathbf{D}\right)^{-1}\left[\mathbf{D}^T\left(\mathbf{y} - \boldsymbol{\eta}_1^k - \boldsymbol{\eta}_2^k - \frac{\boldsymbol{\alpha}_1^{k-1}}{\beta^{k-1}}\right) + \mathbf{g}^{k-1} - \frac{\boldsymbol{\alpha}_3^{k-1}}{\beta^{k-1}}\right]. \tag{20}$$

The subproblem of $\mathbf{g}$ is

$$\begin{aligned}\mathbf{g}^k = &\arg\min_{\mathbf{g}}\ \boldsymbol{\alpha}_3^{k-1^T}\left(\mathbf{x}^k - \mathbf{g}\right) + \boldsymbol{\alpha}_4^{k-1^T}\left(\tilde{\mathbf{g}} - \boldsymbol{\theta}^{k-1}\right)\\ &+ \frac{\beta^{k-1}}{2}\left(\left\|\mathbf{x}^k - \mathbf{g}\right\|_2^2 + \left\|\tilde{\mathbf{g}} - \boldsymbol{\theta}^{k-1}\right\|_2^2\right)\\ \Rightarrow &\left\{\mathbf{g}_i^k\right\}_{i=1}^c = \arg\min_{\mathbf{g}_i}\ \sum_{i=1}^c\left[\frac{\alpha_{4,i}^{k-1} - \beta^{k-1}\theta_i^{k-1}}{2\beta^{k-1}}\left\|\mathbf{g}_i\right\|_2\right.\\ &\left.+ \frac{1}{2}\left\|\mathbf{g}_i - \hat{\mathbf{g}}_i^{k-1}\right\|_2^2\right].\end{aligned} \tag{21}$$

where $\hat{\mathbf{g}}_i^{k-1} = \frac{\beta^{k-1}\mathbf{x}_i^k + \alpha_{3,i}^{k-1}}{2\beta^{k-1}}$. We adopt the soft-thresholding operator to solve (21), resulting in [16]:

$$\mathbf{g}_i^k = \max\left(1 - \frac{\alpha_{4,i}^{k-1} - \beta^{k-1}u_i^{k-1}}{2\beta^{k-1}\left\|\hat{\mathbf{g}}_i^{k-1}\right\|_2}, 0\right)\hat{\mathbf{g}}_i^{k-1}. \tag{22}$$

Afterwards, $\mathbf{g}^k = \left[\mathbf{g}_1^{k^T}, \ldots, \mathbf{g}_c^{k^T}\right]^T$.

For $\boldsymbol{\theta}$-subproblem,

$$\begin{aligned}\boldsymbol{\theta}^k = &\arg\min_{\boldsymbol{\theta}}\boldsymbol{\alpha}_2^{k-1^T}\left(\boldsymbol{\upsilon}^{k-1} - \mathbf{w}\odot\boldsymbol{\theta}\right) + \boldsymbol{\alpha}_4^{k-1^T}\left(\tilde{\mathbf{g}}^k - \boldsymbol{\theta}\right)\\ &+ \frac{\beta^{k-1}}{2}\left(\left\|\boldsymbol{\upsilon}^{k-1} - \mathbf{w}\odot\boldsymbol{\theta}\right\|_2^2 + \left\|\tilde{\mathbf{g}}^k - \boldsymbol{\theta}\right\|_2^2\right)\\ \Rightarrow \boldsymbol{\theta}^k = &\arg\min_{\boldsymbol{\theta}} = \frac{\beta^{k-1}}{2}\left(\left\|\boldsymbol{\upsilon}^{k-1} - \mathbf{w}\odot\boldsymbol{\theta} + \frac{\boldsymbol{\alpha}_2^{k-1}}{\beta^{k-1}}\right\|_2^2\right.\\ &\left.+ \left\|\tilde{\mathbf{g}}^k - \boldsymbol{\theta} + \frac{\boldsymbol{\alpha}_4^{k-1}}{\beta^{k-1}}\right\|_2^2\right),\end{aligned} \tag{23}$$

of which the solution is

$$\boldsymbol{\theta}^k = \left(\mathbf{W}^T\mathbf{W} + \mathbf{I}\right)^{-1}\left(\tilde{\mathbf{g}}^k + \mathbf{W}^T\boldsymbol{\upsilon}^{k-1} + \frac{\boldsymbol{\alpha}_2^{k-1} + \mathbf{W}^T\boldsymbol{\alpha}_4^{k-1}}{\beta^{k-1}}\right), \tag{24}$$

where $\mathbf{W} = \mathrm{diag}(\mathbf{w})$.

As for $\boldsymbol{\upsilon}$-subproblem, we handle it in the same way as (16):

$$\begin{aligned}&\boldsymbol{\upsilon}^k = \arg\min_{\boldsymbol{\upsilon}}\ \lambda_2^k\left\|\boldsymbol{\upsilon}\right\|_0 + \boldsymbol{\alpha}_2^{k-1^T}\left(\boldsymbol{\upsilon} - \mathbf{w}\odot\boldsymbol{\theta}^k\right) + \frac{\beta^{k-1}}{2}\left\|\boldsymbol{\upsilon} - \mathbf{w}\odot\boldsymbol{\theta}^k\right\|_2^2\\ &\Rightarrow \boldsymbol{\upsilon}^k = \arg\min_{\boldsymbol{\upsilon}}\ \left(\mu_{\boldsymbol{\upsilon}}^k\right)^2\left\|\boldsymbol{\upsilon}\right\|_0 + \left\|\boldsymbol{\upsilon} - \hat{\boldsymbol{\upsilon}}^{k-1}\right\|_2^2,\end{aligned} \tag{25}$$

where $\hat{\boldsymbol{\upsilon}}^{k-1} = \mathbf{w}\odot\boldsymbol{\theta}^k - \frac{\boldsymbol{\alpha}_2^{k-1}}{\beta^{k-1}}$ and $\left(\mu_{\boldsymbol{\upsilon}}^k\right)^2 = \frac{2\lambda_2^k}{\beta^{k-1}}$. Similar to (16), the solution of (25) is

$$\begin{aligned}&\upsilon_i^k = \psi_{\mu_{\boldsymbol{\upsilon}}^k}\left(\hat{\upsilon}_i^{k-1}\right), \mu_{\boldsymbol{\upsilon}}^k = \min\left(\hat{\mu}_{\boldsymbol{\upsilon}}^k, \mu_{\boldsymbol{\upsilon}}^{k-1}\right),\\ &\hat{\mu}_{\boldsymbol{\upsilon}}^k = \epsilon_{\boldsymbol{\upsilon}}\times 1.4826\times\mathrm{Med}\left(\left|\hat{\boldsymbol{\upsilon}}^{k-1} - \mathrm{Med}\left(\hat{\boldsymbol{\upsilon}}^{k-1}\right)\right|\right).\end{aligned} \tag{26}$$

Finally, the Lagrange multipliers and β are updated as

$$\boldsymbol{\alpha}_1^k = \boldsymbol{\alpha}_1^{k-1} + \beta^{k-1}\left(\boldsymbol{\eta}_1^k + \boldsymbol{\eta}_2^k - \mathbf{y} + \mathbf{D}\mathbf{x}^k\right),\ \boldsymbol{\alpha}_2^k = \boldsymbol{\alpha}_2^{k-1} + \beta^{k-1}\left(\boldsymbol{\upsilon}^k - \mathbf{w}\odot\boldsymbol{\theta}^k\right),$$
$$\boldsymbol{\alpha}_3^k = \boldsymbol{\alpha}_3^{k-1} + \beta^{k-1}\left(\mathbf{x}^k - \mathbf{g}^k\right),\ \boldsymbol{\alpha}_4^k = \boldsymbol{\alpha}_4^{k-1} + \beta^{k-1}\left(\tilde{\mathbf{g}}^k - \boldsymbol{\theta}^k\right),$$
$$\beta^k = \min\left(\beta_{max}, \rho\beta^{k-1}\right). \tag{27}$$

Algorithm 1 ADMM for Solving (11)

Input: Dictionary $\mathbf{D}$, sample $\mathbf{y}$, ε, ϵ_{η_2}, ϵ_{υ_2}, $\mu_{\eta_2}^0 = 10^3$, $\mu_{\upsilon}^0 = 10^3$, $\rho = 1.01$, $\beta^0 = 1$, $\beta_{max} = 10^3$, tolerance error $\varsigma = 10^{-5}$, maximum iteration number K.

Output: $\mathbf{x}^k$

1: Initialize $\mathbf{x}^0 = \mathbf{D}^{-1}\mathbf{y}$, $\boldsymbol{\eta}_1^0 = \mathbf{0}$, $\boldsymbol{\eta}_2^0 = \mathbf{0}$, $\mathbf{g}^0 = \mathbf{x}$, $\boldsymbol{\alpha}_1^0 = \mathbf{0}$, $\boldsymbol{\alpha}_2^0 = \mathbf{0}$, $\boldsymbol{\alpha}_3^0 = \mathbf{0}$, $\boldsymbol{\alpha}_4^0 = \mathbf{0}$, $k = 1$.

2: **while** $\left\|\mathbf{y} - \mathbf{D}\mathbf{x}^k - \boldsymbol{\eta}_1^k - \boldsymbol{\eta}_2^k\right\|_\infty \le \varsigma$, $\left\|\mathbf{x}^k - \mathbf{g}^k\right\|_\infty \le \varsigma$, $\left\|\boldsymbol{\theta}^k - \tilde{\mathbf{g}}^k\right\|_\infty \le \varsigma$, $\left\|\boldsymbol{\upsilon}^k - \mathbf{w}\odot\boldsymbol{\theta}^k\right\|_\infty \le \varsigma$ or $k > K$ **do**

3: Update $\boldsymbol{\eta}_1^k$ by (14), $\boldsymbol{\eta}_2^k$ by (17), $\mathbf{x}^k$ by (20), $\mathbf{g}^k$ by (22), $\boldsymbol{\theta}^k$ by (24), and $\boldsymbol{\upsilon}^k$ by (26), successively.

4: Update $\boldsymbol{\alpha}_1^k$, $\boldsymbol{\alpha}_2^k$, $\boldsymbol{\alpha}_3^k$, $\boldsymbol{\alpha}_4^k$, and β^k by (27).

5: $k = k + 1$.

6: **end while**

The above updating steps are summarized in Algorithm 1.

2.3. Computational complexity analysis

For the proposed algorithm, the main cost is spent on the updates of $\boldsymbol{\eta}_1^k$, $\boldsymbol{\eta}_2^k$, $\mathbf{x}^k$, $\mathbf{g}^k$, $\boldsymbol{\theta}^k$, and $\boldsymbol{\upsilon}^k$. At each iteration, SVD operation for the update of $\boldsymbol{\eta}_1^k$ consumes $\mathcal{O}(m_1 m_2^2)$ (assuming $m_1 \ge m_2$). The updates of $\boldsymbol{\eta}_2^k$ and $\boldsymbol{\upsilon}^k$ require $\mathcal{O}(mlog(m))$ and $\mathcal{O}(clog(c))$, respectively, which are mainly due to the median operation. Update of $\mathbf{x}^k$ mainly involves the matrix multiplication, corresponding to $\mathcal{O}(N^2)$. The update of $\mathbf{g}^k$ costs at most $\mathcal{O}(\hat{N})$, where $\hat{N} = \max\left\{N_1, \ldots, N_i, \ldots, N_c\right\}$. Computing $\boldsymbol{\theta}^k$ costs $\mathcal{O}(c^2)$, which is similar to $\mathbf{x}^k$. Overall, the computational complexity per iteration is $\mathcal{O}(m_1 m_2^2 + mlog(m) + N^2 + c^2)$.

2.4. Stopping criteria

In order to guarantee the convergence of the proposed algorithm using ADMM, proper stopping criteria should be adopted. According to [28], both the primal residuals and dual residuals must be small. Here, we assign the stopping criteria by restraining the primal residuals:

$$\begin{cases} \left\|\mathbf{y} - \mathbf{D}\mathbf{x}^k - \boldsymbol{\eta}_1^k - \boldsymbol{\eta}_2^k\right\|_\infty \le \varsigma, \\ \left\|\mathbf{x}^k - \mathbf{g}^k\right\|_\infty \le \varsigma, \\ \left\|\boldsymbol{\theta}^k - \tilde{\mathbf{g}}^k\right\|_\infty \le \varsigma, \\ \left\|\boldsymbol{\upsilon}^k - \mathbf{w}\odot\boldsymbol{\theta}^k\right\|_\infty \le \varsigma, \end{cases} \tag{28}$$

where ς denotes the error tolerance and is set as 10^{-5} in our experiments. We measure the primal residuals by ℓ_∞-norm, which equals the largest element of a vector. When the ℓ_∞-norm of residuals are below ς, the primal residuals are considered to be small enough, indicating that the solutions reach a satisfactory accuracy level. The dual residual measures the convergence of dual variables and decreases with primal residuals during iterations. Then, when both primal and dual residuals are small enough, the algorithm converges.

3. Robust SR with application to FR

In this section, to demonstrate the effectiveness of the proposed model, we apply our algorithm to FR.

Fig. 1. Face images under different illuminations from ExYaleB.

Table 1
RRs (%) on Set4 and Set5 of ExYaleB. Set1 is used for training.

Method	Set4	Set5
ACPRC	78.76	25.62
NMR	89.85	43.35
WMNR	91.54	84.76
EGSNR	96.80	93.21
Proposed	**97.18**	**93.49**

Table 2
RRs (%) on Set3 of ExYaleB with different percent block or baboon masks. Set1 is used for training.

Method	Block mask percent			Baboon mask percent		
	40%	50%	60%	40%	50%	60%
ACPRC	50.44	39.25	28.73	69.96	48.47	30.70
NMR	99.12	96.05	87.50	97.15	90.35	70.40
WMNR	98.46	95.39	84.21	**100**	98.25	92.98
EGSNR	**100**	**99.78**	94.52	99.78	99.78	97.59
Proposed	**100**	**99.78**	**97.59**	**100**	**100**	**98.90**

3.1. Experimental sketch

An $m_1 \times m_2$ grayscale face image is represented as a column vector $\mathbf{d} \in \mathbb{R}^m$ with $m = m_1 m_2$ produced by stacking the columns of the face image. Given a face image dataset containing c objects, we select N_i images for each object i and construct $\mathbf{D}_i = [\mathbf{d}_1, \ldots, \mathbf{d}_{N_i}]$. With $\mathbf{D} = \left[\mathbf{D}_1, \ldots, \mathbf{D}_c\right]$, a sparse representation coefficient $\mathbf{x}$ for a test face image $\mathbf{y}$ is obtained by Algorithm 1. Then, we recognize $\mathbf{y}$ as object i according to

$$\min_i \left\|\mathcal{M}\left(\mathbf{y} - \mathbf{D}_i\mathbf{x}_i\right)\right\|_{\text{Log}}, \tag{29}$$

which is similar to the strategy in [15,20]. As for the evaluation metric, we suggest the recognition rate (RR) defined as $\text{RR} = N_{\text{suc}}/N_{\text{test}}$. Here, N_{test} denotes total number of test samples, and N_{suc} is the number of correctly classified samples.

3.2. Experimental results

In this subsection, we compare the proposed algorithm with several SR based FR algorithms, viz. ACPRC [17], nuclear norm-based matrix regression (NMR) [20], weighted mixed-norm regression (WMNR) [19], and enhanced group sparse regularized nonconvex regression (EGSNR) [15] using face image datasets Extended Yale B (ExYaleB) [31] and CMU PIE [32]. The images of ExYaleB and CMU PIE are resized to dimensions 48×42 and 50×40, respectively. For parameter selection of suggested algorithm, ϵ_{η_2} and ϵ_{υ_2} are selected in $[1, 10]$, and proper ε is in $[0.01, 1]$.

We first evaluate the performance of our method under illumination changes on ExYaleB. This dataset contains frontal face images of 38 objects and is divided into 5 sets according to the illumination conditions.

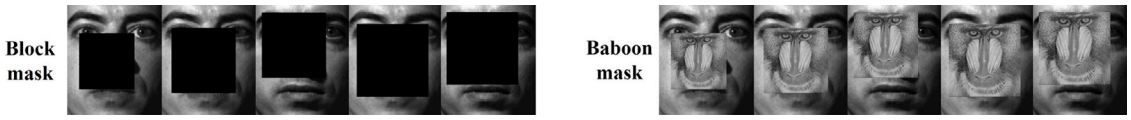

Fig. 2. Face images from ExYaleB with block mask (left) and Baboon mask (right) for different mask percents.

Fig. 3. Face images from ExYaleB with 40% impulsive noise and 40% Baboon mask.

Table 3
RRs (%) on Set3 of ExYaleB with 30% impulsive noise and 30% Baboon mask (denoted as 30%), and 40% impulsive noise and 40% Baboon mask (denoted as 40%). Set1 is used for training.

Method	30%	40%
ACPRC	49.12	41.01
NMR	77.19	64.25
WMNR	84.43	69.52
EGSNR	**94.96**	85.75
Proposed	**94.96**	**85.96**

Fig. 4. Face images with different expressions from CMU PIE.

Table 4
RRs (%) on CMU PIE for objects with different expressions.

Method	CMU PIE
ACPRC	91.85
NMR	92.66
WMNR	92.39
EGSNR	91.85
Proposed	**94.02**

Set1 has the best illumination condition, and the condition of Set5 is the worst. Sample images from Set1, Set4, and Set5 are shown in Fig. 1. We use Set1 for training, Set4 and Set5 for testing. The RRs of different algorithms are tabulated in Table 1. It is observed that the proposed method achieves the best performance. From Fig. 1, we see that some images of Set5 are difficult for human to recognize due to the large areas of shades. For that situation, our method still acquires a high RR.

Next, we conduct test on ExYaleB under occlusions. Set1 is used for training. The test data are generated by randomly adding 40%–60% square block or baboon mask to images of Set3. Sample test images with block mask or baboon mask are displayed in Fig. 2. The RRs are tabulated in Table 2. It is observed that the proposed method is more robust to the occlusion ratio changes and attains higher RRs than other competing algorithms.

Then, we verify the effectiveness of our algorithm on simultaneously tackling correlated interference and sparse outlier. The training data are images of Set1. As for the test data, we add 30% impulsive noise and 30% baboon mask jointly, 40% impulsive noise and 40% baboon mask jointly to images from ExYaleB Set3. Examples of generated test images are shown in Fig. 3. The RRs are tabulated in Table 3, and it is observed that our algorithm performs the best.

One more experiment on CMU PIE is conducted. CMU PIE dataset consists of more than 40,000 face images captured under varying poses, illuminations, as well as expressions. We test the performance of the proposed algorithm under different expressions. We select 46 objects, each of which 11 images with different expressions are employed as the experimental data. Sample images are shown in Fig. 4. For each object, 3 images are randomly chosen as the training data, and the remaining 8 images are used for testing. Comparing the RRs of different algorithms listed in Table 4, we see that our method achieves the best performance.

4. Conclusion

A new model for robust SR is proposed, where the fitting error is decomposed into a low-rank component and a sparse term, which are regularized by the logarithmic norm and ℓ_0-norm, respectively. As for the combination coefficient, we adopt $\ell_{2,0}$-norm minimization for group sparsity. Then, to verify its effectiveness, the proposed robust SR model is applied to FR. The experimental results demonstrate that our algorithm is superior over several competing methods in terms of RR.

CRediT authorship contribution statement

Xiang-Yu Wang: Writing – original draft, Software, Methodology, Investigation, Data curation, Conceptualization. **Xiao-Peng Li:** Writing – review & editing, Supervision, Methodology. **Hing Cheung So:** Writing – review & editing, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgment

The work described in this paper was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 11207922) and a grant from the Young Innovative Talents Project of Guangdong Provincial Department of Education (Natural Science), China (Project No. 2023KQNCX063).

References

[1] X. Wang, W. Wang, J. Liu, X. Li, J. Wang, A sparse representation scheme for angle estimation in monostatic MIMO radar, Signal Process. 104 (2014) 258–263.
[2] X.-P. Li, Z.-L. Shi, L. Huang, A.M.-C. So, H.C. So, ROCS: Robust one-bit compressed sensing with application to direction of arrival, IEEE Trans. Signal Process. (2024) 1–14, (Early Access).
[3] C. Xing, M. Wang, C. Dong, C. Duan, Z. Wang, Joint sparse-collaborative representation to fuse hyperspectral and multispectral images, Signal Process. 173 (2020) 107585.
[4] H. Cheng, Z. Liu, L. Yang, X. Chen, Sparse representation and learning in visual recognition: Theory and applications, Signal Process. 93 (6) (2013) 1408–1425.

[5] E.J. Candes, T. Tao, Near-optimal signal recovery from random projections: Universal encoding strategies? IEEE Trans. Inform. Theory 52 (12) (2006) 5406–5425.
[6] D.L. Donoho, M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization, Proc. Natl. Acad. Sci. USA 100 (5) (2003) 2197–2202.
[7] S. Mukherjee, R. Basu, C.S. Seelamantula, ℓ_1-K-SVD: A robust dictionary learning algorithm with simultaneous update, Signal Process. 123 (2016) 42–52.
[8] A.-K. Seghouane, A. Iqbal, A.M. Rekavandi, RBDL: Robust block-structured dictionary learning for block sparse representation, Pattern Recognit. Lett. 172 (2023) 89–96.
[9] P.A. Forero, S. Shafer, J.D. Harguess, Sparsity-driven Laplacian-regularized outlier identification for dictionary learning, IEEE Trans. Signal Process. 65 (14) (2017) 3803–3817.
[10] A. Iqbal, A.-K. Seghouane, An α-divergence-based approach for robust dictionary learning, IEEE Trans. Image Process. 28 (11) (2019) 5729–5739.
[11] E. Amaldi, V. Kann, On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems, Theoret. Comput. Sci. 209 (1–2) (1998) 237–260.
[12] D. Vidaurre, C. Bielza, P. Larranaga, A survey of L_1 regression, Int. Stat. Rev. 81 (3) (2013) 361–387.
[13] S. Guo, Z. Wang, Q. Ruan, Enhancing sparsity via ℓ^p $(0 < p < 1)$ minimization for robust face recognition, Neurocomputing 99 (2013) 592–602.
[14] Y. Hu, C. Li, K. Meng, J. Qin, X. Yang, Group sparse optimization via $\ell_{p,q}$ regularization, J. Mach. Learn. Res. 18 (1) (2017) 960–1011.
[15] C. Zhang, H. Li, C. Chen, Y. Qian, X. Zhou, Enhanced group sparse regularized nonconvex regression for face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 44 (5) (2022) 2438–2452.
[16] W. Deng, W. Yin, Y. Zhang, Group Sparse Optimization by Alternating Direction Method, Tech. Rep. TR11-06, Dept. Comput. Appl. Math., Rice Univ., Houston, TX, USA, 2011.
[17] J.-X. Mi, Q. Fu, W. Li, Adaptive class preserving representation for image classification, in: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Honolulu, Hawaii, USA, 2017, pp. 7427–7435.
[18] S.O. Rajankar, S.N. Talbar, An electrocardiogram signal compression techniques: A comprehensive review, Analog Integr. Circuits Signal Process. 98 (2019) 59–74.
[19] J. Zheng, K. Lou, X. Yang, C. Bai, J. Tang, Weighted mixed-norm regularized regression for robust face identification, IEEE Trans. Neural Netw. Learn. Syst. 30 (12) (2019) 3788–3802.
[20] J. Yang, L. Luo, J. Qian, Y. Tai, F. Zhang, Y. Xu, Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes, IEEE Trans. Pattern Anal. Mach. Intell. 39 (1) (2016) 156–171.
[21] I. Naseem, R. Togneri, M. Bennamoun, Robust regression for face recognition, Pattern Recognit. 45 (1) (2012) 104–118.
[22] R. He, W.-S. Zheng, B.-G. Hu, Maximum correntropy criterion for robust face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 33 (8) (2011) 1561–1576.
[23] A.B. Ramirez, R.E. Carrillo, G. Arce, K.E. Barner, B. Sadler, An overview of robust compressive sensing of sparse signals in impulsive noise, in: Proc. European Signal Processing Conf., EUSIPCO, Nice, France, 2015, pp. 2859–2863.
[24] J. Tang, X. Shu, Z. Li, G.-J. Qi, J. Wang, Generalized deep transfer networks for knowledge propagation in heterogeneous domains, ACM Trans. Multimed. Comput. Commun. Appl. 12 (4s) (2016) 68.
[25] B. Recht, M. Fazel, P.A. Parrilo, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, SIAM Rev. 52 (3) (2010) 471–501.
[26] A.M. Zoubir, V. Koivunen, E. Ollila, M. Muma, Robust Statistics for Signal Processing, Cambridge University Press, 2018.
[27] L. Chen, X. Jiang, X. Liu, Z. Zhou, Logarithmic norm regularized low-rank factorization for matrix and tensor completion, IEEE Trans. Image Process. 30 (2021) 3434–3449.
[28] 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) 1–122.
[29] M. Yang, Q. Luo, W. Li, M. Xiao, 3-D array image data completion by tensor decomposition and nonconvex regularization approach, IEEE Trans. Signal Process. 70 (2022) 4291–4304.
[30] X.P. Li, Z.-L. Shi, C.-S. Leung, H.C. So, Sparse index tracking with K-sparsity or ε-deviation constraint via ℓ_0-norm minimization, IEEE Trans. Neural Netw. Learn. Syst. 34 (12) (2023) 10930–10943.
[31] K.-C. Lee, J. Ho, D.J. Kriegman, Acquiring linear subspaces for face recognition under variable lighting, IEEE Trans. Pattern Anal. Mach. Intell. 27 (5) (2005) 684–698.
[32] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker, Multi-PIE, Image Vis. Comput. 28 (5) (2010) 807–813.