Low-Rank Matrix Recovery from Noisy via an MDL Framework-based Atomic Norm

Anyong Qin, Lina Xian, Yongliang Yang, Taiping Zhang, and Yuan Yan Tang, Fellow, IEEE

Abstract—The recovery of the underlying low-rank structure of clean data corrupted with sparse noise/outliers is attracting increasing interest. However, in many low-level vision problems, the exact target rank of the underlying structure, the particular locations and values of the sparse outliers are not known. Thus, the conventional methods cannot separate the low-rank and sparse components completely, especially gross outliers or deficient observations. Therefore, in this study, we employ the Minimum Description Length (MDL) principle and atomic norm for low-rank matrix recovery to overcome these limitations. First, we employ the atomic norm to find all the candidate atoms of low-rank and sparse matrix, and then we minimize the description length of the model in order to select the appropriate atoms of low-rank and the sparse matrix, respectively. Our experimental analyses show that the proposed approach can obtain a higher success rate than the state-of-the-art methods even when the number of observations is limited or the corruption ratio is high. Experimental results about synthetic data and real sensing applications (high dynamic range imaging, background modeling, removing shadows and specularities) demonstrate the effectiveness, robustness and efficiency of the proposed method.

Index Terms—atomic norm, low-rank matrix recovery, minimum description length principle, robust principal components analysis.

I. INTRODUCTION

LOW-rank matrix recovery is important in many fields, such as image processing and computer vision [1]–[3], pattern recognition and machine learning [4]–[6], and many other applications [7]–[9]. Due to the sensor or environmental reasons, the observations used in these fields are readily corrupted by noise or outliers, so the given data matrix $Y$ can be decomposed into low-rank and sparse components.

Principal Component Analysis (PCA) [10] has been used widely to search for the best approximation of the underlying structure (unknown low-rank matrix $X$) of the given data. In addition, stable performance can be obtained via the Singular Value Decomposition (SVD) when the data are corrupted only by a small amount of noise. Due to the presence of gross outliers in modern applications, the robust variant of PCA, called robust PCA (RPCA) has also been used to reject the outliers [11], [12]

$$\arg\min_{X,E} \text{rank}(X) + \gamma \|E\|_0, \quad \text{s.t. } Y = X + E,$$

where the parameter $\gamma > 0$ is a regularization parameter, $\text{rank}(X)$ denotes the rank of matrix $X \in \mathbb{R}^{n \times n}$ ($\text{rank}(X) = r$), and $\|E\|_0$ is the number of non-zero entries in the sparse matrix $E$. Unfortunately, solving Eq. (1) is an NP-hard problem. Instead, Candès et al. [12] solved an approximated problem by convex optimization under rather weak assumptions

$$\arg\min_{X,E} \|X\|_* + \gamma \|E\|_1, \quad \text{s.t. } Y = X + E,$$

where $\|X\|_* = \sum_{i} \sigma_i(X)$ is the nuclear norm of $X$ ($\sigma_i(X)$ denotes the $i$-th singular value of $X$), and $\|E\|_1$ represents the $l_1$-norm of the sparse matrix $E$. Various approaches can be used to solve Eq. (2) effectively [13], [14].

Wright et al. [11] and Candès et al. [12] proved that the performance of Eq. (2) will approach stability only by using more observations (larger $n$). However, the number of observations ($n$) is typically limited in many image processing and computer vision problems due to physical constraints. Moreover, when the number ($n$) is very limited, we note that existing methods based on Eq. (2) do not reject some outliers well, such as moving objects in surveillance video [15]–[17], shadows in face images [12], and saturations in low dynamic range (LDR) images [2], [18], [19].

It is well known that the rank ($r$) of $X$ and the $\gamma$ both influence the final results of RPCA decomposition. Unfortunately, the target rank and randomizing parameter $\gamma$ are uncertain in Eq. (1) and Eq. (2), where conventional approaches need to tune the rank of $X$ and $\gamma$ to achieve the desired goal. However, $\gamma = 1/\sqrt{\max\{m,n\}}$, that the typical approaches set, is not the best value [15]. Instead, Ramirez et al. [15], [20] used the Minimum Description Length (MDL) principle [21] to avoid estimating the parameter $\gamma$. The MDL principle selects the best low-rank approximation from an RPCA decomposition sequences, which are obtained via the different values of $\gamma$. Liu et al. [17] employed structured sparse decomposition to solve the regularizing parameter issue in RPCA, where they replaced the static parameter $\gamma$ by adaptive settings for image regions with distinct properties in each frame. However, an accurate rank is crucial for recovering the low-rank matrix and rejecting the outliers completely. An example of a scene
is shown in Fig. 1. The RPCA fails to recover the low-rank matrix and capture the illumination sudden change.

Atoms are the fundamental basis of the representation of a signal. The atomic norm hull is the set of the fundamental elements. Moreover, the atomic norm induced by the convex hull of all unit-norm one-sparse vectors is the $l_1$-norm, and the nuclear norm is induced by taking the convex hull of an atomic set, in which the elements are all unit rank matrices [22]–[24]. To address these issues, such as the limited number of observations, the rank of $X$, and the regularizing parameter $\gamma$, we propose a low-rank model based on the MDL principle within the devised atomic norm (MDLAN), which is also an expanded version of our published conference paper [25]. In our proposed method, we minimize the description length to select the optimum atomic sets for the low-rank matrix ($X$) and structured sparse matrix ($E$), respectively. In contrast to [15], we use the MDL principle to determine the number of atoms in the low-rank matrix, thereby avoiding tuning the rank of low-rank matrix $X$, and we also recover the sparse matrix $E$ via MDL principle. Experimental analyses show that our method can obtain a better approximation of the underlying structure of the given data when the number of observed samples is limited or if the samples have gross outliers. Thus, the proposed framework provides a nonparametric, robust low-rank matrix recovery algorithm.

The main contributions of this paper are summarized as follows:

1. We present an MDL principle based atomic norm method for low-rank matrix recovery. Unlike other model selection algorithms, the proposed MDLAN uses the description length as a cost function to select the two smallest sets of atoms that can span the low-rank matrix and sparse matrix, respectively.

2. We empirically test the MDL framework based atomic norm and find that it outperforms the state-of-the-art methods when the number of observations is limited or if the observations have gross outliers.

3. It is difficult to address the original optimization problem for MDLAN due to the combination of description length and the atomic norm. Thus, we devise a new ADMM based algorithm that considers an approximation of the original non-convex problem.

The remainder of this paper is organized as follows. Section 2 briefly reviews some related researches. In Section 3, we present the unified framework for low-rank matrix recovery based on the atomic norm. In Section 4, we describe the proposed MDLAN method. Section 5 presents the experimental results based on synthetic and real datasets. Finally, we give our conclusions in Section 6.

II. RELATED WORK

In the following, we briefly review recent advances in RPCA and discuss its applications in image processing and computer vision.

To exactly recover $X$, some studies have replaced the rank ($\cdot$) with the nuclear norm and the number of nonzero entries with the $l_1$-norm in Eq. (2). Candès et al. [12] proved that the rank minimization problem can be solved using Eq. (1), and it can be solved in a tractable manner by the convex relaxation version of Eq. (2). They also proved that the unique solution of Eq. (2) corresponds exactly to the solution of the original NP-hard problem in Eq. (1) under suitable conditions.

Recently, the improvements to RPCA are generally divided into two categories. One category is focus on the structured sparse component $E$ in Eq. (2) [26], [27]. For example, Xin et al. [28] replaced $l_1$-norm with an adaptive version of the generalized fused lasso (GFL) regularization [29], which takes into account the spatial neighborhood information of the foregrounds in a video sequence.

$$\arg\min_{X,E} \|X\|_* + \gamma \|E\|_{\text{gfl}}, \quad \text{s.t.} \quad Y = X + E,$$

where the generalized fused lasso $\|E\|_{\text{gfl}}$ can be viewed as a combination of two common regularizer, i.e. the $l_1$-norm and the total variation (TV) penalty [30].

$$\|E\|_{\text{gfl}} = \sum_{l=1}^{n} \left\{ \|e(l)\|_1 + \lambda_1 \sum_{(i,j)\in\mathcal{N}} w_{ij} |f_i^{(l)} - f_j^{(l)}| \right\}$$

where $e(l)$ is the $l$-th column of the sparse matrix $E$, $\mathcal{N}$ is the spatial neighborhood set, $\lambda_1$ is a tuning parameter, and $w_{ij} = \exp(-\|w^{(l)}_i - w^{(l)}_j\|_2^2/\sigma^2) \geq 0$ is a tuning parameter empirically set.

Another category is focus on the low-rank component $X$ in Eq. (2) [34]. For example, Cabral et al. [35] and Guo et al. [36] replaced the $X$ with $UV$, and the relationship $\|X\|_* = \min_{U,V} \|U\|_F^2 + \|V\|_F^2$ holds, where $U \in \mathbb{R}^{m \times r}$, $V \in \mathbb{R}^{r \times n}$ and $\| \cdot \|_F$ represents the Frobenius norm. In addition, Guo et al. [36] employed an entropy term to restrict the support of the outliers. Hu et al. [37] proposed an approximation of the target rank by the truncated nuclear norm, which only minimizes the smallest $\min(m,n) - r$ singular values. T-H Oh et al. [2] proposed to minimize the partial sum of the singular values instead of minimizing the nuclear norm. Thus, the formulation of the partial sum can be written as follows

$$\arg\min_{X,E} |rank(X) - r| + \gamma \|E\|_1, \quad \text{s.t.} \quad Y = X + E$$

The rank minimization algorithms for RPCA have inspired many applications in image processing and computer vision, such as image alignment [38], background subtraction [12], [17], high dynamic range (HDR) imaging [2], [39], and image restoration [40], [41]. However, the clean data are always corrupted by gross noise/outliers or the number of given data is limited due to the sensor reasons or human error [2], [11], [42]. The available methods based on RPCA have difficulty solving these situations. In the present study, we propose an algorithm based on MDL and the atomic norm to overcome these difficulties, i.e., unknown of the target rank $r$, the regularizing parameter $\lambda$, and the deficient observations or gross outliers.
can be expressed as

\[ \| \cdot \|_A \]

the matrix, and

\[ X \]

where the inner product is defined as

\[ \langle \cdot, \cdot \rangle \]

Frobenius norm. For any matrix \( X \), \( \| X \|_F \) is defined by

\[ \| X \|_F = \sqrt{\sum_{i,j} X_{ij}^2} \]

where \( X = \sum_{i,j} X_{ij} E_{ij} \) and the others are zeros. Any \( k \)-sparse matrix in \( \mathbb{R}^{m \times n} \) is a linear combination of \( k \) elements from the atomic set defined above.

**Sparsity inducing norm:** The sparsity inducing atomic set can be expressed as

\[ A_S := \{ \pm E_{ij} \in \mathbb{R}^{m \times n}, i = 1, 2, \ldots, m, j = 1, 2, \ldots, n \} \]

where \( E_{ij} \) denotes a matrix, the \((i,j)\)-th entry of which is 1 and the others are zeros. Any \( k \)-sparse matrix in \( \mathbb{R}^{m \times n} \) is a linear combination of \( k \) elements from the atomic set defined above.

**Low-rankness inducing norm:** The low-rankness inducing atomic set can be written as

\[ A_L := \{ Z \in \mathbb{R}^{m \times n} | \text{rank}(Z) = 1, \| Z \|_F = 1 \} \]

where \( Z \in \mathbb{R}^{m \times n} \) represents a rank-1 matrix with unit Frobenius norm. For any matrix \( X \in \mathbb{R}^{m \times n} \), \( \| X \|_{A_L} = \| X \|_* = \sum_i \sigma_i(X) \) (\( \sigma_i(X) \) denotes the \( i \)-th singular value of the matrix \( X \)).

### III. LOW RANK MATRIX RECOVERY VIA THE ATOMIC NORM

#### A. Atomic norm

First, we provide a definition of an atomic norm and some assumptions regarding the set of atoms \( (A) \). We also assume that the set \( A \) is origin-symmetric (i.e., \( A \in A \) if and only if \(-A \in A \)). The atomic norm [22] is the gauge functional induced by \( A \):

\[ \| X \|_A := \inf_{t > 0} \{ t : X \in t \cdot \text{conv}(A) \} \]

where \( \text{conv}(A) \) denotes the convex hull of \( A \). In fact, the atomic norm is changed into many familiar norms when we specify the atomic set. The dual norm of \( \| \cdot \|_A \) is defined by

\[ \| X \|_A^* := \sup \{ \langle X, A \rangle, a \in A \} \]

where the inner product is defined as \( \langle X, A \rangle = \text{tr}(X^T A) \) for the matrix, and \( \text{tr}(\cdot) \) denotes the trace of a matrix. The dual atomic norm is crucial for producing the atomic set in our case.

**Sparsity inducing norm:** The sparsity inducing atomic set can be expressed as

\[ A_S := \{ \pm E_{ij} \in \mathbb{R}^{m \times n}, i = 1, 2, \ldots, m, j = 1, 2, \ldots, n \} \]

where \( E_{ij} \) denotes a matrix, the \((i,j)\)-th entry of which is 1 and the others are zeros. Any \( k \)-sparse matrix in \( \mathbb{R}^{m \times n} \) is a linear combination of \( k \) elements from the atomic set defined above.

**Low-rankness inducing norm:** The low-rankness inducing atomic set can be written as

\[ A_L := \{ Z \in \mathbb{R}^{m \times n} | \text{rank}(Z) = 1, \| Z \|_F = 1 \} \]

where \( Z \in \mathbb{R}^{m \times n} \) represents a rank-1 matrix with unit Frobenius norm. For any matrix \( X \in \mathbb{R}^{m \times n} \), \( \| X \|_{A_L} = \| X \|_* = \sum_i \sigma_i(X) \) (\( \sigma_i(X) \) denotes the \( i \)-th singular value of the matrix \( X \)).

#### B. Atomic norm based low-rank matrix recovery

The unified form of the low-rank model (Eq. (2)) based on the atomic norm can also be expressed as follows

\[ \arg\min_{X, E} \| X \|_{A_L} + \lambda \| E \|_{A_S} \quad \text{s.t.} \quad Y = X + E \quad (5) \]

where \( \lambda \) is the regularizing parameter. To simplify the presentation, we define a linear operator as follows.

**Definition 1:** Given a set \( \Psi = \{ \psi_1, \ldots, \psi_{|\Psi|} \} \subset \mathbb{R}^{m \times n} \), we define a linear operator \( F_\Psi : \mathbb{R}^{|\Psi|} \rightarrow \mathbb{R}^{m \times n} \) by

\[ F_\Psi \alpha = \sum_{k=1}^{|\Psi|} \alpha_k \psi_k \quad \forall \alpha \in \mathbb{R}^{|\Psi|} \quad (6) \]

From Eq. (6), it follows that the adjoint operator \( F_\Psi^* : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{|\Psi|} \) is given by

\[ F_\Psi^* X = [\langle X, \psi_1 \rangle, \ldots, \langle X, \psi_{|\Psi|} \rangle] \quad \forall X \in \mathbb{R}^{m \times n} \quad (7) \]

From Definition 1, it follows that the specific forms of the atomic norm in Eq. (5) are given by

\[ \| X \|_{A_L} := \inf_{\| \alpha \|} \{ \sum_i \alpha_i : X = F_\Psi \alpha, \alpha_i \geq 0, \Psi \subset A_L \} \quad (8) \]

\[ \| E \|_{A_S} := \inf_{\| \beta \|} \{ \sum_i \beta_i : E = F_\Psi^\beta, \beta_i \geq 0, \Phi \subset A_S \} \quad (9) \]

where \( \alpha, \beta \) are the vectors of the scalar coefficients, and \( \alpha = \{ \alpha_1, \alpha_2, \ldots, \alpha_{|\Psi|} \}, \beta = \{ \beta_1, \beta_2, \ldots, \beta_{|\Psi|} \} \).

### IV. MDL PRINCIPLE BASED LOW-RANK RECOVERY

In this section, we first present the concept of MDL principle and its background. We then propose a new low-rank matrix recovery method (MDLAN) based on the MDL principle and atomic norm, as well as the optimization algorithm.

Fig. 1: Recovered background and detection of outliers. Three frames from an 80-frame sequences taken in a lobby. (a) Three frames from the original video \( Y \), low-rank component \( X \) (b,d) and structured sparse component \( E \) (c,e) obtained by RPCA (b,c), and the proposed approach (d,e), respectively. The rank estimated by RPCA is 7, so ghosting appeared in the background. By contrast, the rank estimated by our approach is 1.
A. Minimum description length principle

The MDL principle works as an objective function that balances a measure of the goodness of fit with the model complexity, and searches for a model $M$ from the set of possible models, $\mathcal{M}$. In the MDL framework, a model $M$ in $\mathcal{M}$ that describes the given data $Y$ completely with the fewest number of bits is considered the best. The MDL problem is formulated as follows:

$$\hat{M} = \arg\min_{M \in \mathcal{M}} L(Y, M)$$  \hspace{1cm} (10)

where the codelength assignment function $L(Y, M)$ defines the theoretical codelength required to describe $(Y, M)$ uniquely.

A common implementation of the MDL framework uses the Ideal Shannon Codelength Assignment [43, Ch.5] to define $L(Y, M)$ in terms of a probability assignment $P(Y, M)$ as

$$L(Y, M) = -\log P(Y, M) = -\log(P(Y|M)P(M))$$

Thus, we obtain the MDL framework

$$\hat{M} = \arg\min_{M \in \mathcal{M}} -\log P(M) - \log P(Y|M)$$  \hspace{1cm} (11)

where $-\log P(M)$ represents the model complexity and $-\log P(Y|M)$ represents a measure of the goodness of fit.

B. The proposed method

Our family of models for expressing the low-rank matrix recovery problem are defined by $\mathcal{M} = \{(X, E) : Y \leftarrow X + E, \text{rank}(X) = r, \|E\|_0 = k\}$, where $r$ is the truthful rank of low-rank matrix $X$, and $k$ represents the truthful number of non-zero entries in the sparse matrix $E$. Using these definitions, our objective function in the MDL framework can be formulated as follows

$$(\hat{X}, \hat{E}) = \arg\min_{X, E} L(Y, M)$$

$$= \arg\min_{X, E} L(X) + L(E) + L(Y - X - E)$$  \hspace{1cm} (12)

Combining Eq. (8), Eq. (9) and Eq. (12) yields the following MDL based atomic norm for low-rank matrix recovery (MDLAN) model

$$\arg\min_{\Psi, \Phi} \sum_i L(\alpha_i \psi_i) + L(\mathcal{F}_q \beta) + L(Y - \mathcal{F}_q \alpha - \mathcal{F}_q \beta)$$  \hspace{1cm} (13)

The basic idea of the proposed MDLAN is to find two smallest sets $\Psi$ and $\Phi$, while minimizing the $L(Y - \mathcal{F}_q \alpha - \mathcal{F}_q \beta)$. The cost function in Eq. (13) is non-convex in $(\Psi, \Phi)$ and we relax it with an alternative objective function in order to effectively handle the proposed problem.

C. Encoding scheme

In our MDL framework, we need to encode the low-rank matrix ($\sum \alpha_i \psi_i$) and the sparse matrix ($L(\mathcal{F}_q \beta)$), respectively. It is usual to extend the ideal codelength to continuous random variables $x$ with a probability assignment $P(x)$ as $L(x) = -\log P(x) \approx -\log(p(x)\delta)$, and $p(x)$ is the probability density function of variables $x$. To losslessly encode the finite-precision obtained variables, we quantize the variables with step $\delta = 1$ [20].

![Fig. 2: The encoding procedure of prediction scheme](image)

1) Encoding the sparse matrix: For sake of simplicity, we set the elements in atomic set $\Phi$ to be positive signs, and the scalar coefficients $\beta$ to be mix signs. We assume that the scalar coefficients $\beta$ comprise a sequence of Laplace random variables [20]

$$L(\mathcal{F}_q \beta) = L(\beta | \Phi) + L(\Phi)$$

$$= -\log P(\beta | \theta) + k \log(\text{mn})$$

$$= \theta \|\beta\|_1 + c$$

where each atom $\phi_i \in \Phi$ has only one nonzero entry. The $\phi_i$ only describes the index of the nonzero position, and therefore $c$ is a fixed constant (the description length of $\phi_i$ is $\log(\text{mn})$). Moreover, $\theta^t = \frac{1}{k} \sum_{i=1}^{k} |\beta_i|^{-1}$ is the MLE of $\theta$ (the parameter of the Laplacian) based on $\beta^t$. So minimizing the description length of sparse matrix ($L(\mathcal{F}_q \beta)$) is replaced by minimizing $\theta \|\beta\|_1 (\|E\|_1 = \|\mathcal{F}_q \beta\|_1 = \|\beta\|_1)$.

2) Encoding the low-rank matrix: It is not surprising that each atom represents numerous Eigen-information of the low-rank matrix. In other words, in the case of our real world applications, we can suppose that the columns of an atom are standard static images which are piecewise smooth. So we should efficiently exploit the smoothness of the atoms via employing a prediction scheme. Concretely, to describe the each column ($\psi_i = [a_1, a_2, \ldots, a_m]$) of the atoms, we reshape it ($a_j$) as an image or frame of the same size as the original images or frames in the observed matrix $Y$, respectively. Then we employ a causal bilinear kernel with zero-padding to produce a predicted vector $\tilde{a}_j$ (each element is given by $\text{north}\_\text{element} + \text{west}\_\text{element} - \text{northwest}\_\text{element}$), obtaining the residual $a_j = a_j - \tilde{a}_j$. In particular, the residual can be written as $\tilde{a}_j = W a_j$, and the matrix residual can be formed as $\tilde{\psi}_i = W \psi_i$, where $W \in \mathbb{R}^{m \times m}$ is lower triangular. The detailed procedure is depicted in Fig. 2. We refer to [15], [20] for details on these results.

We also assume the prediction residual $\tilde{\psi}_i$ to be a sequence of LG distributed continuous random variables [15], [20]. Compared to the codelength of $\psi_i$, the codelength of scalar coefficient $\alpha_i$ is inconsequential for our model. So the codelength of low-rank matrix $X$ can be written as $\sum_i L(W \psi_i)$. 

D. Optimization by ADMM

Just like other researches [1], [2], [42], in this work the equation \( Y = X + E \) still holds. Then the original problem (Eq. (13)) can be reformed as follows
\[
(X, E) = \arg\min_{X, E} \sum_i \mathcal{L}(W \psi_i) + \theta \|E\|_1
\]
subject to \( Y = X + E \) \( \tag{15} \)

where the matrix \( X = F_{\Phi} \alpha \) and set \( \Psi = \{\psi_1, \psi_2, \ldots, \psi_r\} \). Here we employ the alternating direction method of multipliers (ADMM) method [13], [44] to solve this constrained optimization problem. The Augmented Lagrangian function of Eq. (15) is
\[
L_{\mu}(X, E, U) = \sum_i \mathcal{L}(W \psi_i) + \theta \|E\|_1 + \langle U, Y - X - E \rangle
+ \mu \|Y - X - E\|_F^2 \tag{16} \]

where \( \mu \) is a positive scalar, \( U \in \mathbb{R}^{m \times n} \) is the Lagrange multiplier and \( \langle \cdot, \cdot \rangle \) denotes the inner product operator. The ADMM consists of the following iterations
\[
X^{t+1} = \arg\min_X L_{\mu t}(X, E^t, U^t) \tag{17} \\
E^{t+1} = \arg\min_E L_{\mu t}(X^{t+1}, E, U^t) \tag{18} \\
U^{t+1} = U^t + \mu (Y - X^{t+1} - E^{t+1}) \tag{19} \\
\]

the two subproblems (Eq. (17) and Eq. (18)) are convex optimization problems while fixing the other variables. Algorithm 1 summarizes the whole recovery procedure, recovering the low-rank matrix and rejecting the sparse outliers alternately. The logic that underlies the proposed MDLAN method is that the code-length cost of adding a new atom to the model is usually very high, so adding a new atom is only reasonable if its contribution is sufficiently high to produce the largest decrease in the other part, i.e., the constrained term \( Y - X - E = 0 \).

1) Recovering the low-rank matrix: The subproblem in Eq. (17) can be formulated as follows
\[
\arg\min_X \frac{1}{\mu^t} \sum_i \mathcal{L}(W \psi_i) + \frac{1}{2} \|X - G^t\|_F^2 \tag{20} \]

where \( G^t = Y - E^t + \frac{1}{\mu^t} U^t \). When we obtain the candidate atomic set of the low-rank matrix, we only need to select the suitable atoms. Since the low-rank matrix is a combination of \( r \) atoms, so we first determine the candidate set \( \Psi \) by the dual atomic norm.
\[
\|X\|_{A_L} := \sup \{\langle X, \psi \rangle, \psi \in A_L\} \tag{21} \]

This is equivalent to finding at most \( \hat{r} = \min\{m, n\} \) atoms to maximize
\[
\arg\max_{\Psi \subset A_L} \{\langle G^t, \Psi \rangle : |\Psi| \leq \hat{r} \} \tag{22} \]

By the Eckart-Young theorem, the atoms \( \Psi \) are obtained from the SVD of \( G^t \), as \( \Psi = \{u_i v_i^T\}_{i=1}^{\hat{r}} \), where \( u_i \) and \( v_i \) are the \( i \)-th principal left and right singular vectors, respectively (the singular value \( \alpha_i \) is the coefficient of the corresponding atom \( \psi_i \) and \( \alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_{\hat{r}} \)). This result ensures that the selection atoms achieve the supremum in Eq. (21) and the optimal solution will actually lie in the set \( \Psi \). Minimizing Eq. (20) and estimating the rank of the low-rank matrix, indicate that the selection atoms must compromise between minimizing the code-length \( \mathcal{L} \) and being near to \( G^t \).

We can add a new atom to the low-rank matrix \( X \) in proper order, to move in the opposite to the worst possible direction of the optimization problem (20). To address this optimization problem efficiently, we propose a weighted formulation [45] of description length minimization designed to democratically penalize the code-length of selection atoms.
\[
\arg\min_X \frac{1}{\mu^t} \sum_i v_i s_i + \frac{1}{2} \|X - G^t\|_F^2 \tag{23} \]

where \( v_i \in \{0, 1\} \) denotes the \( i \)-th element of vector \( v \) and vector \( s = \{\mathcal{L}(W \psi_1), \mathcal{L}(W \psi_2), \ldots, \mathcal{L}(W \psi_r)\} \). \( v_i = 1 \) indicates that the atom \( \psi_i \) is selected to add to the low-rank model and the atom \( \psi_i \) has a high enough contribution to decrease the term \( \|X - G^t\|_F^2 \). \( v_i = 0 \) indicates that the atom \( \psi_i \) is not selected. So the subproblem (23) has the closed-form solution by variant of shrinkage operator, i.e, \( X = F_{\Phi} (\alpha, I_{\mu^t}(s)) \), where the \( I_r(x) \) is a variant of shrinkage operator, defined as
\[
I_r(x) = \begin{cases} 1, & x - \tau \geq 0, \\ 0, & \text{otherwise}. \end{cases} \tag{24} \]

It turns out that the number of selection atoms is the rank of the truthful low-rank matrix, \( r \).

2) Rejecting the sparse outliers: The subproblem in Eq. (18) can also be formulated as follows
\[
\arg\min_E \frac{\theta^t}{\mu^t} \|E\|_1 + \frac{1}{2} \|E - (Y - X^{t+1} + \frac{1}{\mu^t} U^t)\|_F^2 \tag{25} \]

To efficiently minimizing the \( l_1 \)-norm and the proximity term in Eq. (25), the soft-thresholding (shrinkage) method is employed. We can obtain the solution of the subproblem in Eq. (25) as \( S_{\theta^t/\mu^t}[Y - X^{t+1} + \frac{1}{\mu^t} U^t] \), where \( S_r[x] = \text{sign}(x) \max(|x| - \tau, 0) \) is the soft-thresholding operator [46].

E. Discussion

Why does the proposed MDLAN recover the best approximation of the low-rank and sparse matrices even though the number of observations is limited or the observations have gross outliers? In our MDL framework, recovering the low-rank matrix \( X \) by solving Eq. (17) or Eq. (20) aims to find the smallest set of atoms in \( A_L \) that can span \( X \), so it is equivalent to
\[
\text{atoms}(X) = \min_{\Psi} \{|\Psi| : \Psi \subset A_L, X \in \text{span}(\Psi)\} \tag{26} \]

and recovering the sparse matrix by solving Eq. (18) or Eq. (25) is equivalent to
\[
\text{atoms}(E) = \min_{\phi} \{|\Phi| : \Phi \subset A_S, E \in \text{span}(\Phi)\} \tag{27} \]

where we note that \( \text{rank}(X) = \text{atoms}(X) \) and \( \|E\|_0 = \text{atoms}(E) \). Thus, this theory (Eq. (26) and Eq. (27)) can
Algorithm 1 ADMM for the MDLAN method

**Input:** Observation \( Y \in \mathbb{R}^{m \times n} \), \( \hat{r} = \min \{m, n\} \).

1. Initial \( X^1 = \theta(m, n), E^1 = \theta(m, n), t \leftarrow 1, \mu^1 > 0, \rho > 1, \theta^1 = 1 \).

2. **repeat**

   3. //Lines 4 - 7 solve Eq. (20).
   4. \( G^t = Y - E^t + \frac{1}{\mu^t} U^t \).
   5. \( \Psi, \alpha \leftarrow \max \{\langle \hat{G}^t, \Psi \rangle : \|\Psi\| \leq \hat{r}\} \).
   6. \( s^t = \langle L(W\psi_1), L(W\psi_2), \ldots, L(W\psi_{\hat{r}}) \rangle \).
   7. \( X^{t+1} \leftarrow F_{\Psi}(\alpha, I_{\hat{r}}(s^t)) \).
   8. //Line 9 solves Eq. (25).
   9. \( E^{t+1} \leftarrow S_{\mu^t}[Y - X^{t+1} + \frac{1}{\mu^t} U^t] \).
   10. \( U^{t+1} \leftarrow U^t + \mu(Y - X^{t+1} - E^{t+1}) \).
   11. \( \mu^{t+1} \leftarrow \rho\mu^t \).
   12. \( \theta^{t+1} \leftarrow \text{the mean of } |E^{t+1}| \).
   13. \( t \leftarrow t + 1 \).

3. **until** converged.

**Output:** optimal \( X^t, E^t \)

ensure that the proposed algorithm recovers the low-rank matrix accurately and rejects the outliers.

As shown in Algorithm 1, the proposed MDLAN can find the candidate atoms for the truthful low-rank matrix and sparse outliers, respectively, and then decides which atom to add to the model according to the MDL principle. Estimating the rank of the truthful low-rank matrix \( X \) correctly is the key to recovering the low-rank matrix accurately and it also contributes to rejecting all the outliers. Similarly, rejecting all the outliers will contribute to the search for the best approximation of truthful low-rank matrix \( X \).

V. EXPERIMENTS

We evaluate the proposed method using both synthetic data sets and real sensing application examples to verify its effectiveness and robustness. In all the experiments, we use the default parameters for the methods compared.

A. Experiments with synthetic data

To compare the proposed method (MDLAN) with state-of-the-art methods on synthetic data, we synthesize a ground-truth low-rank matrix \( X_0 \in \mathbb{R}^{m \times n} \) of rank-\( r \) and a sparse matrix \( E_0 \in \mathbb{R}^{m \times n} \) with \( k \) nonzero entries, that simulates the affected data due to sensor malfunction. The low-rank matrix is a linear combination of \( r \) arbitrary orthogonal basis vectors, and the weights used to span the vector are sampled randomly from the uniform distribution \( U(0, 5) \). The \( k \) entries from \( X_0 \) are corrupted by random noise from \( N(0, 1) \). We refer to \( \frac{\|X_0 - X\|_F}{\|X_0\|_F} \) as the normalized root mean squared error (NRMSE).

1) Comparison of the success ratio: We use the recoverability results to verify the robustness of RPCA and our method (MDLAN) with respect to the number of samples (\( n \)), synthetic data dimension (\( m \)), and corruption ratio (\( p \)). For each pair, \( (n, p) \) and \( (m, p) \), we run 50 trials and report the average NRMSE of the trials. If the recovered low-rank matrix \( X \) has an NRMSE value smaller than \( \varepsilon (\varepsilon = 0.01) \), we consider that recovery is successful. The magnitudes of the colors in Fig. 3 and Fig. 4 indicate the success probability. The larger red areas indicate the more robust performance of the algorithm.

Fig. 3 shows the success ratio using RPCA and the proposed method with rank 2, 4, 6, and 8. We fix \( m = 4900 \) and vary \( n \) and \( p \). When the number of observations is deficient or the corruption ratio is large, the proposed method can obtain a competitive results. Both methods exhibit similar behavior when more samples are available or the corruption ratio is small. We also perform experiments where we fix \( n = 15 \) and vary \( m \) and \( p \). As shown in Fig. 4, the proposed method yields more robust results than RPCA for the rank 1 and 3 cases. Fig. 4 shows that the dimension (\( m \)) do not have a particularly significant effect on the results. However, the number of observations and corruption ratio severely affect the final recovery results.

2) Comparisons with other low-rank matrix approximations: We also perform experimental comparisons of a rank minimum based method (RPCA) [12], MDL principle based method (LR-MDL) [15], atomic norm based method (CoGEnT) [23], generalized fused lasso foreground modeling (BSGFL) [28] and partial sum of singular values based method (PSSV) [2]. We verify the robustness of RPCA, LR-MDL, CoGEnT, BSGFL, PSSV, and the proposed method (MDLAN) with respect to the corruption ratio. We fix \( m = 108, n = 100, r = 4 \) and vary the corruption ratio \( p \in [0.01, 0.8] \). To show more detail obtained by RPCA, LR-MDL, CoGEnT, PSSV and MDLAN in Fig. 5, the results of BSGFL is not shown (Due to considering the spatial neighborhood information in sparse matrix, the BSGFL fails to recover the synthetic sparse matrix).

Fig. 5(a) shows the NRMSE of low-rank matrix for each method as a function of the corruption ratio based on the synthetic data averaged over 50 random runs. As shown in Fig. 5(a), when the outlier ratio is lower than 0.3, the proposed method obtains similar results to PSSV and RPCA, which are better than those produced by the other methods (LR-MDL and CoGEnT). When the outlier ratio is more than 0.3, MDLAN achieves much higher accuracy than RPCA, LR-MDL, CoGEnT, and PSSV. It is clear that gross outliers exist and thus the existing methods do not capture all the energy of the underlying structure. The results shown in Fig. 5(c) demonstrate that only the proposed method estimates the rank of the underlying structure correctly (rank-4). As stated in the previous section, the proposed method finds all the candidate atoms of low-rank matrix via the atomic norm and then selects the most appropriate atoms via the MDL principle. Estimating the rank of the underlying structure correctly is crucial for recovering the low-rank matrix accurately and also benefits the outliers estimation.

The NRMSE of sparse matrix obtained by our method in Fig. 5(b) have smaller errors than those produced by RPCA, LR-MDL, CoGEnT, and PSSV when the outlier ratio is more than 0.3. The proposed method can search for the best approximation of the sparse structure via the MDL, so it can obtain more accurate results. Moreover, compared with the
other methods, the proposed approach estimates the number of nonzero entries in the sparse matrix more accurately, even when the corruption ratio is up to 0.55, as shown in Fig. 5(d) (the number of nonzero entries recovered by LR-MDL is always mn). When the corruption ratio is more than 0.55, the number of nonzero entries estimated by the proposed method is still close to the original number. To reject the outliers completely, it is necessary to recover the locations and the corresponding values of the nonzero entries accurately, which we achieved by solving Eq. (25) in our MDL framework.

Table I shows the recovery results averaged over 50 random runs, where we fix the corruption ratio to $p = 0.05$ or 0.5. When the data are corrupted with 50% outliers, the average NRMSE for the low-rank matrix using the proposed method is 0.01 and the average NRMSE for the sparse noise matrix is 0.019. In addition, MDLAN preforms better than LR-MDL, CoGEnT and BSGFL when the corruption ratio is only 0.05. In summary, the experimental results on synthetic data suggest that MDLAN performs better at recovering the low-rank matrix and rejecting the outliers from the corrupted data compared with the other state-of-the-art methods.

B. Real-world sensing applications

1) High dynamic range (HDR) imaging: Low dynamic range (LDR) images of a scene are usually captured by a sensor with different bracketing exposures. We formulate the HDR image generation problem as a rank minimization problem, where the moving objects, noise, and other nonlinear artifacts are considered as sparse outliers, and our goal is to merge several LDR images into the final HDR images. We know that LDR images are linearly dependent due to the continuous camera response. Thus, we construct three observed intensity matrices $Y \in \mathbb{R}^{m \times n} = [\text{vec}(I_1) , \cdots , \text{vec}(I_n)]$ by stacking the vectorized input images (processing each color channel respectively), where $m$ and $n$ represent the number of pixels and images, respectively, and $I_i$ denotes the input image. We apply the rank minimization methods to the three corrupted matrices to separate the outliers and the background scene (low-rank term).
Fig. 5: (a) Average NRMSE for the low-rank matrix, (b) Average NRMSE for the sparse matrix, (c) the rank (r) of recovery for the low-rank matrix, (d) Average number of nonzero entries of the recovered sparse matrix.

TABLE I: Quantitative comparison of NRMSE for the low-rank and sparse noise matrices

| Ratio  | 0.05 Low-rank | 0.05 Sparse | 0.5 low-rank | 0.5 sparse |
|--------|---------------|-------------|--------------|------------|
|        | RPCA          | LR-MDL      | CoGEnT       | BSGFL      | PSSV       | MDLAN      |
| 0.05 Low-rank | 7.45e-06      | 0.006±0.024 | 2.38±0.004  | 2.31±0.004 | 0.02±0.004 | 1.96e-06   |
| 0.05 Sparse | 1.25±0.005    | 0.200±0.024 | 1.218±0.004 | 0.275±0.063| 1.72±0.05  | 2.64±0.06  |
| 0.5 low-rank| 0.25±0.005    | 0.379±0.034 | 1.604±0.585 | 0.351±0.026| 0.207±0.002| 0.01±0.007 |
| 0.5 sparse | 0.591±0.014   | 0.490±0.024 | 1.541±0.144 | 10.04±1.423| 0.486±0.005| 0.019±0.013|

We apply the proposed approach to the three observed matrices $Y \in \mathbb{R}^{699392 \times 4}$ using a set of LDR images comprising four pictures taken in a forest [18]. The images contain artifacts caused by a person walking in the scene. Moreover, the wind makes the branches move and thus there are shadows due to the wind blowing. The final HDR results are shown in Fig. 6. Compared with the results obtained by RPCA, LR-MDL, CoGEnT, BSGFL, and PSSV, the proposed method can recover the low-rank component (artifact-free in Fig. 6(g)) and reject more outliers, even with only four input images ($n = 4$). The detailed comparison in Fig. 7 shows that our method can reject the outliers, such as ghosting and shadows, which are caused by person and the wind respectively.

2) Background modeling based on video sensor: We adopt the $F$-measure as the quantitative metric for performance evaluation of the background modeling. The $F$-measure which combines precision and recall, is calculated as follows

$$F$-measure = \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$(28)

where $\text{precision} = \frac{TP}{TP+FP}$ and $\text{recall} = \frac{TP}{TP+FN}$. TP, FP, TN, and FN, denote the numbers of true positives, false positives, true negatives, and false negatives, respectively. The higher the $F$-measure, the more accurate the outliers (foreground objects) are detected [47].

In background modeling, it is difficult to determine the correlations between video frames, as well as modeling background variations and the foreground activity. It is reasonable to assume that these background variations are low-rank, while the objects moving in the foreground are large in magnitude.
and sparse in the spatial domain. Background estimation is complex due to the presence of foreground activity such as moving people and variations in illumination.

We first consider the example video introduced by Li et al. [48], which comprises a sequence of 1186 grayscale frames obtained from a busy shopping center. Multiple people move in the scene, so the shadows on the ground surface vary significantly in the image sequences. To verify the effectiveness of the proposed method when the number of observations is limited, we only utilize a small number of continuous frames ($n=100$). Each frame has resolution of $256 \times 320$ and we stack the frames as the columns in our observed matrix $Y \in \mathbb{R}^{245760 \times 100}$.

The results are displayed in Fig. 8, which show that all the methods successfully detect the moving people. However, many shadows are present in the low-rank background recovered by RPCA, LR-MDL, CoGEnT, BSGFL and PSSV, as shown in Fig. 8(b)-(f). By contrast, our proposed method correctly model the background scene and gives better foreground with fewer false detections.

We then consider two sequences from the SABS data set, including a "Basic" sequence and a "Clutter" sequence. The "Clutter" category of sequences contain a large number of foreground moving objects occluding a large portion of the background, which is very challenging and we also only utilize 100 continuous frames. The results of all models on an example frame are indicated in Fig. 9 and Fig. 10. As shown in Fig. 9, the proposed method obtains a cleaner background (no ghosting) and detects more outliers compared against the other models, when the corruption ratio is high. Fig. 10 demonstrates that the proposed MDLAN can recover the low-rank background (no shadow) and almost cuts a foreground correctly, compared against other models.

The average F-measures and running time (on a 3 GHz Core(TM) i7 CPU) of all the models on the three sequences are shown in Table II. As illustrated in Figs. 8-10, the shadow is included in the sparse component, which makes the value of F-measure relatively low. Table II indicates that...
the proposed method can achieve the highest F-measure on the three sequences and has shown to be computationally efficient.

3) Removing shadows and specularities from faces: Basri et al. [49] stated that the face recognition problem in computer vision is a low-dimensional linear model and showed that under certain idealized circumstances, images captured by a sensor which is under variable illumination, lie near an approximately nine-dimensional linear subspace known as the harmonic plane. However, due to the presence of shadows and specularities, real face images often violate the aforementioned low-rank model. It is reasonable to consider that outliers such as shadows, specularities, and saturations are sparse in the spatial domain. Thus, we aimed to recover a low-rank model from the corrupted face images. The images have a resolution of 96 × 84 and we stack 20 face images as the columns in our observed matrix $Y \in \mathbb{R}^{8064 \times 20}$.

Fig. 11(a) shows three images from the Extended Yale B database [50], and Fig. 11(b)-(g) and Fig. 12(a)-(f) show the recovered low-rank and corresponding sparse components, respectively. The sparse terms $E$ obtained by CoGEnT and BSGFL are all zeros. Thus, the CoGEnT and BSGFL methods fail to reject the outliers $E$. Unlike the other methods, when the shaded area is small, MDLAN removes the shadows around the nose region (see the first and second rows in Fig. 11(g)). When the shaded area is large, the proposed method still removes more shadows than RPCA, LR-MDL, CoGEnT, BSGFL and PSSV (see the third row in Fig. 11(g)). Thus, our technique may be useful for pre-processing training images in face recognition systems by removing such outliers.

VI. CONCLUSION

In this study, we introduce the MDL principle and atomic norm into the field of low-rank matrix recovery, and we propose a novel nonparametric low-rank matrix approximation method called MDLAN. The existing algorithms have difficulty tackling the proposed optimization problem, so we consider an approximation of the original problem. Our method selects the best atoms to search for the best approximation of low-rank matrix, and it also can find the sparse noise simultaneously. We compare the proposed approach with state-of-the-art methods using synthetic data and three real sensing low-rank applications, i.e., HDR imaging, background modeling based on video sensor, and removing shadows and specularities from face images. The experimental results using the synthetic and real sensing datasets demonstrate the effectiveness and robustness of the proposed approach.

REFERENCES

[1] Q. Zhao, D. Meng, Z. Xu, W. Zuo, and L. Zhang, “Robust principal component analysis with complex noise,” in The International Conference on Machine Learning, 2014, pp. 55–63.

[2] T. H. Oh, Y. W. Tai, J. C. Bazin, H. Kim, and I. S. Kweon, “Partial sum minimization of singular values in robust pca: Algorithm and applications.” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 4, pp. 744–758, 2016.

[3] G. Xia, H. Sun, B. Chen, Q. Liu, and R. Hang, “Nonlinear low-rank matrix completion for human motion recovery,” IEEE Transactions on Image Processing, vol. 27, no. 99, pp. 3011–3024, 2018.

[4] L. Zhuang, S. Gao, J. Tang, J. Wang, Z. Lin, Y. Ma, and N. Yu, “Constructing a nonnegative low-rank and sparse graph with data-adaptive features,” IEEE Transactions on Image Processing, vol. 24, no. 11, pp. 3717–3728, 2015.

[5] Hengmin, Zhang, Chen, Gong, Jianjun, Qian, Bob, Chunyan, Xu, and J. and, “Efficient recovery of low-rank matrix via double nonconvex nonsmooth rank minimization,” IEEE Transactions on Neural Networks and Learning Systems, 2019.

[6] H. Zhang, J. Qian, B. Zhang, J. Yang, C. Gong, and Y. Wei, “Low-rank matrix recovery via modified schatten- p norm minimization with convergence guarantees,” IEEE Transactions on Image Processing, vol. 29, pp. 3132–3142, 2020.

[7] I. Chen, J. Zhou, and J. Ye, “Integrating low-rank and group-sparsity structures for robust multi-task learning,” in ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2011, pp. 42–50.

[8] W. Chen, “Simultaneously sparse and low-rank matrix reconstruction via nonconvex and nonseparable regularization,” IEEE Transactions on Signal Processing, vol. 66, no. 99, pp. 5313–5323, 2018.

[9] X. Xie, J. Wu, G. Liu, and J. Wang, “Matrix recovery with implicitly low-rank data,” Neurocomputing, vol. 334, no. MAR.21, pp. 219–226, 2019.

[10] I. T. Jolliffe, “Principal component analysis,” Springer, vol. 14, no. 27, pp. 231–246, 1986.

[11] A. Ganesh, Y. Ma, S. Rao, and J. Wright, “Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization,” Advances in Neural Information Processing Systems, vol. 87, no. 4, pp. 20:3–20:56, 2009.

[12] Candes, E. J. S., X. Li, Y. Ma, and J. Wright, “Robust principal component analysis,” Journal of the Acm, vol. 58, no. 3, pp. 1–7, 2011.

[13] M. Chen, Z. Lin, Y. Ma, and L. Wu, “The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices,” Evrim Arxiv, vol. 9, 2009.

[14] J. Wright, “Low-rank matrix recovery and completion via convex optimization,” http://perception.csi.illinois.edu/matrix-rank/home. html, 2012.

[15] I. Ramírez and G. Sapiro, “Low-rank data modeling via the minimum description length principle,” in IEEE International Conference on Acoustics, Speech, and Signal Processing, 2011, pp. 2165–2168.

[16] T. Bouwmans and E. H. Zahzah, “Robust pca via principal component pursuit: A review for a comparative evaluation in video surveillance,” Computer Vision and Image Understanding, vol. 122, no. 4, pp. 22–34, 2014.

[17] X. Liu, G. Zhao, J. Yao, and C. Qi, “Background subtraction based on low-rank and structured sparse decomposition,” IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, vol. 24, no. 8, pp. 2502–2515, 2015.

[18] O. Gallo, “Artifact-free high dynamic range imaging,” in Proc. IEEE Int. Conf. Comput. Photography, 2009, pp. 1–7.

[19] A. T. Celebi, R. Duvar, and O. Urhan, “Fuzzy fusion based high dynamic range imaging using adaptive histogram separation,” IEEE Transactions on Consumer Electronics, vol. 61, no. 1, pp. 119–127, 2015.

[20] I. Ramírez and G. Sapiro, “An mdl framework for sparse coding and dictionary learning.” IEEE Transactions on Signal Processing, vol. 60, no. 6, pp. 2913–2927, 2011.

[21] J. Rissanen, “Paper: Modeling by shortest data description,” Automatica, vol. 14, no. 5, pp. 465–471, 1978.

[22] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, “The convex geometry of linear inverse problems,” Foundations of Computational Mathematics, vol. 12, no. 6, pp. 805–849, 2012.

[23] N. Rao, P. Shah, and S. Wright, “Forward-backward greedy algorithms for atomic norm regularization,” Signal Processing IEEE Transactions on, vol. 63, no. 21, pp. 5798–5811, 2014.

[24] Wang, Yulong, Tang, Y. Yan, and Luoqing, “Minimum error entropy based sparse representation for robust subspace clustering,” IEEE Transactions on Signal Processing: A publication of the IEEE Signal Processing Society, 2015.

[25] A. Qin, Z. Shang, T. Zhang, Y. Ding, and Y. Y. Tang, “Minimum description length principle based atomic norm for synthetic low-rank matrix recovery,” in 2016 7th International Conference on Cloud Computing and Big Data (CCBD), 2016, pp. 7–12.

[26] X. Zhou, C. Yang, and W. Yu, “Moving object detection by detecting contiguous outliers in the low-rank representation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 3, pp. 597–610, 2013.

[27] J. Xu, V. K. Ithapu, L. Mukherjee, J. M. Rehg, and V. Singh, “Gossus: Grassmannian online subspace updates with structured-sparcity,” in
Fig. 8: Background modeling results on Li data set. (a) One frame of the original video $Y$ (top) and the ground truth (bottom). Low-rank component $X$ (top) and sparse component $E$ (bottom) obtained by RPCA (b), LR-MDL (c), CoGEnT (d), BSGFL (e) PSSV (f), and the proposed approach (g).

Fig. 9: Background modeling results on SABS data set. (a) One frames of the original video $Y$ (top) and the ground truth (bottom). Low-rank component $X$ (top) and sparse component $E$ (bottom) obtained by RPCA (b), LR-MDL (c), CoGEnT (d), BSGFL (e) PSSV (f), and the proposed approach (g).

Fig. 10: Background modeling results on SABS data set. (a) One frames of the original video $Y$ (top) and the ground truth (bottom). Low-rank component $X$ (top) and sparse component $E$ (bottom) obtained by RPCA (b), LR-MDL (c), CoGEnT (d), BSGFL (e) PSSV (f), and the proposed approach (g).

[28] B. Xin, Y. Tian, Y. Wang, and W. Gao, “Background subtraction via generalized fused lasso foreground modeling,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 4676–4684.

[29] B. Xin, Y. Kawahara, Y. Wang, and W. Gao, “Efficient generalized fused lasso and its application to the diagnosis of alzheimer’s disease,” in Twenty-Eighth AAAI Conference on Artificial Intelligence, 2014, pp. 2163–2169.

[30] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” in Eleventh International Conference of the Center for Nonlinear Studies on Experimental Mathematics : Computational Issues in Nonlinear Science: Computational Issues in Nonlinear Science, 1992, pp. 259–268.

[31] S. Erfanian, Ebadi, Ebroul, and Izquierdo, “Foreground segmentation with tree-structured sparse rpca.” IEEE transactions on pattern analysis and machine intelligence, vol. 40, no. 9, pp. 2273–2280, 2018.

[32] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Sisstrunk, “Slic superpixels compared to state-of-the-art superpixel methods,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 11, pp. 2274–2282, 2012.

[33] S. Shah, T. Goldstein, and C. Studer, “Estimating sparse signals with smooth support via convex programming and block sparsity,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 5906–5915.

[34] Z. Zha, X. Liu, X. Huang, H. Shi, Y. Xu, Q. Wang, L. Tang, and X. Zhang, “Analyzing the group sparsity based on the rank minimization methods,” in Multimedia and Expo (ICME), 2017 IEEE International Conference on. IEEE, 2017, pp. 883–888.

[35] R. Cabral, F. D. L. Torre, J. P. Costeira, and A. Bernardino, “Unifying nuclear norm and bilinear factorization approaches for low-rank matrix decomposition,” in IEEE International Conference on Computer Vision, 2013, pp. 2488–2495.

[36] X. Guo, Z. Lin, and C. M. I. Center, “Route: Robust outlier estimation for low rank matrix recovery,” in Proceedings of the Twenty-Sixth
TABLE II: Quantitative evaluation of the background modeling, given as the F-measure and running time.

| Data sets | RPCA | LR-MDL | CoGEnT | BSGFL | PSSV | MSLAN |
|-----------|------|--------|--------|-------|------|-------|
| Shopping Mall | F-measure | 0.6975 | 0.5021 | 0.6846 | 0.6406 | 0.7060 | **0.7536** |
| | Times(sec) | 585 | 912 | 1534 | 11628 | 20.1 | 28.6 |
| HumanBody2 | F-measure | 0.6881 | 0.6294 | 0.0676 | 0.6806 | 0.7377 | **0.7941** |
| | Times(sec) | 985 | 106 | 586 | 2219 | 37.1 | 40.7 |
| MPEG | F-measure | 0.7437 | 0.7977 | 0.7735 | 0.7987 | 0.7756 | **0.8264** |
| | Times(sec) | 4087 | 446 | 6498 | 14942 | 207 | 227 |

Fig. 11: Removing shadows and specularities from face images. (a) Cropped and aligned images from the Extended Yale B database of a person’s face under different illumination. The size of each image is 96×84 pixels and 20 different illumination settings in total were used for the person. Low-rank term $X$ obtained by RPCA (b), LR-MDL (c), CoGEnT (d), BSGFL (e) PSSV (f), and the proposed approach (g).

Fig. 12: Sparse component $E$ obtained by RPCA (a), LR-MDL (b), CoGEnT (c), BSGFL (d) PSSV (e), and the proposed approach (f).

[37] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He, “Fast and accurate matrix completion via truncated nuclear norm regularization,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 9, pp. 2117–2130, 2013.
[38] X. Guo, R. Zhao, G. An, and Y. Cen, “An algorithm of face alignment and recognition by sparse and low rank decomposition,” in *2014 12th International Conference on Signal Processing (ICSP)*, Hangzhou, 2014, pp. 1036–1040.
[39] T. H. Oh, J. Y. Lee, Y. W. Tai, and I. S. Kweon, “Robust high dynamic range imaging by rank minimization,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 6, pp. 1219–32, 2015.
[40] Y. Peng, J. Suo, Q. Dai, and I. S. Xu, Wenli, “Reweighted low-rank matrix recovery and its application in image restoration,” *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2418–2430, 2014.