Article

Low-Rank Matrix Recovery from Noise via an MDL Framework-Based Atomic Norm

Anyong Qin [1,*] **, Lina Xian** [2] **, Yongliang Yang** [2] **, Taiping Zhang** [3] **and Yuan Yan Tang** [4]

[1] School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

[2] School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; 2017211059@stu.cqupt.edu.cn (L.X.); 2018211426@stu.cqupt.edu.cn (Y.Y.)

[3] College of Computer Science, Chongqing University, Chongqing 400030, China; tpzhang@cqu.edu.cn

[4] Faculty of Science and Technology, University of Macau, Macau 999078, China; yytang@umac.mo

* Correspondence: qinay@cqupt.edu.cn

check for updates

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 and 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 in the case of 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 terms, and then we minimize the description length of the model in order to select the appropriate atoms of low-rank and the sparse matrices, 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 utilizing synthetic data and real sensing applications (high dynamic range imaging, background modeling, removing noise and shadows) demonstrate the effectiveness, robustness and efficiency of the proposed method.

Keywords: atomic norm; low-rank matrix recovery; minimum description length principle; robust principal components analysis

1. 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, and so the given data matrix Y can be decomposed into low-rank and sparse components.

Principal components 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 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 outliers [11,12]:

$$\underset{X,E}{\text{argmin}} \, rank(X) + \gamma\|E\|_0, \quad s.t. \, Y = X + E, \tag{1}$$

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

$$\underset{X,E}{\text{argmin}} \, \|X\|_* + \gamma\|E\|_1, \quad s.t. \, Y = X + E, \tag{2}$$

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 Equation (2) effectively [13,14].

Wright et al. [11] and Candès et al. [12] proved that the performance of Equation (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 Equation (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 γ both influence the final results of RPCA decomposition. Unfortunately, the target rank and regularizing parameter γ are uncertain in Equations (1) and (2), where conventional approaches need to tune the rank of X and γ to achieve the desired goal. However, $\gamma = 1/\sqrt{max\{m,n\}}$, which is set by the typical approaches, is not the best value [15]. The goal of model selection is to select the most appropriate model from the set of candidates; for example, selecting the appropriate parameters. The information theoretic criteria (ITC) is often used to solve the model selection problems by minimizing a penalized likelihood function via a specific criterion, such as the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). The minimum description length (MDL) principle is also motivated by an information theoretic perspective and can avoid assumptions regarding prior distribution [20,21]. Instead, Ramirez et al. [15,22] used the minimum description length (MDL) principle [23] to avoid estimating the parameter γ. The MDL principle selects the best low-rank approximation from RPCA decomposition sequences, which are obtained via different values of γ. Liu et al. [17] employed structured sparse decomposition to solve the regularizing parameter issue in RPCA, where they replaced the static parameter γ 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 Figure 1. The RPCA fails to recover the low-rank matrix and capture the illumination sudden changes.
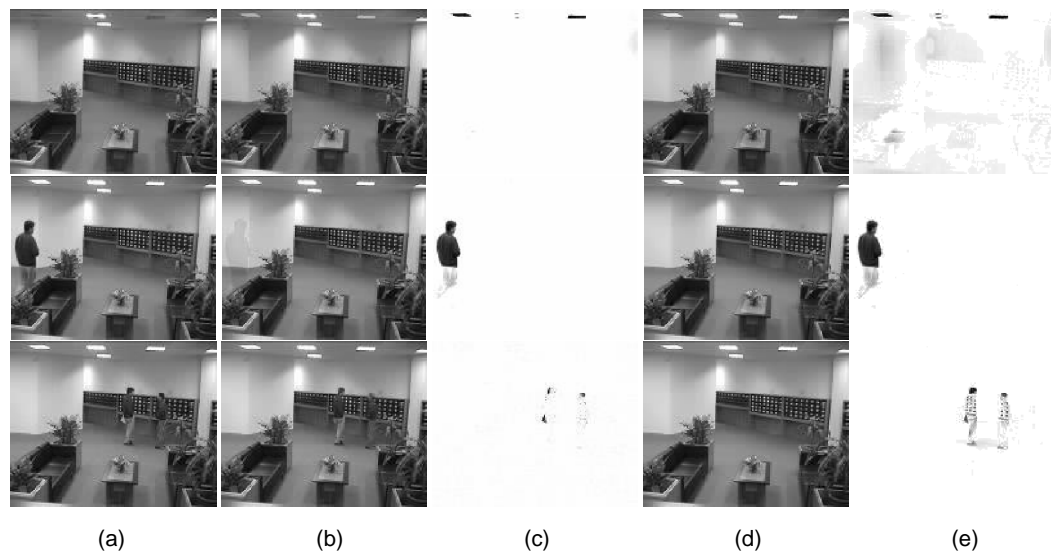
Figure 1. Recovered background and detection of outliers. Three frames from an 80-frame sequence taken in a lobby are presented. (**a**) Three frames from the original video *Y*, low-rank component *X* (**b**,**d**) and structured sparse component *E* (**c**,**e**) obtained by robust principal components analysis (RPCA) (**b**,**c**), and the proposed approach (**d**,**e**), respectively. The rank estimated by RPCA is 7, and so ghosting appeared in the background. By contrast, the rank estimated by our approach is 1.

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 [24–26]. To address issues such as the limited number of observations, the rank of *X* and the regularizing parameter γ, 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 [27]. 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 the 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 study 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 alternating direction method of multipliers (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 research works. In Section 3, we describe the proposed MDLAN method. Section 4 presents the

experimental results based on the synthetic and real datasets. Finally, we give our conclusions in Section 5.

2. Related Works

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 have also replaced the number of nonzero entries with the l_1-norm, as shown in Equation (2). Candès et al. [12] proved that the rank minimization problem can be solved using Equation (1), and it also can be solved in a tractable manner by the convex relaxation version of Equation (2). They also proved that the unique solution of Equation (2) corresponded exactly to the solution of the original NP-hard problem in Equation (1) under suitable conditions.

Recently, the improvements to RPCA have been generally divided into two categories. One category focuses on the structured sparse component E in Equation (2) [28,29]. For example, Xin et al. [30] replaced l_1-*norm* with an adaptive version of generalized fused lasso (GFL) regularization [31], which takes into account the spatial neighborhood information of the foregrounds in a video sequence.

$$\underset{X,E}{\operatorname{argmin}} \|X\|_* + \gamma \|E\|_{gfl}, \ \ s.t. \ Y = X + E, \tag{3}$$

where the generalized fused lasso $\|E\|_{gfl}$ can be viewed as a combination of two common regularizers; i.e., the l_1-*norm* and the total variation (TV) penalty [32].

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

where $e^{(l)}$ is the l-th column of the sparse matrix E, $\mathcal{N}$ is the spatial neighborhood set, λ_1 is a tuning parameter and $w_{ij} = exp(\frac{-\|y_i^{(l)} - y_j^{(l)}\|_2^2}{2\sigma^2})$ ($\sigma \geq 0$ is a tuning parameter (empirically set). Ebadi et al. [33] dynamically estimated the support of the sparse matrix E via a superpixel generation step [34] to impose the spatial coherence onto the structured sparse outliers. Shah et al. [35] replaced the l_1-*norm* in Equation (2) with hybrid l_1/l_2-*norm*, which can promote the spatial smoothness in the support set of the structured sparse outliers.

Another category focuses on the low-rank component X in Equation (2) [36]. For example, Cabral et al. [37] and Guo et al. [38,39] replaced the X with UV, and the relationship $\|X\|_* = \underset{U,V}{\min} \frac{1}{2}\|U\|_F^2 + \frac{1}{2}\|V\|_F^2$ 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. [38,39] employed an entropy term to restrict the support of the outliers. Hu et al. [40] proposed an approximation of the target rank by the truncated nuclear norm, which only minimized the smallest $min(m, n) - r$ singular values. T-H Oh et al. [2] proposed the minimization of 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:

$$\underset{X,E}{\operatorname{argmin}} |rank(X) - r| + \gamma \|E\|_1, \ \ s.t. \ Y = X + E \tag{4}$$

The rank minimization algorithms for RPCA have inspired many applications in image processing and computer vision, such as image alignment [41], background subtraction [12,17], high dynamic range (HDR) imaging [2,42] and image restoration [43,44]. However, the clean data are easily corrupted by gross noise/outliers, or the amount of given data can be limited by factors related to the sensor or human error [2,11,45]. The available methods based on RPCA have difficulty solving these problems. In the present study, we propose an algorithm based on MDL and the atomic norm to overcome these difficulties; i.e., an unknown target rank r, the regularizing parameter λ and deficient observations or gross outliers.

3. An MDL Principle-Based Atomic Norm for Low-Rank Matrix Recovery

In this section, we will present the concept of the atomic norm and the MDL principle, respectively. We also provide the unified form of the low-rank model, which is based on the atomic norm. We then propose the new low-rank matrix recovery method (MDLAN) based on the MDL principle and atomic norm, as well as the optimization algorithm.

3.1. Atomic Norm

First, we provide a definition of an atomic norm and some assumptions regarding the set of atoms ($\mathcal{A}$). We also assume that the set $\mathcal{A}$ is origin-symmetric (i.e., $A \in \mathcal{A}$ if and only if $-A \in \mathcal{A}$). The atomic norm [24] is the gauge function induced by $\mathcal{A}$:

$$\|X\|_{\mathcal{A}} := \inf_{t>0}\{t : X \in t \cdot conv(\mathcal{A})\}$$

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

$$\|X\|_{\mathcal{A}}^* := \sup\{\langle X, A\rangle, a \in \mathcal{A}\}$$

where the inner product is defined as $\langle X, A\rangle = tr(X^T A)$ for the matrix and $tr(\cdot)$ 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

$$\mathcal{A}_S := \{\pm E_{ij} \in \mathbb{R}^{m \times n}, i = 1, 2, \cdots, m, j = 1, 2, \cdots, n\}$$

where E_{ij} denotes a matrix, where the *(i,j)-th* entry of the matrix 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

$$\mathcal{A}_L := \{Z \in \mathbb{R}^{m \times n} | rank(Z) = 1, \|Z\|_F = 1\}$$

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

3.2. Atomic Norm-Based Low-Rank Matrix Recovery

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

$$\underset{X,E}{\operatorname{argmin}} \|X\|_{\mathcal{A}_L} + \lambda \|E\|_{\mathcal{A}_S} \ \ s.t. \ Y = X + E \tag{5}$$

where λ 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 $\mathcal{F}_\Psi : \mathbb{R}^{|\Psi|} \to \mathbb{R}^{m \times n}$ by*

$$\mathcal{F}_\Psi \boldsymbol{\alpha} = \sum_{k=1}^{|\Psi|} \alpha_k \psi_k \quad \forall \boldsymbol{\alpha} \in \mathbb{R}^{|\Psi|} \tag{6}$$

From Equation (6), it follows that the adjoint operator $\mathcal{F}_\Psi^* : \mathbb{R}^{m \times n} \to \mathbb{R}^{|\Psi|}$ is given by

$$\mathcal{F}_\Psi^* X = [\langle X, \psi_1\rangle, \ldots, \langle X, \psi_{|\Psi|}\rangle] \quad \forall X \in \mathbb{R}^{m \times n} \tag{7}$$

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

$$\|X\|_{\mathcal{A}_L} := \inf\{\sum_{i=1}^{|\Psi|} \alpha_i : X = \mathcal{F}_\Psi \boldsymbol{\alpha}, \ \alpha_i \geq 0, \Psi \subset \mathcal{A}_L\} \tag{8}$$

$$\|E\|_{\mathcal{A}_S} := \inf\{\sum_{i=1}^{|\Phi|} \beta_i : E = \mathcal{F}_\Phi \boldsymbol{\beta}, \ \beta_i \geq 0, \Phi \subset \mathcal{A}_S\} \tag{9}$$

where $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ are the vectors of the scalar coefficients, and $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \cdots, \alpha_i, \cdots\}$, $\boldsymbol{\beta} = \{\beta_1, \beta_2, \cdots, \beta_i, \cdots\}$.

3.3. 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:

$$\hat{M} = \underset{M \in \mathcal{M}}{\operatorname{argmin}} \mathcal{L}(Y, M) \tag{10}$$

where the codelength assignment function $\mathcal{L}(Y, M)$ defines the theoretical codelength required to describe (Y, M) uniquely. A common implementation of the MDL framework uses the Ideal Shannon Codelength Assignment ([46], Ch.5) to define $\mathcal{L}(Y, M)$ in terms of a probability assignment $P(Y, M)$ as $\mathcal{L}(Y, M) = -logP(Y, M) = -log(P(Y|M)P(M))$. Thus, we obtain the MDL framework

$$\hat{M} = \underset{M \in \mathcal{M}}{\operatorname{argmin}} -logP(M) - logP(Y|M) \tag{11}$$

where $-logP(M)$ represents the model complexity and $-logP(Y|M)$ represents the measure of the goodness of fit.

3.4. The Proposed Method

Our family of models for expressing the low-rank matrix recovery problems are defined by $\mathcal{M} = \{(X, E) : Y \leftarrow X + E, rank(X) = r, \|E\|_0 = k\}$, 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:

$$\begin{aligned}(\hat{X}, \hat{E}) &= \underset{M \in \mathcal{M}}{\operatorname{argmin}} \mathcal{L}(Y, M) \\ &= \underset{X, E}{\operatorname{argmin}} \mathcal{L}(X) + \mathcal{L}(E) + \mathcal{L}(Y - X - E)\end{aligned} \tag{12}$$

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

$$\underset{\Psi, \Phi}{\operatorname{argmin}} \sum_i \mathcal{L}(\alpha_i \psi_i) + \mathcal{L}(\mathcal{F}_\Phi \boldsymbol{\beta}) + \mathcal{L}(Y - \mathcal{F}_\Psi \boldsymbol{\alpha} - \mathcal{F}_\Phi \boldsymbol{\beta}) \tag{13}$$

The basic idea of the proposed MDLAN is to find two smallest sets Ψ and Φ, while minimizing the $\mathcal{L}(Y - \mathcal{F}_\Psi \boldsymbol{\alpha} - \mathcal{F}_\Phi \boldsymbol{\beta})$. The cost function in Equation (13) is non-convex in (Ψ, Φ), and we relax it with an alternative objective function in order to effectively handle the proposed problem. The codelength of low-rank matrix X can be written as $\sum_i \mathcal{L}(W\psi_i)$, where $W \in \mathbb{R}^{m \times m}$ is in a lower triangular form. Minimizing the description length of sparse matrix ($\mathcal{L}(\mathcal{F}_\Phi \boldsymbol{\beta})$) is replaced by minimizing $\theta \|E\|_1$, where $\theta = \frac{1}{k}\sum_{i=1}^{k} |\beta_i|$. The encoding schemes of the low-rank matrix and sparse matrix are given in Appendix A.

3.5. Optimization by ADMM

As with other research works [1,2,45], in this work, the equation $Y = X + E$ still holds. Then, the original problem (Equation (13)) can be reformed as follows:

$$(\hat{X}, \hat{E}) = \underset{X,E}{\arg\min} \sum_i \mathcal{L}(W\psi_i) + \theta \|E\|_1$$
$$s.t. \ Y = X + E$$

(14)

Here, we employ the alternating direction method of multipliers (ADMM) method [13,47] to solve this constrained optimization problem. The augmented Lagrangian function of Equation (14) is

$$L_\mu(X, E, U) = \sum_i \mathcal{L}(W\psi_i) + \theta \|E\|_1 + \langle U, Y - X - E \rangle$$
$$+ \frac{\mu}{2} \|Y - X - E\|_F^2$$

(15)

where μ 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:

$$X^{t+1} = \underset{X}{\arg\min} \ L_{\mu^t}(X, E^t, U^t)$$

(16)

$$E^{t+1} = \underset{E}{\arg\min} \ L_{\mu^t}(X^{t+1}, E, U^t)$$

(17)

$$U^{t+1} = U^t + \mu(Y - X^{t+1} - E^{t+1})$$

(18)

The two subproblems (Equations (16) and (17)) are convex optimization problems which are solved while fixing another variable. Algorithm 1 summarizes the whole recovery procedure of recovering the low-rank matrix and rejecting the sparse outliers alternately. The logic that underlies the proposed MDLAN method is that the codelength cost of adding a new atom to the model is usually very high, and 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 $Y - X - E = 0$.

Algorithm 1 ADMM [13,47] for the MDLAN method.

Input: Observation $Y \in \mathbb{R}^{m \times n}, \hat{r} = min\{m, n\}$.
 1: Initial $X^1 = \mathbf{0}(m, n), E^1 = \mathbf{0}(m, n), t \leftarrow 1, \mu^1 > 0, \rho > 1, \theta^1 = 1$.
 2: **repeat**
 3:     *//Lines 4 − 7 solve Equation (19).*
 4:     $G^t = Y - E^t + \frac{1}{\mu^t} U^t$.
 5:     $\Psi, \alpha \leftarrow \underset{\Psi \subset \mathcal{A}_L}{\max} \{ \langle G^t, \Psi \rangle : |\Psi| \leq \hat{r} \}$
 6:     $s^t \leftarrow (\mathcal{L}(W\psi_1), \mathcal{L}(W\psi_2), \cdots, \mathcal{L}(W\psi_{\hat{r}}))$
 7:     $X^{t+1} \leftarrow \mathcal{F}_\Psi \langle \alpha, \Gamma_{\frac{1}{\mu^t}}(s^t) \rangle$
 8:     *//Line 9 solves Equation (24).*
 9:     $E^{t+1} \leftarrow \mathcal{S}_{\frac{\theta^t}{\mu^t}} [Y - X^{t+1} + \frac{1}{\mu^t} U^t]$
10:     $U^{t+1} \leftarrow U^t + \mu(Y - X^{t+1} - E^{t+1})$
11:     $\mu^{t+1} \leftarrow \rho\mu^t$
12:     $\theta^{t+1} \leftarrow$ *the mean of* $|E^{t+1}|$
13:     $t \leftarrow t + 1$.
14: **until** converged.
Output: optimal X^t, E^t

3.5.1. Recovering the Low-Rank Matrix

The subproblem in Equation (16) can be formulated as follows:

$$\underset{X}{\operatorname{argmin}} \frac{1}{\mu^t} \sum_i \mathcal{L}(W\psi_i) + \frac{1}{2} \left\| X - G^t \right\|_F^2 \tag{19}$$

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, we first determine the candidate set Ψ by the dual atomic norm.

$$\|X\|_{\mathcal{A}_L}^* := \sup\{\langle X, \psi \rangle, \psi \in \mathcal{A}_L\} \tag{20}$$

This is equivalent to finding at most $\hat{r} = min\{m, n\}$ atoms to maximize

$$\underset{\Psi \subset \mathcal{A}_L}{\operatorname{argmax}}\{\langle G^t, \Psi \rangle : |\Psi| \le \hat{r}\} \tag{21}$$

By the Eckart–Young theorem, the atoms Ψ 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 α_i is the coefficient of the corresponding atom ψ_i and $\alpha_1 \ge \alpha_2 \ge \cdots \ge \alpha_{\hat{r}}$). This result ensures that the selection atoms achieve the supremum in Equation (20) and that the optimal solution will actually lie in the set Ψ. Minimizing Equation (19) and estimating the rank of the truthful low-rank matrix indicates that the selection atoms must compromise between minimizing the codelength $\mathcal{L}$ and being near to G^t. We can add a new atom to the low-rank matrix X in proper order to move opposite to the worst possible direction of the optimization problem (19). To address this optimization problem efficiently, we propose a weighted formulation [48] of description length minimization that is designed to democratically penalize the codelength of selected atoms.

$$\underset{X}{\operatorname{argmin}} \frac{1}{\mu^t} \sum_i v_i s_i + \frac{1}{2} \left\| X - G^t \right\|_F^2 \tag{22}$$

where $v_i \in \{0, 1\}$ denotes the i-th element of vector v and vector $s = (\mathcal{L}(W\psi_1), \mathcal{L}(W\psi_2), \cdots, \mathcal{L}(W\psi_{\hat{r}}))$. $v_i = 1$ indicates that the atom ψ_i is selected to add to the low-rank model and the atom ψ_i makes a sufficiently high contribution to decrease the term $\left\| X - G^t \right\|_F^2$. $v_i = 0$ indicates that the atom ψ_i is not selected. Thus, the subproblem (22) has a closed-form solution by varying the shrinkage operator; i.e, $X = \mathcal{F}_\Psi \langle \alpha, I_{\frac{1}{\mu^t}}(s) \rangle$. Where $I_\tau(x)$ is a variant of the shrinkage operator, defined as

$$I_\tau(x) = \begin{cases} 1, & x - \tau \ge 0, \\ 0, & \text{otherwise}, \end{cases} \tag{23}$$

it has been shown that the number of selection atoms is the rank of the truthful low-rank matrix, r.

3.5.2. Rejecting the Sparse Outliers

The subproblem in Equation (17) can also be formulated as follows:

$$\underset{E}{\operatorname{argmin}} \frac{\theta^t}{\mu^t} \|E\|_1 + \frac{1}{2} \left\| E - (Y - X^{t+1} + \frac{1}{\mu^t} U^t) \right\|_F^2 \tag{24}$$

To efficiently minimize the l_1-*norm* and the proximity term in Equation (24), the soft-thresholding (shrinkage) method is employed. We can obtain the solution of the subproblem in Equation (24) as $\mathcal{S}_{\frac{\theta^t}{\mu^t}}[Y - X^{t+1} + \frac{1}{\mu^t} U^t]$, where $\mathcal{S}_\tau[x] = sign(x)max(|x| - \tau, 0)$ is the soft-thresholding operator [49].

3.6. 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, the recovery of the low-rank matrix X by solving Equation (16) or Equation (19) is performed with the aim of finding the smallest set of atoms in $\mathcal{A}_L$ that can span X, so it is equivalent to

$$atoms(X) = \min_{\Psi}\{|\Psi| : \Psi \subset \mathcal{A}_L, X \in span(\Psi)\} \tag{25}$$

and recovering the sparse matrix by solving Equation (17) or Equation (24) is equivalent to

$$atoms(E) = \min_{\Phi}\{|\Phi| : \Phi \subset \mathcal{A}_S, E \in span(\Phi)\} \tag{26}$$

where we note that $rank(X) = atoms(X)$ and $\|E\|_0 = atoms(E)$. Thus, this theory (Equations (25) and (26)) can 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 outliers. Similarly, rejecting all outliers will contribute to the search for the best approximation of the truthful low-rank matrix X.

4. Experiments

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

4.1. 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 bad 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).

4.1.1. 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 overall average NRMSE of the trials. If the recovered low-rank matrix X has an NRMSE value smaller than ε ($\varepsilon = 0.01$), we consider that recovery is successful. The magnitudes of the colors in Figures 2 and 3 indicate the success probability. The larger red areas indicate the more robust performance of the algorithm.

Figure 2 shows the success ratio using RPCA and the proposed method with ranks 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 competitive results. Both methods exhibit similar behaviors when more samples are available or the corruption ratio is small.
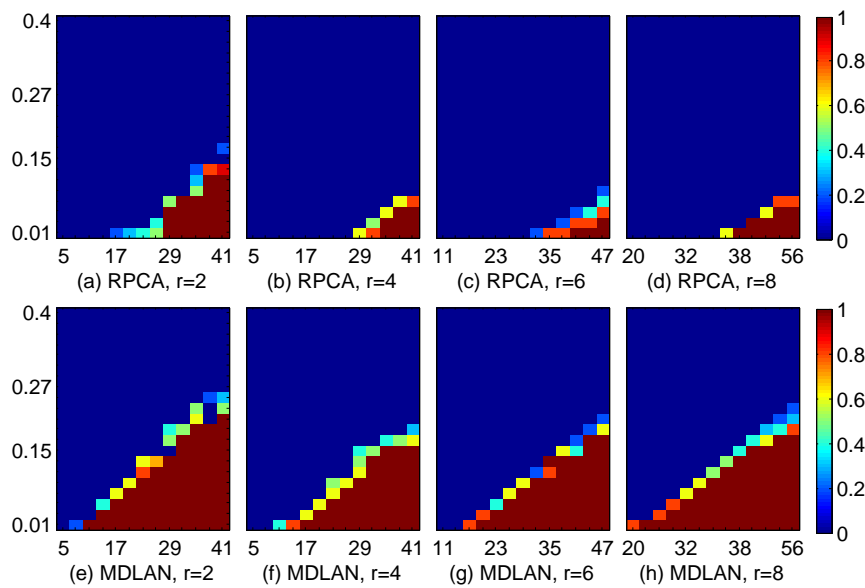
Figure 2. Recovery results with various numbers of observations (*n*). Comparison of robust principal components analysis (RPCA) and the proposed minimum description length atomic norm (MDLAN) method for rank 2, 4, 6 and 8 cases. The *X*-axis represents the number of samples (*n*) and the *Y*-axis represents the corruption ratio $p \in [0.01, 0.41]$. The color magnitude represents the success ratio [0,1].
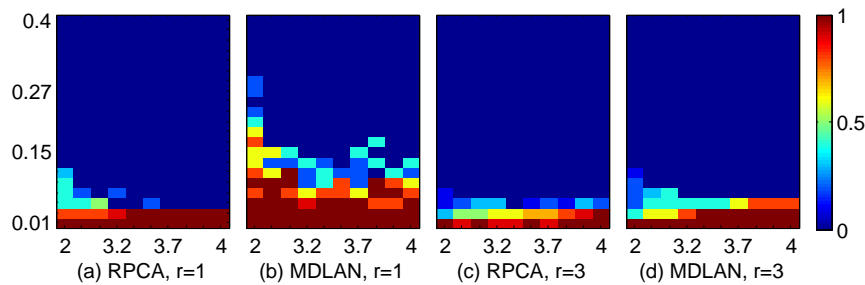

Figure 3. Recovery results with various numbers of dimensions (*m*). Comparison of RPCA and the proposed MDLAN method for rank 1 and 3 cases. The *X*-axis represents the log-scale row size ($log_{10}m \in [log_{10}100, log_{10}100,000]$) and the *Y*-axis represents the corruption ratio $p \in [0.01, 0.41]$. The color magnitude represents the success ratio [0,1].

We also performed experiments in which we fixed $n = 15$ and vary m and p. As shown in Figure 3, the proposed method yields more robust results than RPCA for the rank 1 and 3 cases. Figure 3 shows that the dimension (*m*) does not have a particularly significant effect on the results. However, the number of observations and corruption ratio severely affect the final recovery results.

4.1.2. Comparisons with Other Low-Rank Ratrix Approximations

We also perform experimental comparisons of a rank minimum-based method (RPCA) [12], MDL principle-based method (LR-MDL) [15], conditional gradient with enhancement and truncation based on atomic norm (CoGEnT) [25], generalized fused lasso foreground modeling (BSGFL) [30], partial sum of singular values-based method (PSSV) [2], low-rank matrix recovery via robust outlier estimation (ROUTE) [39], factor group-sparse regularization for low-rank matrix recovery (FGSR) [50] and low-rank matrix recovery via subgradient method (SubGM) [51]. We verify the robustness of RPCA, LR-MDL, CoGEnT, BSGFL, PSSV, ROUTE, FGSR, SubGM and the proposed method (MDLAN) with respect to the corruption ratio. We fix $m = 108$, $n = 100$ (except for SubGM, where we set $n = 108$), $r = 4$ and vary the corruption ratio $p \in [0.01, 0.8]$. To show more of the detail obtained by RPCA, LR-MDL, CoGEnT, PSSV, ROUTE, FGSR, SubGM and MDLAN in Figure 4, the results of BSGFL

are not shown (as the spatial neighborhood information is considered in a sparse matrix, the BSGFL fails to recover the synthetic sparse matrix).
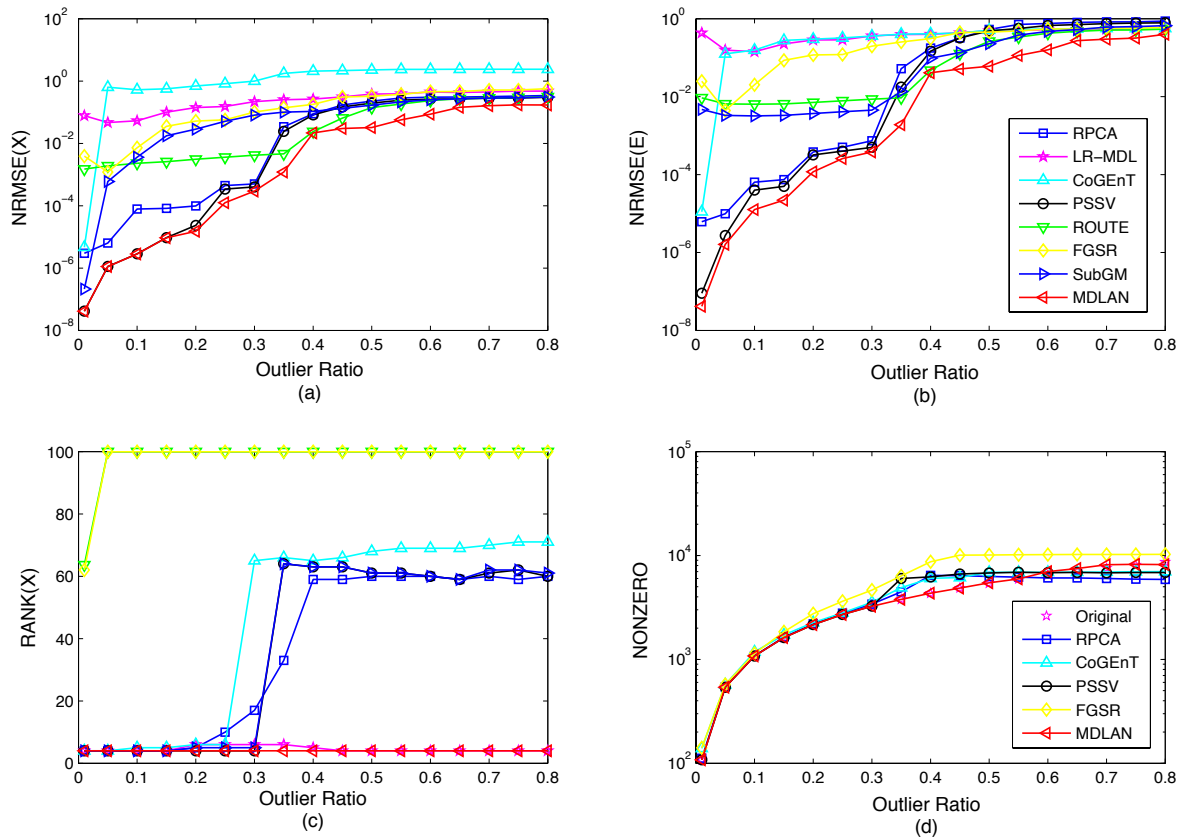

Figure 4. (**a**) Average normalized root mean square error (NRMSE) for the low-rank matrix, (**b**) average NRMSE for the sparse matrix, (**c**) the rank (*r*) of recovery for the low-rank matrix, (**d**) average number of nonzero entries of the recovered sparse matrix.

Figure 4a shows the NRMSE of the 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 Figure 4a, 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, CoGEnT, ROUTE, FGSR and SubGM). When the outlier ratio is more than 0.3, MDLAN achieves much higher accuracy than RPCA, LR-MDL, CoGEnT, PSSV, ROUTE, FGSR and SubGM. 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 Figure 4c 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 outlier estimation.

The NRMSE of the sparse matrix obtained by our method in Figure 4b has smaller errors than those produced by RPCA, LR-MDL, CoGEnT, PSSV, ROUTE, FGSR and SubGM 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 Figure 4d (the number of nonzero entries recovered by LR-MDL, ROUTE and SubGM are 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 Equation (24) in our MDL framework.

Table 1 shows the recovery results averaged over 50 random runs, where the corruption ratio is fixed 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, BSGFL, ROUTE, FGSR and SubGM when the corruption ratio is only 0.05. In summary, the experimental results for 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.

Table 1. Quantitative comparison of NRMSE for the low-rank and sparse noise matrices. LR-MDL: low-rank minimum description length; CoGEnT: conditional gradient with enhancement and truncation; BSGFL: generalized fused lasso foreground modeling; PSSV: partial sum of singular values; ROUTE: low-rank matrix recovery via robust outlier estimation; FGSR: factor group-sparse regularization; SubGM: subgradient method.

	0.05		0.5	
	Low-Rank	**Sparse**	**Low-Rank**	**Sparse**
RPCA	**0.000007 ± 0.0**	**0.00001 ± 0.0**	0.25 ± 0.005	0.591 ± 0.014
LR-MDL	0.061 ± 0.024	0.200 ± 0.072	0.379 ± 0.034	0.490 ± 0.024
CoGEnT	0.082 ± 0.032	1.218 ± 0.082	1.604 ± 0.585	1.541 ± 0.144
BSGFL	0.030 ± 0.004	0.275 ± 0.063	0.351 ± 0.026	10.04 ± 1.423
PSSV	**0.000002 ± 0.0**	**0.00002 ± 0.0**	0.207 ± 0.002	0.486 ± 0.005
ROUTE	0.002 ± 0.0001	0.007 ± 0.0001	0.141 ± 0.009	0.264 ± 0.015
FGSR	0.001 ± 0.0001	0.004 ± 0.0001	0.324 ± 0.007	0.454 ± 0.008
SubGM	**0.0006 ± 0.0**	0.003 ± 0.0001	0.168 ± 0.039	0.227 ± 0.005
MDLAN	**0.000002 ± 0.0**	**0.000003 ± 0.0**	**0.01 ± 0.007**	**0.019 ± 0.013**

4.2. Real-World Sensing Applications

4.2.1. 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 $Y \in \mathbb{R}^{m \times n} = [vec(I_1), \cdots, vec(I_n)]$ by stacking the vectorized input images (processing each color channel individually), 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).

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. The final HDR results are shown in Figure 5. Compared with the results obtained by RPCA, LR-MDL, CoGEnT, BSGFL, PSSV, ROUTE and FGSR, the proposed method can recover the low-rank component (artifact-free in Figure 5g) and reject more outliers, even with only four input images ($n = 4$). The detailed comparison in Figure 6 shows that our method can reject the outliers, such as ghosting and shadows, which are caused by the person and the wind, respectively. The reason for this is that 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. Furthermore,

the proposed method, utilizing the MDL principle to select the optimal atoms, can search for the best approximation of the sparse structure. Figure 4b–d also shows that the proposed MDLAN can estimate the intensity and the number of the nonzero entries in the sparse matrix.
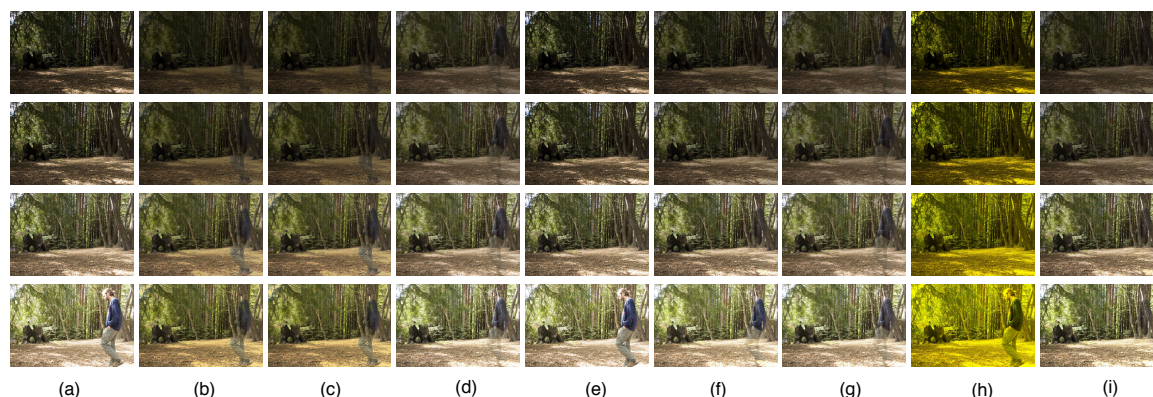

(a)　(b)　(c)　(d)　(e)　(f)　(g)　(h)　(i)

Figure 5. High dynamic range imaging. (**a**) Input images (four) and low-rank term *X* obtained by RPCA (**b**), LR-MDL (**c**), CoGEnT (**d**), BSGFL (**e**), PSSV (**f**), ROUTE (**g**), FGSR (**h**) and the proposed approach (**i**).

(a)　(b)

Figure 6. Detailed comparison of the branches and their shadows. Low-rank component obtained by RPCA (**a**) and the proposed approach (**b**).

4.2.2. Background Modeling Based on Video Sensor

We adopt the *F*-measure as the quantitative metric for the performance evaluation of the background modeling. The *F*-measure, which combines precision and recall, is calculated as follows:

$$F\text{-}measure = 2\frac{precision \cdot recall}{precision + recall} \tag{27}$$

where $precision = \frac{TP}{TP+FP}$ and $recall = \frac{TP}{TP+FN}$, 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 accurately the outliers (foreground objects) are detected [52].

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 moving objects 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. [53], which comprises a sequence of 1186 grayscale frames obtained from a busy shopping center. Multiple people move in the scene, and 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 256×320, and we stack the frames as the columns in our observed matrix $Y \in \mathbb{R}^{245760 \times 100}$.

The results are displayed in Figure 7, 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, PSSV, ROUTE and FGSR, as shown in Figure 7b–h. By contrast, our proposed method correctly models the background scene and gives a better foreground with fewer false detections.

We then consider two sequences from the stuttgart artificial background subtraction (SABS) data set, including a "Basic" sequence and a "Clutter" sequence. The "Clutter" category of sequences contains 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 Figures 8 and 9. As shown in Figure 8, the proposed method obtains a cleaner background (no ghosting) and detects more outliers compared to the other models when the corruption ratio is high. Figure 9 demonstrates that the proposed MDLAN can recover the low-rank background (no shadow) and almost cuts the foreground correctly compared to the 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 2. As illustrated in Figures 7–9, the shadow is included in the sparse component, which makes the value of the *F*-measure relatively low. Table 2 indicates that the proposed method can achieve the highest *F*-measure for the three sequences and also shows better computational efficiency.
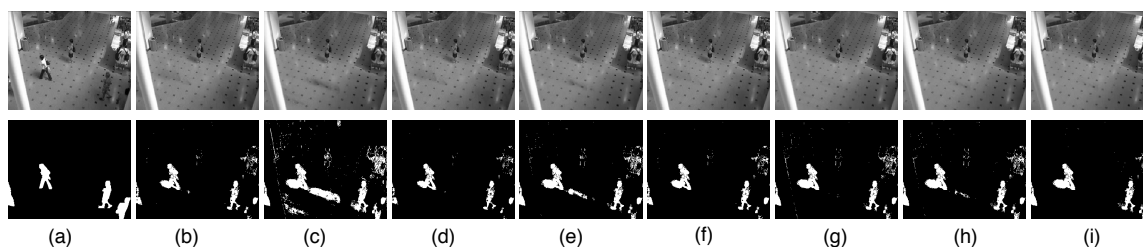

Figure 7. Background modeling results on the Li data set. (**a**) One frame of the original video Y (top) and the ground truth (bottom). Low-rank component X (top) and sparse component E (bottom) obtained by RPCA (**b**), LR-MDL (**c**), CoGEnT (**d**), BSGFL (**e**) PSSV (**f**), ROUTE (**g**), FGSR (**h**) and the proposed approach (**i**).
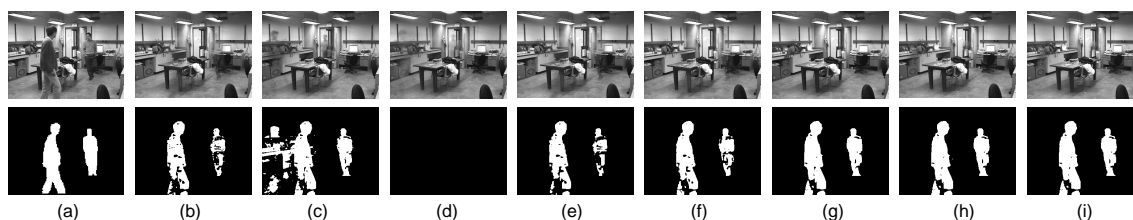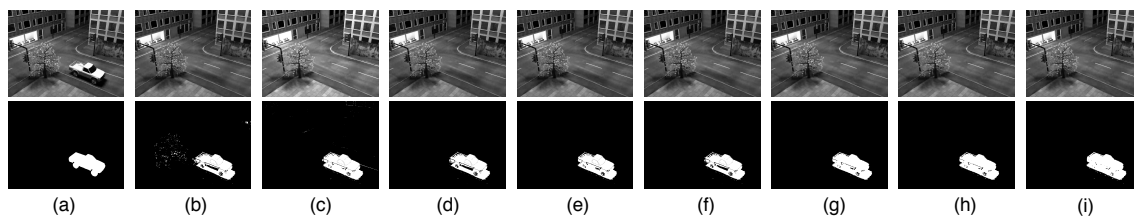

Figure 8. Background modeling results on the SABS data set. Each frame has a resolution of 240×320, and we stack the frames as the columns in our observed matrix $Y \in \mathbb{R}^{76800 \times 100}$. (**a**) One frame of the original video Y (top) and the ground truth (bottom). Low-rank component X (top) and sparse component E (bottom) obtained by RPCA (**b**), LR-MDL (**c**), CoGEnT (**d**), BSGFL (**e**) PSSV (**f**), ROUTE (**g**), FGSR (**h**) and the proposed approach (**i**).

Figure 9. Background modeling results on the SABS data set. Each frame has a resolution of 600×800, and we stack the frames as the columns in our observed matrix $Y \in \mathbb{R}^{480000 \times 100}$. (**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**), ROUTE (**g**), FGSR (**h**) and the proposed approach (**i**).

Table 2. Quantitative evaluation of the background modeling, given as the F-measure and running time.

	Shopping Mall		HumanBody2		MPEG	
	F-Measure	Times (s)	F-Measure	Times (s)	F-Measure	Times (s)
RPCA	0.6975	585	0.6881	985	0.7437	4087
LR-MDL	0.5021	912	0.6294	106	0.7977	446
CoGEnT	0.6846	1534	0.0676	586	0.7735	6498
BSGFL	0.6406	11628	0.6806	2219	0.7987	14942
PSSV	0.7060	20.1	0.7377	37.1	0.7756	207
ROUTE	0.7181	62.4	0.7801	41.5	0.8110	340
FGSR	0.7175	61.8	0.7837	50.2	0.8141	304
MDLAN	**0.7536**	28.6	**0.7941**	40.7	**0.8264**	227

4.2.3. Removing Noise and Shadows From Faces

Basri et al. [54] 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 $Y \in \mathbb{R}^{8064 \times 20}$.

Figure 10a shows three images from the Extended Yale B database [55], Figure 10b–i shows the recovered low-rank components and Figure 11a–h shows the corresponding sparse components. 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 Figure 10i). When the shaded area is large, the proposed method still removes more shadows than RPCA, LR-MDL, CoGEnT, BSGFL, PSSV, ROUTE and FGSR (see the third row in Figure 10g). In addition, we add salt and pepper noise to each observed image, and the noise density is 0.2. Figure 12b–i shows the recovered low-rank components. Compared to the above methods, the proposed MDLAN can remove both noise and shadows. Thus, our technique may be useful for pre-processing training images in face recognition systems by removing such noise/outliers.
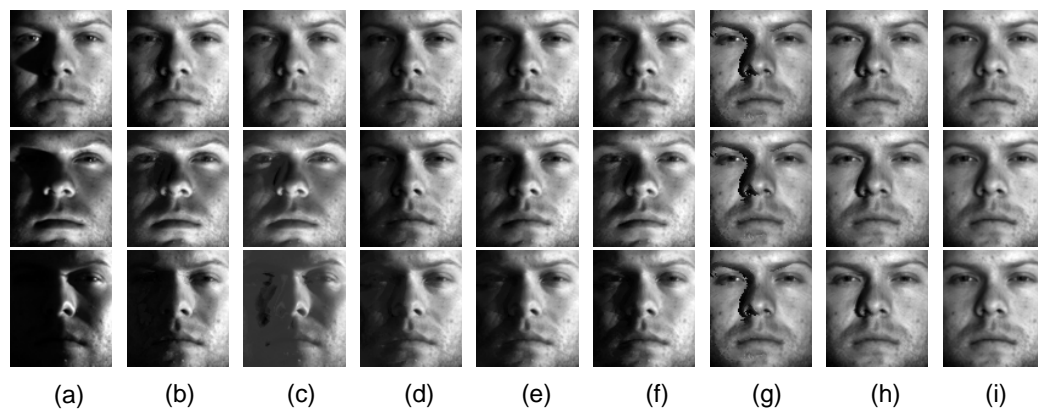
Figure 10. Removing shadows and specularities from face images. (**a**) Cropped and aligned images from the Extended Yale B database of a person's face under different illumination. The size of each image is 96 × 84 pixels and 20 different illumination settings in total were used for the person. Low-rank term *X* obtained by RPCA (**b**), LR-MDL (**c**), CoGEnT (**d**), BSGFL (**e**), PSSV (**f**), ROUTE (**g**), FGSR (**h**) and the proposed approach (**i**).
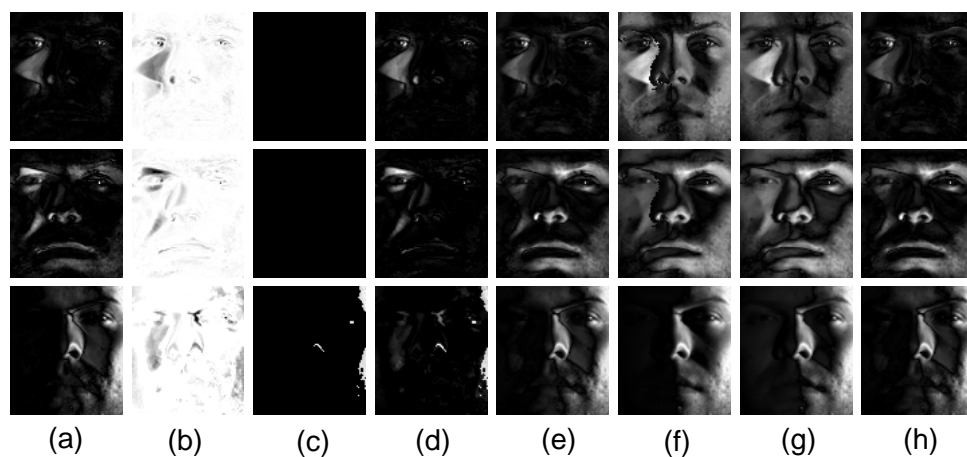
Figure 11. Sparse component *E* obtained by RPCA (**a**), LR-MDL (**b**), CoGEnT (**c**), BSGFL (**d**), PSSV (**e**), ROUTE (**f**), FGSR (**g**) and the proposed approach (**h**).
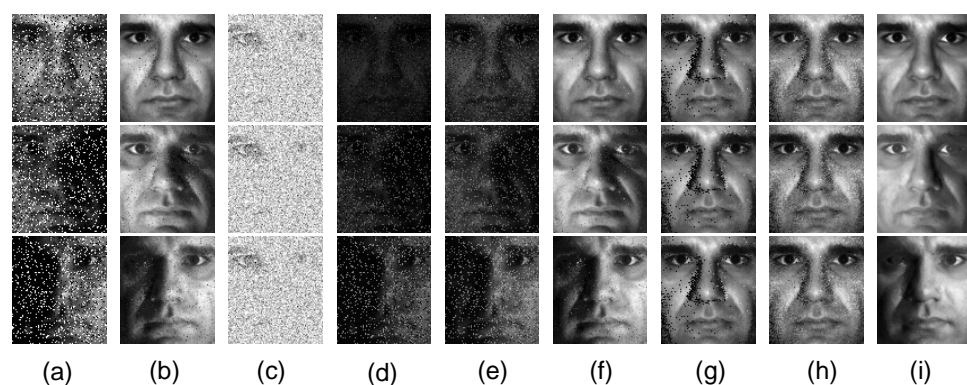

Figure 12. Removing noise, shadows and specularities from face images. (**a**) Cropped and aligned images from the Extended Yale B database of a person's face under different illumination. The size of each image is 96 × 84 pixels, and 20 different illumination settings in total were used for the person. We also added salt and pepper noise to each image. Low-rank term *X* obtained by RPCA (**b**), LR-MDL (**c**), CoGEnT (**d**), BSGFL (**e**), PSSV (**f**), ROUTE (**g**), FGSR (**h**) and the proposed approach (**i**).

5. Conclusions

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; thus, we consider an approximation of the original problem. Our method selects the best atoms to search for the best approximation of the low-rank matrix, and it also can find 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 a video sensor and the removal of noise and shadows from face images. The experimental results using the synthetic and real sensing data sets demonstrate the effectiveness and robustness of the proposed approach.

Author Contributions: A.Q. is responsible for all the theoretical work, the implementation of the experiments and the writing of the manuscript. L.X. and Y.Y. are responsible for the literature search, data collection and production of charts. T.Z. and Y.Y.T. revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Encoding Scheme

In our MDL framework, we need to encode the low-rank matrix ($\sum_i \mathcal{L}(\alpha_i \psi_i)$) and the sparse matrix ($\mathcal{L}(\mathcal{F}_\Phi \beta)$), respectively. It is usual to extend the ideal codelength to continuous random variables x with a probability assignment $P(x)$ as $\mathcal{L}(x) = -log P(x) \approx -log(p(x)\delta)$, and $p(x)$ is the probability density function of variables x. To losslessly encode the finite-precision obtained variables, we quantize the variables with the step $\delta = 1$ [22].

Appendix A.1. Encoding the Sparse Matrix

For the sake of simplicity, we set the elements in atomic set Φ to be positive signs and the scalar coefficients β to be mixed signs. We assume that the scalar coefficients β comprise a sequence of Laplace random variables [22]:

$$
\begin{aligned}
\mathcal{L}(\mathcal{F}_\Phi \beta) &= \mathcal{L}(\beta|\Phi) + \mathcal{L}(\Phi) \\
&= -log\, P(\beta; \theta) + k\, log(mn) \\
&= \theta \|\beta\|_1 + c
\end{aligned}
\tag{A1}
$$

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

Appendix A.2. Encoding the Low-Rank Matrix

It is not surprising that each atom represents numerous pieces of eigeninformation 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 smooth in a piecewise manner. Thus, we should efficiently exploit the smoothness of the atoms by employing a prediction scheme. Concretely, to describe each column ($\psi_i = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n]$) of the atoms, we reshape it ($\mathbf{a}_j$) 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 $\hat{\mathbf{a}}_j$ (each element is given by *north_element* + *west_element* − *northwest_element*), obtaining the residual $\bar{\mathbf{a}}_j = \mathbf{a}_j - \hat{\mathbf{a}}_j$. In particular, the residual can be written as $\bar{\mathbf{a}}_j = W\mathbf{a}_j$, and the matrix residual can be formed as $\bar{\psi}_i = W\psi_i$, where $W \in \mathbb{R}^{m \times m}$ is in a lower triangular form. The detailed procedure is depicted in Figure A1. We refer to [15,22] for details on these results.

　　We also assume the prediction residual $\bar{\psi}_i$ to be a sequence of *LG*-distributed continuous random variables [15,22]. Compared to the codelength of $\bar{\psi}_i$, the codelength of scalar coefficient α_i is inconsequential for our model. Thus, the codelength of low-rank matrix X can be written as $\sum_i \mathcal{L}(W\psi_i)$.

causal bilinear kernel

−1	1
1	x

zero-padding

0	0	0	0
0	4	3	2
0	4	3	2
0	4	3	1

$$
\begin{bmatrix} 4 \\ 0 \\ 0 \\ -1 \\ 0 \\ 0 \\ -1 \\ 0 \\ 1 \end{bmatrix}
=
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & -1 & 0 & -1 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & -1 & 0 & -1 & 1
\end{bmatrix}
\times
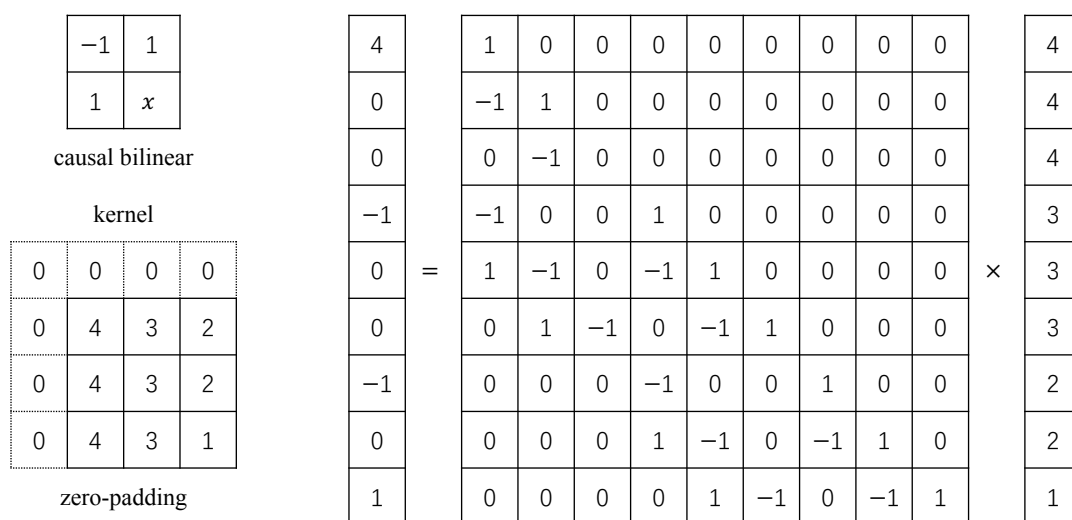\begin{bmatrix} 4 \\ 4 \\ 4 \\ 3 \\ 3 \\ 3 \\ 2 \\ 2 \\ 1 \end{bmatrix}
$$

Figure A1. The encoding procedure of the prediction scheme. The column of an atom is arranged as a 3×3 matrix, and the elements outside of the range are assumed to be zero. The causal bilinear predictor is assumed to be a 2×2 template, and the mapping matrix W is of the size 9×9.

References

1.　Zhao, Q.; Meng, D.; Xu, Z.; Zuo, W.; Zhang, L. Robust Principal Component Analysis with Complex Noise. In Proceedings of the 31st International Conference on Machine Learning (ICML 2014), Beijing, China, 21–26 June 2014.

2.　Oh, T.H.; Tai, Y.W.; Bazin, J.C.; Kim, H.; Kweon, I.S. Partial Sum Minimization of Singular Values in Robust PCA: Algorithm and Applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 744–758. [CrossRef] [PubMed]

3.　Xia, G.; Sun, H.; Chen, B.; Liu, Q.; Hang, R. Nonlinear Low-Rank Matrix Completion for Human Motion Recovery. *IEEE Trans. Image Process.* **2018**, *27*, 3011–3024. [CrossRef] [PubMed]

4.　Zhuang, L.; Gao, S.; Tang, J.; Wang, J.; Lin, Z.; Ma, Y.; Yu, N. Constructing a Nonnegative Low-Rank and Sparse Graph With Data-Adaptive Features. *IEEE Trans. Image Process.* **2015**, *24*, 3717–3728. [CrossRef] [PubMed]

5.　Zhang, H.; Gong, C.; Qian, J.; Zhang, B.; Xu, C.; Yang, J. Efficient Recovery of Low-Rank Matrix via Double Nonconvex Nonsmooth Rank Minimization. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 2916–2925. [CrossRef] [PubMed]

6.　Zhang, H.; Qian, J.; Zhang, B.; Yang, J.; Gong, C.; Wei, Y. Low-Rank Matrix Recovery via Modified Schatten-p Norm Minimization With Convergence Guarantees. *IEEE Trans. Image Process.* **2020**, *29*, 3132–3142. [CrossRef] [PubMed]

7.　Chen, J.; Zhou, J.; Ye, J. Integrating low-rank and group-sparse structures for robust multi-task learning. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2011), San Diego, CA, USA, 21–24 August 2011.

8. Chen, W. Simultaneously Sparse and Low-Rank Matrix Reconstruction via Nonconvex and Nonseparable Regularization. *IEEE Trans. Signal Process.* **2018**, *66*, 5313–5323. [CrossRef]

9. Xie, X.; Wu, J.; Liu, G.; Wang, J. Matrix Recovery with Implicitly Low-Rank Data. *Neurocomputing* **2019**, *334*, 219–226. [CrossRef]

10. Jolliffe, I.T. *Principal Component Analysis*; Springer: Berlin/Heidelberg, Germany, 1986; Volume 14, pp. 231–246.

11. Ganesh, A.; Ma, Y.; Rao, S.; Wright, J. Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices via Convex Optimization. *Adv. Neural Inf. Process. Syst.* **2009**, *87*, 20:3–20:56.

12. Candes, S.E.J.; Li, X.; Ma, Y.; Wright, J. Robust principal component analysis. *J. Acm* **2011**, *58*, 1–73. [CrossRef]

13. Chen, M.; Lin, Z.; Ma, Y.; Wu, L. The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices. Available online: https://people.eecs.berkeley.edu/~yima/psfile/Lin09-MP.pdf (accessed on 25 October 2020)

14. Wright, J. Low-Rank Matrix Recovery and Completion via Convex Optimization. Available online: http://perception.csl.illinois.edu/matrix-rank/home.html (accessed on 25 October 2020)

15. Ramirez, I.; Sapiro, G. Low-rank data modeling via the minimum description length principle. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Prague, Czech Republic, 22–27 May 2011.

16. Bouwmans, T.; Zahzah, E.H. Robust PCA via Principal Component Pursuit: A review for a comparative evaluation in video surveillance. *Comput. Vis. Image Underst.* **2014**, *122*, 22–34. [CrossRef]

17. Liu, X.; Zhao, G.; Yao, J.; Qi, C. Background subtraction based on low-rank and structured sparse decomposition. *IEEE Trans. Image Process.* **2015**, *24*, 2502–2514. [CrossRef] [PubMed]

18. Gallo, O.; Gelfandz, N. Artifact-free High Dynamic Range Imaging. In Proceedings of the 2009 IEEE International Conference on Computational Photography (ICCP), San Francisco, CA, USA, 16–17 April 2009.

19. Celebi, A.T.; Duvar, R.; Urhan, O. Fuzzy fusion based high dynamic range imaging using adaptive histogram separation. *IEEE Trans. Consum. Electron.* **2015**, *61*, 119–127. [CrossRef]

20. Mariani, A.; Giorgetti, A.; Chiani, M. Model Order Selection Based on Information Theoretic Criteria: Design of the Penalty. *IEEE Trans. Signal Process.* **2015**, *63*, 2779–2789. [CrossRef]

21. Ding, J.; Tarokh, V.; Yang, Y. Model Selection Techniques: An Overview. *IEEE Signal Process. Mag.* **2018**, *35*, 16–34. [CrossRef]

22. Ramirez, I.; Sapiro, G. An MDL Framework for Sparse Coding and Dictionary Learning. *IEEE Trans. Signal Process.* **2011**, *60*, 2913–2927. [CrossRef]

23. Rissanen, J. Modeling by shortest data description. *Automatica* **1978**, *14*, 465–471. [CrossRef]

24. Chandrasekaran, V.; Recht, B.; Parrilo, P.A.; Willsky, A.S. The Convex Geometry of Linear Inverse Problems. *Found. Comput. Math.* **2012**, *12*, 805–849. [CrossRef]

25. Rao, N.; Shah, P.; Wright, S. Forward-Backward Greedy Algorithms for Atomic Norm Regularization. *IEEE Trans. Signal Process.* **2014**, *63*, 5798–5811. [CrossRef]

26. Wang, Y.; Tang, Y.Y.; Li, L. Minimum Error Entropy Based Sparse Representation for Robust Subspace Clustering. *IEEE Trans. Signal Process.* **2015**, *63*, 4010–4021. [CrossRef]

27. Qin, A.; Shang, Z.; Zhang, T.; Ding, Y.; Tang, Y.Y. Minimum Description Length Principle Based Atomic Norm for Synthetic Low-Rank Matrix Recovery. In Proceedings of the 2016 7th International Conference on Cloud Computing and Big Data (CCBD), Macau, China, 16–18 November 2016.

28. Zhou, X.; Yang, C.; Yu, W. Moving object detection by detecting contiguous outliers in the low-rank representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 597–610. [CrossRef]

29. Xu, J.; Ithapu, V.K.; Mukherjee, L.; Rehg, J.M.; Singh, V. GOSUS: Grassmannian online subspace updates with structured-sparsity. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013.

30. Xin, B.; Tian, Y.; Wang, Y.; Gao, W. Background subtraction via generalized fused lasso foreground modeling. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.

31. Xin, B.; Kawahara, Y.; Wang, Y.; Gao, W. Efficient generalized fused lasso and its application to the diagnosis of Alzheimer's disease. In Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, Québec City, QC, Canada, 27–31 July 2014.

32. Rudin, L.I.; Osher, S.; Fatemi, E. Nonlinear total variation based noise removal algorithms. *Physics D* **1992**, *60*, 259–268. [CrossRef]

33. Erfanian, S.; Ebadi; Ebroul; Izquierdo. Foreground Segmentation with Tree-Structured Sparse RPCA. *IEEE TTrans. Pattern Anal. Mach. Intell.* **2018**, *40*, 2273–2280.

34. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2274–2282. [CrossRef] [PubMed]

35. Shah, S.; Goldstein, T.; Studer, C. Estimating sparse signals with smooth support via convex programming and block sparsity. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.

36. Zha, Z.; Liu, X.; Huang, X.; Shi, H.; Xu, Y.; Wang, Q.; Tang, L.; Zhang, X. Analyzing the group sparsity based on the rank minimization methods. In Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), Hong Kong, China, 10–14 July 2017.

37. Cabral, R.; Torre, F.D.L.; Costeira, J.P.; Bernardino, A. Unifying Nuclear Norm and Bilinear Factorization Approaches for Low-Rank Matrix Decomposition. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013.

38. Guo, X.; Lin, Z.; Center, C.M.I. ROUTE: Robust Outlier Estimation for Low Rank Matrix Recovery. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, Melbourne, Australia, 19–25 August 2017.

39. Guo, X.; Lin, Z. Low-rank matrix recovery via robust outlier estimation. *IEEE Trans. Image Process.* **2018**, *27*, 5316–5327. [CrossRef] [PubMed]

40. Hu, Y.; Zhang, D.; Ye, J.; Li, X.; He, X. Fast and Accurate Matrix Completion via Truncated Nuclear Norm Regularization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2117–2130. [CrossRef]

41. Guo, X.; Zhao, R.; An, G.; Cen, Y. An algorithm of face alignment and recognition by sparse and low rank decomposition. In Proceedings of the 2014 12th International Conference on Signal Processing (ICSP), Hangzhou, China, 19–23 October 2014.

42. Oh, T.H.; Lee, J.Y.; Tai, Y.W.; Kweon, I.S. Robust High Dynamic Range Imaging by Rank Minimization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1219–1232. [CrossRef]

43. Peng, Y.; Suo, J.; Dai, Q.; Xu, W. Reweighted low-rank matrix recovery and its application in image restoration. *IEEE Trans. Cybern.* **2014**, *44*, 2418–2430. [CrossRef]

44. Huang, C.; Ding, X.; Fang, C.; Wen, D. Robust Image Restoration Via Adaptive Low-Rank Approximation and Joint Kernel Regression. *IEEE Trans. Image Process.* **2014**, *23*, 5284–5297. [CrossRef]

45. Gao, P.; Wang, R.; Meng, C.; Joe, H. Low-Rank Matrix Recovery From Noisy, Quantized, and Erroneous Measurements. *IEEE Trans. Signal Process.* **2018**, *66*, 2918–2932. [CrossRef]

46. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley: New York, NY, USA, 2005; Volume 1, pp. 1–748.

47. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. Available online: https://web.stanford.edu/~boyd/papers/pdf/admm_distr_stats.pdf (accessed on 25 October 2020)

48. Candès, E.J.; Wakin, M.B.; Boyd, S.P. Enhancing Sparsity by Reweighted l_1 Minimization. *J. Fourier Anal. Appl.* **2008**, *14*, 877–905. [CrossRef]

49. Hale, E.T.; Yin, W.; Zhang, Y. Fixed-Point Continuation for ℓ_1-Minimization: Methodology and Convergence. *Siam J. Optim.* **2008**, *19*, 1107–1130. [CrossRef]

50. Fan, J.; Ding, L.; Chen, Y.; Udell, M. Factor group-sparse regularization for efficient low-rank matrix recovery. In Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, BC, Canada, 8–14 December 2019.

51. Li, X.; Zhu, Z.; Man-Cho So, A.; Vidal, R. Nonconvex robust low-rank matrix recovery. *Siam J. Optim.* **2020**, *30*, 660–686. [CrossRef]

52. Davis, J.; Goadrich, M. The relationship between Precision-Recall and ROC curves. In Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA, USA, 25–29 June 2006.

53. Li, L.; Huang, W.; Gu, Y.H.; Tian, Q. Statistical Modeling of Complex Backgrounds for Foreground Object Detection. *IEEE Trans. Image Process.* **2004**, *13*, 1459–1472. [CrossRef] [PubMed]

54. Basri, R.; Jacobs, D. Lambertian Reflectance and Linear Subspaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 218–233. [CrossRef]

55. Georghiades, A.S.; Belhumeur, P.N.; Kriegman, D.J. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *23*, 643–660. [CrossRef]