Multispectral Images Denoising by Intrinsic Tensor Sparsity Regularization

Qi Xie[1], Qian Zhao[1], Deyu Meng[1,*], Zongben Xu[1], Shuhang Gu[2], Wangmeng Zuo[3], Lei Zhang[2]
[1]Xi'an Jiaotong University; [2]The Hong Kong Polytechnic University; [3]Harbin Institute of Technology
xq.liwu@stu.xjtu.edu.cn timmy.zhaoqian@gmail.com {dymeng zbxu}@mail.xjtu.edu.cn
shuhanggu@gmail.com wmzuo@hit.edu.cn cslzhang@comp.polyu.edu.hk

Abstract

Multispectral images (MSI) can help deliver more faithful representation for real scenes than the traditional image system, and enhance the performance of many computer vision tasks. In real cases, however, an MSI is always corrupted by various noises. In this paper, we propose a new tensor-based denoising approach by fully considering two intrinsic characteristics underlying an MSI, i.e., the global correlation along spectrum (GCS) and nonlocal self-similarity across space (NSS). In specific, we construct a new tensor sparsity measure, called intrinsic tensor sparsity (ITS) measure, which encodes both sparsity insights delivered by the most typical Tucker and CANDECOMP/PARAFAC (CP) low-rank decomposition for a general tensor. Then we build a new MSI denoising model by applying the proposed ITS measure on tensors formed by non-local similar patches within the MSI. The intrinsic GCS and NSS knowledge can then be efficiently explored under the regularization of this tensor sparsity measure to finely rectify the recovery of a MSI from its corruption. A series of experiments on simulated and real MSI denoising problems show that our method outperforms all state-of-the-arts under comprehensive quantitative performance measures.

1. Introduction

A multispectral image (MSI) consists of multiple images of a real scene captured by sensors over various discrete bands. As compared with traditional image collecting systems which integrate the product of the intensity at only a few typical band intervals, MSI facilitates a fine delivery of more faithful knowledge under real scenes. Such full knowledge representation capability of MSI has been substantiated to greatly enhance the performance of various computer vision tasks, such as superresolution [13], inpainting [7], and tracking [22].

In real cases, however, due to the acquisition errors conducted by sensor, an MSI generally contains certain extent of noises, which inclines to negatively influence the subsequent MSI processing tasks. Therefore, MSI denoising has become a critical and inevitable issue for MSI analysis.

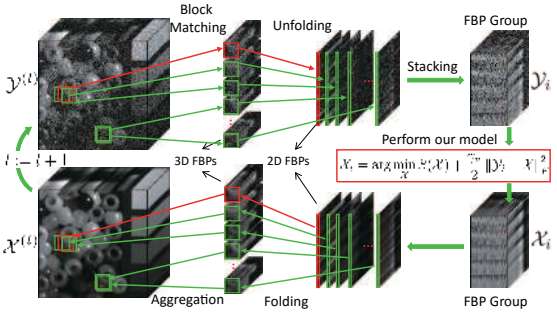


Figure 1. Flowchart of the proposed MSI denoising algorithm.

The most significant issue of recovering a clean MSI from its corruption is to rationally extract prior structure knowledge under a to-be-reconstructed MSI, and fully utilize such prior information to rectify the configuration of the recovered MSI in a sound manner. The most commonly utilized prior structures for MSI recovery include its global correlation along spectrum (GCS) and nonlocal self-similarity across space (NSS). Specifically, the GCS prior indicates that an MSI contains a large amount of spectral redundancy and the images obtained across the spectrum of an MSI are generally highly correlated. And the NSS prior refers to the fact that for a given local fullband patch (FBP) of an MSI (which is stacked by patches at the same location of MSI over all bands), there are many FBPs similar to it. It has been extensively shown that such two kinds of prior knowledge are very helpful for various MSI recovery problems [13, 7, 35, 23].

Albeit demonstrated to be effective to certain MSI denoising cases, most of the current methods to this task only consider one such prior knowledge in their model, like BM3D [8] only considering the NSS and PARAFAC [16]) only considering the GCS. Their potential capacity thus still has room to be further enhanced. The TDL [23] method was recently proposed by taking both priors into account and achieved the state-of-the-art MSI denoising performance. The method, however, coarsely encodes the NSS prior under relatively small amount of FBP clusters while does not

*Corresponding author.

fully consider the entire NSS knowledge across all FBPs. Besides, its realization is relatively heuristic and short of a concise formulation to abstract such latent priors underlying an MSI, which makes the methodology hard to be extended to other MSI recovery problems.

To alleviate this problem, this paper proposes a new MSI denoising technique which not only fully takes both GCS and NSS knowledge into account, but also is with a concise formulation to regularize such priors which can be easily transferred to general MSI restoration problems. Specifically, we regard each FBP as a matrix with a spatial mode and a spectral mode, and build a 3-order tensor by stacking all its non-local similar FBPs (see the upper part of Fig. 1). Such a tensor naturally forms a faithful representation to deliver both the latent GCS and NSS knowledge underlying the MSI. Since GCS and NSS imply the correlation along the spectral and FBP-number modes of this tensor, respectively, the key problem is then transferred to how to construct a rational sparsity measure to reflect such correlation and use it to regularize the MSI recovery from corrupted one.

To handle the aforementioned issues, this paper makes the following three-fold contributions. Firstly, a new measure for tensor sparsity is proposed. Beyond traditional tensor sparsity measures without an evident physical meaning, this new measure can be easily interpreted as a regularization for the number of rank-1 Kronecker bases for representing this tensor. Such measure not only unifies the traditional understanding of sparsity from vector (1-order tensor) to matrix (2-order tensor), but also encodes both sparsity insights delivered by the most typical Tucker and CP low-rank decomposition for a general tensor. We thus call it intrinsic tensor sparsity (ITS) for convenience.

Secondly, we propose a new tensor-based denoising model by performing tensor recovery with the proposed ITS measure to encode the inherent spatial and spectral correlation of the nonlocal similar FBP groups. The model is with a concise formulation and can be easily extended to solving other MSI recovery problems.

Thirdly, we design an effective alternating direction method of multipliers (ADMM)[2, 15] based algorithm for solving the model, and deduce the closed-form equations for updating each involved parameter, which makes it able to be efficiently implemented. Experiments on benchmark and real MSI data show that the proposed method achieves the state-of-the-art performance on MSI denoising among various quality assessments.

Throughout the paper, we denote scalar, vector, matrix and tensor as non-bold lower case, bold lower case, upper case and calligraphic upper case letters, respectively.

2. Notions and preliminaries

A tensor, shown as a multi-dimensional data array, is a multilineal mapping over a set of vector spaces. A tensor of order N is denoted as $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots I_N}$. Elements of $\mathcal{A}$ are denoted as $a_{i_1 \cdots i_n \cdots i_N}$ where $1 \leq i_n \leq I_n$. The mode-n vectors of an N-order tensor $\mathcal{A}$ are the I_n dimensional vectors obtained from $\mathcal{A}$ by varying index i_n while keeping the others fixed. The unfolding matrix $A_{(n)} = \text{unfold}_n(\mathcal{A}) \in \mathbb{R}^{I_n \times (I_1 \cdots I_{n-1}, I_{n+1} \cdots I_N)}$ is composed by taking the mode-n vectors of $\mathcal{A}$ as its columns. This matrix can also be naturally seen as the mode-n flattening of the tensor $\mathcal{A}$. Conversely, the unfolding matrices along the n^{th} mode can be transformed back to the tensor by $\mathcal{A} = \text{fold}_n\left(A_{(n)}\right), 1 \leq n \leq N$. The n-rank $\mathcal{A}$, denoted as r_n, is the dimension of the vector space spanned by the mode-n vectors of $\mathcal{A}$.

The product between matrices can be generalized to the product of a tensor and a matrix. The mode-n product of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots I_n}$ by a matrix $B \in \mathbb{R}^{J_n \times I_n}$, denoted by $\mathcal{A} \times_n B$, is an N-order tensor $\mathcal{C} \in \mathbb{R}^{I_1 \times \cdots \times J_n \times \cdots I_N}$, whose entries are computed by

$$c_{i_1 \times \cdots i_{n-1} \times j_n \times i_{n+1} \ldots i_N} = \sum\nolimits_{i_n} a_{i_1 \cdots i_n \cdots i_N} b_{j_n i_n}.$$

The mode-n product $\mathcal{C} = \mathcal{A} \times_n B$ can also be calculated by the matrix multiplication $C_{(n)} = BA_{(n)}$, followed by the re-tensorization of undoing the mode-n flattening. The Frobenius norm of an tensor $\mathcal{A}$ is $\|\mathcal{A}\|_F = \left(\sum_{i_1, \cdots i_n} |a_{i_1, \cdots i_N}|^2\right)^{1/2}$.

We call a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots I_N}$ is rank-1 if it can be written as the outer product of N vectors, i.e.,

$$\mathcal{A} = \mathbf{a}^{(1)} \circ \mathbf{a}^{(2)} \circ \cdots \circ \mathbf{a}^{(N)},$$

where $\circ$ represents the vector outer product. This means that each element of the tensor is the product of the corresponding vector elements:

$$a_{i_1, i_2, \cdots, i_N} = a^{(1)}_{i_1} a^{(2)}_{i_2} \ldots a^{(N)}_{i_N} \quad \forall\, 1 \leq i_n \leq I_n. \qquad (1)$$

such a simple rank-1 tensor is also called a Kronecker basis in the tensor space. For example, in a 2D case, a Kronecker basis is a rank-1 matrix expressed as the outer product $\mathbf{u}\mathbf{v}^T$ of two vectors $\mathbf{u}$ and $\mathbf{v}$.

3. Related work

Approaches for MSI denoising can be generally grouped into two categories: the 2D extended approach and the tensor-based approach.

2D extended approach. As a classical problem in computer vision, 2D image denoising has been studied for more than 50 years and a large amount of methods have been proposed on this problem, such as NLM [3], K-SVD [10] and BM3D [8]. These methods can be directly applied to MSI denoising by treating the images located at different bands separately. This extension, however, neglects the intrinsic properties of MSIs and generally cannot attain good perfor-

mance in real applications. Another more reasonable extension is specifically designed for the patch-based image denoising methods, which takes the small local patches of the image into consideration. By building small 3D cubes of an MSI instead of 2D patches of a traditional image, the corresponding 3D-cube-based MSI denoising algorithm can then be constructed [24]. The state-of-the-art of 3D-cube-based approach is represented by the BM4D method [18, 19], which exploits the 3D NSS of MSI to remove noise in similar MSI 3D cubes collaboratively. The deficiency of these methods is that they neglect the useful GCS knowledge underlying an MSI, and still have not essentially reached the full potential for handling this task.

Tensor-based approach. An MSI is composed by a stack of 2D images, which can be naturally regarded as a 3-order tensor. The tensor-based approach implements the MSI denoising by applying the tensor factorization techniques to the MSI tensor. Along this research line, Renard et al. [21] presented a low-rank tensor approximation (LRTA) method by employing the Tucker decomposition [27] to obtain the low-rank approximation of the input MSI. Liu et al. [16] designed the PARAFAC method by utilizing the parallel factor analysis [6]. The advantage of both methods is that they take the correlation between MSI images over different bands into consideration, and try to eliminate the spectral redundancy of MSIs. However, they have not utilized the NSS prior of MSI. The state-of-the-art method of this category is represented by tensor dictionary learning (TDL) [23] which takes both GCS and NSS under MSI into account. This method, however, only consider NSS among several FBP clusters while not fully utilize the fine-grained NSS structures across all FBPs over the tensor space. There is thus still much room for further improvement.

4. MSI denoising by intrinsic tensor sparsity regularization

4.1. GCS and NSS modeling for MSI denoising

We first briefly introduce a general NSS-based framework for image denoising, which has been adopted by multiple literatures in image cases [12], aiming to reconstruct the original image Z from its noisy observation Y. Separating Y into a set of image patches $\Omega = \{y_i \in \mathbb{R}^d\}_{i=1}^N$ (where d is the pixel number of each patch) with overlap, and by performing block matching [8], a set of patches which is most similar to each patch y_i can be extracted. By stacking all these patches to form a matrix $Y_i \in \mathbb{R}^{d\times n}$, where n is the number of these nonlocal similar patches, we can then recover the corresponding original nonlocal-similar-patch-matrix X_i through

$$X_i = \arg\min_X S(X) + \frac{\gamma}{2}\|Y_i - X\|_F^2, \tag{2}$$

where $S(X)$ denotes the 2-order sparsity measure on the true matrix X and γ is the compromise parameter. The matrix rank is generally recognized as a rational sparsity measure for matrix [31], and it as well as its relaxations can thus be readily adopted into the model for implementation. When all X_is are obtained, the recovered image Z can then be estimated by aggregating X_i at each pixels.

The similar denoising model can be easily extended to MSI cases. Denote d_H, d_W and d_S as the spatial height, spatial width and spectral band number of an MSI, and we can express it as a 3-order tensor $\mathcal{Y} \in \mathbb{R}^{d_H\times d_W\times d_S}$ with two spatial modes and one spectral mode. By sweeping all across the MSI with overlaps, we can build a group of 2D FBPs $\{P_{ij}\}_{1\le i\le d_H-d_h, 1\le j\le d_W-d_w} \subset \mathbb{R}^{d_h d_w\times d_S}$ ($d_h < d_H$, $d_w < d_W$) to represent the MSI, where each band of a FBP is ordered lexicographically as a column vector. We can now reformulate all FBPs as a group of 2D patches $\Omega_{\mathcal{Y}} = \{Y_i \in \mathbb{R}^{d_h d_w\times d_S}\}_{i=1}^N$, where $N = (d_H - d_h + 1)\times(d_W - d_w + 1)$ is the number of patches over the whole MSI. According to the NSS of MSI, for a given local FBP Y_i, we can find a collection of FBPs similar to it from $\Omega_{\mathcal{Y}}$ in a non-local neighboring area of it. Denote $\mathcal{Y}_i \in \mathbb{R}^{d_h d_w\times d_S\times d_n}$ (where d_n is the number of non-local similar FBPs of Y_i) as the 3-order tensor stacked by Y_i and its non-local similar FBPs in $\Omega_{\mathcal{Y}}$, and then both GCS and NSS knowledge are well preserved and reflected by such representation, along its spectral and nonlocal-similar-patch-number modes, respectively.

Then, similar to the image cases, we can estimate the corresponding true nonlocal similarity FBPs $\mathcal{X}_i$ from its corruption $\mathcal{Y}_i$ by solving the following optimization problem:

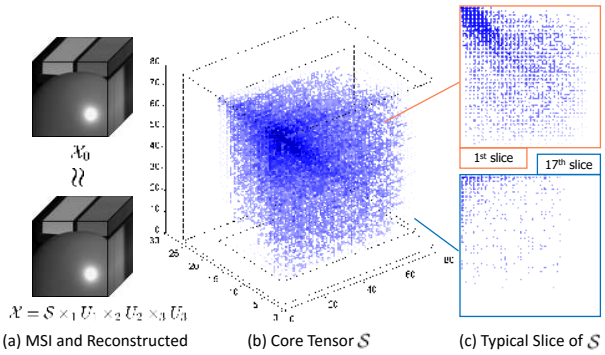
$$\mathcal{X}_i = \arg\min_{\mathcal{X}} S(\mathcal{X}) + \frac{\gamma}{2}\|\mathcal{Y}_i - \mathcal{X}\|_F^2, \tag{3}$$

where $S(\mathcal{X})$ is the sparsity measure imposed on $\mathcal{X}$. By aggregating all reconstructed $\mathcal{X}_i$s we can reconstruct the estimated MSI. The whole denoising progress can be easily understood by seeing Fig. 1. Obviously, the key issue now is to design an appropriate tensor sparsity measure on $\mathcal{X}$.

Different from the vector/matrix cases, where the sparsity measure can be easily constructed as nonzero-element-number/matrix-rank based on very direct intuitions, constructing a rational tensor sparsity is a relatively more difficult task. Most of the current work directly extended the 2-order sparsity measure to higher-order cases by easily ameliorating it as the weighted sum of ranks (or its relaxations) along all tensor modes [16, 25, 29, 5], i.e.,

$$S(\mathcal{X}) = \sum\nolimits_{i=1}^{d} w_i \text{rank}(X_{(i)}). \tag{4}$$

Such formulation, however, on one hand is short of a clear physical meaning for general tensors, and on the other hand lacks a consistent relationship with previous defined sparsity measures for vector/matrix. To ameliorate this issue,



(a) MSI and Reconstructed (b) Core Tensor $\mathcal{S}$ (c) Typical Slice of $\mathcal{S}$

Figure 2. (a) An MSI $\mathcal{X}_0 \in \mathbb{R}^{80\times80\times26}$ (upper) and a nearly perfect reconstruction $\mathcal{X}_0$ (PSNR=61.25). (b) Core tensor $\mathcal{S} \in \mathbb{R}^{69\times71\times17}$ of $\mathcal{X}$. Note that 78.4% of its elements are zeroes and more than half of them are very small. (c) Typical Slices of $\mathcal{S}$, where deeper color of the element represents a larger value of it.

we attempt to propose a new measure for more rationally assessing tensor sparsity.

4.2. Intrinsic tensor sparsity measure

We first briefly review two particular forms of tensor decomposition, both containing insightful understanding of tensor sparsity: Tucker decomposition [27] and CP decomposition [14].

In Tucker decomposition, an N-order tensor $\mathcal{X} \in \mathbb{R}^{I_1\times I_2\times\ldots\times I_N}$ is written as the following form:

$$\mathcal{X} = \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 \ldots \times_N U_N, \tag{5}$$

where $\mathcal{S} \in \mathbb{R}^{r_1\times r_2\times 3\ldots\times_N r_N}$ is called the core tensor, and $U_i \in \mathbb{R}^{I_i\times r_i}(1 \le i \le N)$ is composed by the r_i orthogonal bases along the i^{th} mode of $\mathcal{X}$. Tucker decomposition considers the low-rank property of the vector subspace unfolded along each of its modes. Such a sparsity understanding naturally conducts a block shape for the coefficients affiliated from all combinations of tensor subspace bases, represented by the core tensor term. It, however, has not considered the fine-grained sparsity configurations inside this coefficient tensor. Specifically, if we assume that the subspace bases along each mode have been sorted based on their importance for tensor representation, then the value of elements of the core tensor will show an approximate descending order along each of tensor modes. Along some modes, the corresponding tensor factor might have strong correlations across data, and then the coefficients in the core tensor along this mode tends to be decreasing very fast to zeroes. While for those modes with comparatively weaker correlation, albeit still approximately decreasing along the mode, the core tensor elements might be all non-zeroes. Fig. 2 depicts a visualization for facilitating the understanding of the above analysis. Tucker decomposition cannot well describe such an elaborate information, and thus is still hard to conduct a rational measure for comprehensively delivering the sparsity knowledge underlying a tensor.

CP decomposition attempts to decompose an N-order tensor $\mathcal{X} \in \mathbb{R}^{I_1\times I_2\times\ldots\times I_N}$ into the linear combination of a series of Kroneker bases, written as:

$$\mathcal{X} = \sum_{i=1}^{r} c_i \mathcal{V}_i = \sum_{i=1}^{r} c_i \mathbf{v}_{i_1} \circ \mathbf{v}_{i_2} \circ \ldots \circ \mathbf{v}_{i_N}, \tag{6}$$

where c_i denotes the coefficient imposed on the Kroneker basis $\mathcal{V}_i$. By arranging each Kroneker coefficients c_i into its corresponding $i_1, i_2, \cdots, i_N$ position of a core tensor, CP decomposition can be equivalently formulated as a Tucker decomposition form. Yet the core tensor will always be highly sparse since generally only a small amount of affiliated combinations of tensor bases are involved. Opposite to Tucker cases, such a tensor reformulation, however, ignores the low-rank property of the tensor subspaces along its modes. An extreme example is that when the core tensor obtained from a CP transformation on a tensor is approximately diagonal, the subspace along each tensor mode induced from this decomposition will not be low-rank, although the core tensor is very sparse. This deviates most real scenarios that the data representation along a meaningful factor should always has an evident correlation and thus a low-rank structure. Such a useful knowledge, however, cannot be well expressed by CP decomposition. By integrating rational sparsity understanding elements from both decomposition forms, we attempt to construct the following quantity, which we call intrinsic tensor sparsity (ITS) for convenience, for measuring the sparsity of a tensor $\mathcal{X}$:

$$S(\mathcal{X}) = t\,\|\mathcal{S}\|_0 + (1-t)\prod_{i=1}^{N} \text{rank}\left(X_{(i)}\right), \tag{7}$$

where $\mathcal{S}$ is the core tensor of $\mathcal{X}$ obtained from the Tucker decomposition, $0 < t < 1$ is a tradeoff parameter to compromising the role of two terms.

Note that the new ITS measure takes both sparsity insights of Tucker and CP decompositions into consideration. Its first term constrains the number of Kronecker bases for representing the objective tensor, complying with intrinsic mechanism of the CP decomposition. The second term is physically interpreted as the size of the core tensor in Tucker decomposition. It inclines to regularize the low-rank property of the subspace spanned upon each tensor mode. Such integrative consideration in the new measure facilitates a tensor with both inner sparsity configurations in the core tensor and low-rank property of the tensor subspace along each tensor mode, and thus is hopeful to alleviate the limitations in both Tucker and CP decompositions as aforementioned.

It should be noted that very recently, Zhao et al. [34] also proposed a tensor sparsity measure by only considering the second (rank-product) term of (7). Such a measure can only provide a coarse regularization for rectifying the tensor sparsity (i.e., the block size of the core tensor) while

cannot finely rectify the fine-grained tensor sparsity inside the coefficient tensor. The neglection of the important CP sparsity element tends to evidently degenerate its MSI denoising performance, as verified by our experiments.

Apart from the above insight of the proposed ITS measure, its superiority also lies on the following two aspects as compared with the conventional tensor sparsity measures.

On one hand, it has a natural physical interpretation. As shown in Fig. 2, when the rank of a d-order tensor along its i^{th} mode is r_i, the second term of the proposed tensor sparsity (7) not only corresponds to the low-rank sparsity of the subspace spanned upon each tensor mode, but also corresponds to a upper bound of the number of Kronecker bases for representing this tensor, and the first term further more accurately describes the intrinsic Kronecker basis number utilized for this tensor representation. This means that the new tensor sparsity quantity represents a reasonable proxy for measuring the capacity of tensor space, in which the entire tensor located, by taking Kronecker basis as the fundamental representation component.

On the other hand, it unifies the sparsity measures throughout vector to matrix. The sparsity of a vector is conventionally measured by the number of the bases (from a predefined codebook/dictionary) that can represent the vector as the linear combination of them. Since in vector case, a Kronecker basis is just a common vector, this measurement is just the number of Kronecker bases required to represent the vector, which fully complies with our proposed sparsity measure and its physical meaning. The sparsity of a matrix is conventionally assessed by its rank. Actually there are the following results: (1) if the ranks of a matrix along its two dimensions are r_1 and r_2, respectively, then $r_1 = r_2 = r$, implying the second term in (7) is r^2; (2) if the matrix is with rank r, then it can be represented as r Kronecker bases (each with form $\mathbf{u}\mathbf{v}^T$), implying the first term in (7) is r. This means that the ITS measure is also proportional to the conventional one in matrix cases.

4.3. Relaxation

Note that the l_0 and rank terms in (7) can only take discrete values, and lead to combinatorial optimization problem in applications which is hard to solve. We thus relax the ITS as a log-sum form to ease computation. Such relaxation has been substantiated as an effective strategy to solve l_0 [4, 26] or rank minimization [17, 12] problems.

Instead of solving (3), our aim is then changed to solving the following optimization problem:

$$\min_{\mathcal{X}} P_{ls}(\mathcal{S}) + \lambda \prod_{j=1}^{3} P_{ls}^{*}\left(X_{(j)}\right) + \frac{\beta}{2}\left\|\mathcal{Y}_i - \mathcal{X}\right\|_F, \quad (8)$$

where $\lambda = \frac{1-t}{t}$,$\beta = \frac{\gamma}{t}$, and

$$P_{ls}(\mathcal{A}) = \sum\nolimits_{i_1,i_2,i_3} \log(|a_{i_1,i_2,i_3}| + \varepsilon),$$

$$P_{ls}^{*}(A) = \sum\nolimits_{j} \log(\sigma_j(A) + \varepsilon),$$

which are the log-sum forms of the vector and matrix sparsities, respectively. ε is a small positive number, and $\sigma_j(A)$ defines the j^{th} singular of $A \in \mathbb{R}^{m\times n}$. An efficient algorithm is then proposed in the following section to solve the problem.

4.4. ADMM algorithm

We apply the alternating direction method of multipliers (ADMM) [2, 15], an effective strategy for solving large scale optimization problems, to solving (8). Firstly, we need to introduce 3 auxiliary tensors $\mathcal{M}_j$ $(j = 1, 2, 3)$ and equivalently reformulate (8) as follows:

$$\min_{\mathcal{S},\mathcal{M}_j,U_j} P_{ls}(\mathcal{S}) + \lambda \prod_{j=1}^{3} P_{ls}^{*}\left(M_{j(j)}\right) + \frac{\beta}{2}\|\mathcal{Y}_i - \mathcal{S}\times_1 U_1 \times_2 U_2 \times_3 U_3\|_F^2$$

$$\text{s.t. } \mathcal{S}\times_1 U_1 \times_2 U_2 \times_3 U_3 - \mathcal{M}_j = 0,\ U_j^T U_j = I,\ \ j = 1, 2, 3,$$

where $M_{j(j)} = \text{unfold}_j(\mathcal{M}_j)$. Then its augmented Lagrangian function is with the form:

$$\begin{aligned} L_\mu(\mathcal{S}, \mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3, U_1, U_2, U_3) = {} & P_{ls}(\mathcal{S}) + \lambda \textstyle\prod_{j=1}^{3} P_{ls}^{*}\left(M_{j(j)}\right) \\ & + \frac{\beta}{2}\left\|\mathcal{Y}_i - \mathcal{S}\times_1 U_1 \times_2 U_2 \times_3 U_3\right\|_F^2 \\ & + \textstyle\sum_{j=1}^{3} \langle \mathcal{S}\times_1 U_1 \times_2 U_2 \times_3 U_3 - \mathcal{M}_j, \mathcal{P}_j\rangle \\ & + \textstyle\sum_{j=1}^{3} \frac{\mu}{2}\left\|\mathcal{S}\times_1 U_1 \times_2 U_2 \times_3 U_3 - \mathcal{M}_j\right\|_F^2, \end{aligned}$$

where $\mathcal{P}_j$s are the Lagrange multipliers, μ is a positive scalar and U_j must satisfy $U_j^T U_j = I, \forall j = 1, 2, 3$. Now we can solve the problem within the ADMM framework.

With other parameters fixed, $\mathcal{S}$ can be updated by solving $\min_{\mathcal{S}} L_\mu(\mathcal{S}, \mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3, U_1, U_2, U_3)$, which is equivalent to the following sub-problem:

$$\min_{\mathcal{S}} bP_{ls}(\mathcal{S}) + \frac{1}{2}\left\|\mathcal{S}\times_1 U_1 \times_2 U_2 \times_3 U_3 - \mathcal{O}\right\|_F^2, \quad (9)$$

where $b = \frac{1}{\beta+3\mu}$ and $\mathcal{O} = \frac{\beta\mathcal{Y}_i + \sum_j(\mu\mathcal{M}_j - \mathcal{P}_j)}{\beta+3\mu}$. Since

$$\|\mathcal{D}\times V\|_F^2 = \|\mathcal{D}\|_F^2,\ \forall\, V^T V = I, \quad (10)$$

by mode-j producting U_j^T on each mode, Eq. (9) turns to the following problem:

$$\min_{\mathcal{S}} bP_{ls}(\mathcal{S}) + \frac{1}{2}\left\|\mathcal{S} - \mathcal{Q}\right\|_F^2, \quad (11)$$

where $\mathcal{Q} = \mathcal{O}\times_1 U_1^T \times_2 U_2^T \times_3 U_3^T$, which has been proved to have a closed-form solution [11]:

$$\mathcal{S}^{+} = \mathrm{D}_{b,\varepsilon}(\mathcal{Q}). \quad (12)$$

Here, $\mathrm{D}_{b,\varepsilon}(\cdot)$ is the thresholding operator defined as:

$$\mathrm{D}_{b,\varepsilon}(x) = \begin{cases} 0 & \text{if} \quad c_2 \le 0 \\ \text{sign}(x)\left(\frac{c_1 + \sqrt{c_2}}{2}\right) & \text{if} \quad c_2 > 0 \end{cases} \quad (13)$$

with that $c_1 = |x| - \varepsilon$, $c_2 = (c_1)^2 - 4(b - \varepsilon|x|)$.

When updating U_j $(1 \leq j \leq 3)$ with other parameters fixed, we can also deduce its closed-form solution. Let's take U_1 as an example. With U_2, U_3 and other parameters fixed, we update U_1 by solving $\min_{U_1} L_\mu(\mathcal{S}, \mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3, U_1, U_2, U_3)$, which is equivalent to the following problem:

$$\min_{U_1} \|\mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 U_3 - \mathcal{O}\|_F^2. \tag{14}$$

By employing (10) and the following equation

$$\mathcal{B} = \mathcal{D} \times_n V \iff B_{(n)} = VD_{(n)}, \tag{15}$$

we can obtain that (14) is equivalent to

$$\max_{U_1^T U_1 = I} \langle A_1, U_1 \rangle, \tag{16}$$

where $A_1 = \left(O_{(1)}\text{unfold}_1(\mathcal{S} \times_2 U_2 \times_3 U_3)\right)$. It can be easily seen that U_2 and U_3 can be updated by solving

$$\max_{U_k^T U_k = I} \langle A_k, U_k \rangle. \tag{17}$$

We can use the following theorem to obtain the closed-from solution of (17).

Theorem 1. *$\forall\, A \in \mathbb{R}^{m \times n}$, the following problem:*

$$\max_{U^T U = I} \langle A, U \rangle, \tag{18}$$

has the closed-form solution $\hat{U} = BC^T$, where $A = BDC^T$ is the SVD decomposition of A.

The proof of Theorem 1 can be found in supplementary material. We can now update U_k by the following formula:

$$U_k^+ = B_k {C_k}^T. \tag{19}$$

where $A_k = B_k D C_k^T$ is the SVD decomposition of A_k.

With $\mathcal{M}_j (j \neq k)$ and other parameters fixed, $\mathcal{M}_k$ can be updated by solving the following problem:

$$\min_{\mathcal{M}_k} a_k P_{ls}^* \left(M_{k(k)}\right) + \frac{1}{2}\|\mathcal{L} + \frac{1}{\mu}\mathcal{P}_k - \mathcal{M}_k\|_F^2, \tag{20}$$

where $a_k = \left(\frac{\lambda}{\mu} \prod_{j \neq k} P_{ls}^* \left(M_{j(j)}\right)\right)$ and $\mathcal{L} = \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 U_3$. This sub-problem can be easily solved by virtue of the following theorem:

Theorem 2. *Given $Y \in \mathbb{R}^{m \times n}$, $m \geq n$, let $Y = U\text{diag}(\sigma_1, \sigma_2, ..., \sigma_n)V^T$ be the SVD of Y. Let $0 < \lambda$, $0 < \varepsilon < \min\left(\sqrt{\lambda}, \frac{\lambda}{\sigma_1}\right)$, and define d_i as the i-th singular value of X, the solution to the following problem:*

$$\min_{X \in \mathbb{R}^{m \times n}} \lambda \sum\nolimits_{i=1}^{n} \log(d_i + \varepsilon) + \frac{1}{2}\|X - Y\|_F^2 \tag{21}$$

can be expressed as $\hat{X} = U\text{diag}(\hat{d}_1, \hat{d}_2, ..., \hat{d}_n)V^T$, where $\hat{d}_i = \text{D}_{\lambda,\varepsilon}(\sigma_i), i = 1, 2, ..., n$.

The proof of Theorem 2 can be found in the supplementary material. We can now update $\mathcal{M}_k$ by following equation:

$$\mathcal{M}_k^+ = \text{fold}_k\left(V_1 \Sigma_{a_k} V_2^T\right), \tag{22}$$

where $\Sigma_{a_k} = \text{diag}\left(\text{D}_{a_k,\varepsilon}(\sigma_1), \text{D}_{a_k,\varepsilon}(\sigma_2), \cdots, \text{D}_{a_k,\varepsilon}(\sigma_n)\right)$ and $V_1 \text{diag}(\sigma_1, \sigma_2, ..., \sigma_n) V_2^T$ is the SVD of $\text{unfold}_k\left(\mathcal{L} + \frac{1}{\mu}\mathcal{P}_k\right)$.

The proposed algorithm for MSI denoising can be summarized in Algorithm 1, and we denote the proposed method as ITSReg (Intrinsic Tensor Sparsity Regularization) for convenience.

Algorithm 1 Algorithm for MSI Denoising

Input: Noisy MSI $\mathcal{Y}$

Initialize $\mathcal{X}^{(0)} = \mathcal{Y}$
for $l = 1 : L$ **do**
  Calculate $\mathcal{Y}^{(l)} = \mathcal{X}^{(l-1)} + \delta\left(\mathcal{Y} - \mathcal{X}^{(l-1)}\right)$
  Construct the entire FBP set $\Omega_{\mathcal{Y}^{(l)}}$
  Construct the set of similar FBP group set $\{\mathcal{Y}_i\}_{i=1}^K$
  for each FBP groups $\mathcal{Y}_i$ **do**
    //Solve the problem (8) by ADMM
    while not convergence **do**
      Update $\mathcal{S}$ by (12)
      Update U_k by (19), $\forall k = 1, 2, 3$
      Update $\mathcal{M}_k$ by (22), $\forall k = 1, 2, 3$
      Update the multipliers and let $\mu := \rho\mu$
    end while
    Calculate $\mathcal{X}_i = \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 U_3$
  end for
  Aggregate $\{\mathcal{X}_i\}_{i=1}^K$ to form the clean image $\mathcal{X}^{(l)}$
end for

Output: Denoised MSI $\mathcal{X}^{(L)}$

5. Experimental results

To validate the effectiveness of the proposed method for MSI denoising, we perform both simulated and real data experiments and compare the experimental results both quantitatively and visually.

5.1. Simulated MSI denoising

Columbia Datasets. The Columbia MSI Database [32][1] is utilized in our simulated experiment. This MSI data set contains 32 real-world scenes of a wide variety of real-world materials and objects, each with spatial resolution 512×512 and spectral resolution 31, which includes full spectral resolution reflectance data collected from 400nm to 700nm in 10nm steps. In our experiments, each of these MSIs is scaled into the interval $[0, 1]$.

Implementation details. Additive Gaussian noises with variance v are added to these testing MSIs to generate the noisy observations with v ranging from 0.1 to 0.3. There are two parameters λ and β in our model. The former λ is used to balance two parts in the same order of magnitude, and we just simply set $\lambda = 10$ in all our experiments, β is dependent on v, and we let $\beta = cv$, where c is set as the constant 10^{-3}. More clarifications on such parameter settings are provided in the supplementary material.

[1] http://www1.cs.columbia.edu/CAVE/databases/multispectral

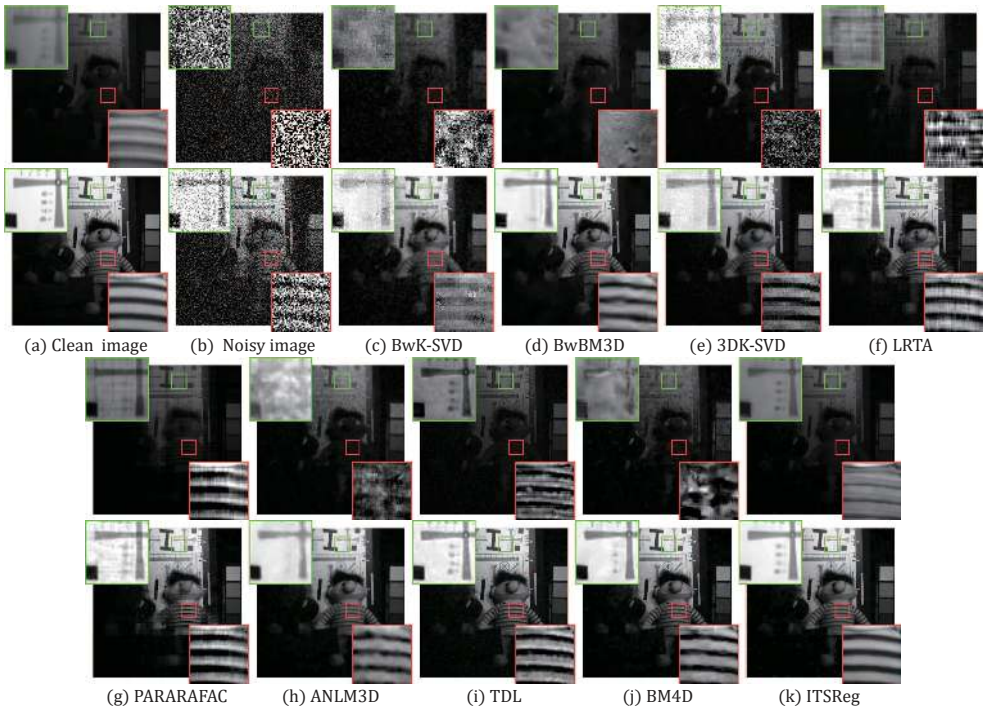


Figure 3. (a) The image at two bands (400nm and 700nm) of chart and staffed toy; (b) The corresponding images corrupted by Gaussian noise with variance $v = 0.2$, (c)-(k) The restored image obtained by the 9 utilized MSI denoising methods. Two demarcated areas in each image are amplified at a 4 times larger scale and the same degree of contrast for easy observation of details.

Table 1. Average performance of 9 competing methods with respect to 4 PQIs. For both settings, the results are obtained by averaging through the 32 scenes and the varied parameters.

	PSNR	SSIM	FSIM	ERGAS
$v = 0.1, 0.15, 0.2, 0.25, 0.3$				
Nosiy image	14.59±3.38	0.06±0.05	0.47±0.15	1151.54±534.17
BwK-SVD	27.77±2.01	0.47±0.10	0.81±0.06	234.58±66.73
BwBM3D	34.00±3.39	0.86±0.06	0.92±0.03	116.91±42.76
3DK-SVD	30.31±2.23	0.69±0.06	0.89±0.03	176.58±43.78
LRTA	33.78±3.37	0.82±0.09	0.92±0.03	120.79±46.06
PARAFAC	31.35±3.42	0.72±0.12	0.89±0.04	160.66±66.95
ANLM3D	34.12±3.19	0.86±0.07	0.93±0.03	117.01±35.79
TDL	35.71±3.09	0.87±0.07	0.93±0.04	96.21±34.36
BM4D	36.18±3.03	0.86±0.07	0.94±0.03	91.20±29.70
ITSReg	**37.78±3.32**	**0.90±0.07**	**0.96±0.02**	**77.35±30.16**

Comparison methods. The comparison methods include: band-wise K-SVD [1][2] and band-wise BM3D [8][3], representing state-of-the-arts for the 2D extended band-wise approach; 3D-cube K-SVD [10][4], ANLM3D [20][5] and BM4D [18, 19][6], representing state-of-the-arts for the 2D extended 3D-cube-based approach; LRTA [21], PARAFAC[16][7] and TDL[23][8] representing state-of-the-arts for the tensor-based approach. All parameters involved in the competing algorithms were optimally assigned or automatically chosen as described in the reference papers.

Evaluation measures. Four quantitative picture quality indices (PQI) are employed for performance evaluation, including peak signal-to-noise ratio (PSNR), structure similarity (SSIM [30]), feature similarity (FSIM [33]), erreur relative globale adimensionnelle de synthèse (ERGAS [28]). PSNR and SSIM are two conventional PQIs in image processing and computer vision. They evaluate the similarity between the target image and the reference image based on MSE and structural consistency, respectively. FSIM emphasizes the perceptual consistency with the reference image. The larger these three measures are, the closer the target MSI is to the reference one. ERGAS measures fidelity of the restored image based on the weighted sum of MSE in each band. Different from the former three measures, the smaller ERGAS is, the better does the target MSI estimate the reference one.

Performance evaluation. For each noise setting, all of the four PQI values for each competing MSI denoising methods on all 32 scenes have been calculated and recorded. Table 1 lists the average performance (over different scenes and noise settings) of all methods. From these quantitative comparison, the advantage of the proposed method can be evidently observed. Specifically, our method can significantly outperform other competing methods with respective to all evaluation measures, e.g., the PSNR obtained by our

[2] http://www.cs.technion.ac.il/~elad/software
[3] http://www.cs.tut.fi/~foi/GCF-BM3D/
[4] http://www.cs.technion.ac.il/~elad/software
[5] http://personales.upv.es/jmanjon/denoising/arnlm.html
[6] http://www.cs.tut.fi/~foi/GCF-BM3D/
[7] http://www.sandia.gov/tgkolda/TensorToolbox/index-2.5.html
[8] http://gr.xjtu.edu.cn/web/dymeng/2

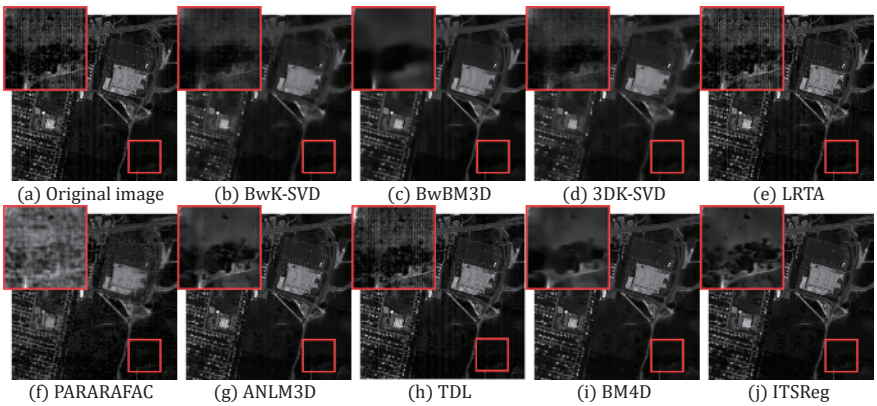


Figure 4. (a) The Original image located at the 1^{st} band in HYDICE urban data set; (b)-(j) The restored image obtained by the 9 utilized MSI denoising methods. The demarcated areas in each image are amplified at a 2.5 times larger scale and the same degree of contrast.

method are more than 1.5 larger and ERGAS is around 10 smaller than the second best methods. More details are listed in our supplementary material.

To further depict the denoising performance of our method, we show in Fig. 3 two bands in *chart and stuffed toy* that centered at 400nm (the darker one) and 700nm (the brighter one), respectively. From the figure, it is easy to observe that the proposed method evidently performs better than other competing ones, both in the recovery of finer-grained textures and coarser-grained structures. Especially, when the band energy is low, most competing methods begin to fail, while our method still performs consistently well in such harder cases.

5.2. Real MSI denoising

In this section, urban area HYDICE MSI of natural scenes[9] is used in our experiments. The original MSI is of the size $304 \times 304 \times 210$. As the bands 76, 100-115, 130-155 and 201-210 are seriously polluted by the atmosphere and water absorption and can provide little useful information, we remove them and leave the remaining test data with a size of $304 \times 304 \times 157$.

We scale the MSI into the interval [0, 1], and employed the similar implementation strategies and parameter settings for all competing methods as previous simulated experiments. Since the noise level is unknown for real noisy images, we use an off-the-shelf noise estimation method [9] to estimate it.

We illustrate the experimental results at the first band of the test MSI in Fig. 4. It can be easily observed that the image restored by our method is capable of properly removing the stripes and Gaussian noise while finely preserving the structure underlying the MSI. BM3D and BM4D can perform comparatively better in stripes noise removing, but their results contain evident blurry area , where our method finely recovers the details hiding under the corrupted MSI.

[9] http://www.tec.army.mil/Hypercube

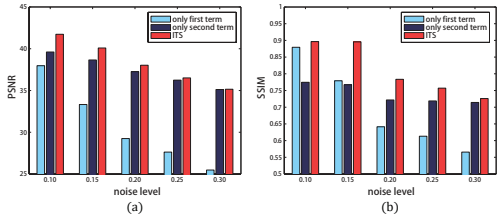


Figure 5. PSNR (a) and SSIM (b) of resorted MSI obtained by the 3 sparsity measures over different noise levels.

5.3. Analysis of the new tensor sparsity measure

Since our the proposed ITS measure is composed of two terms, we give some experimental analysis to verify if both terms could contribute to the denoising performance. To this aim, we perform MSI denoising within the proposed framework, while using sparsity measure $S(\mathcal{X})$ containing only its first term, second term (the work of Zhao et al. [34] corresponds to only considering the second term) and both terms. Fig. 5 shows the denoising results on *face* MSI from Columbia Dataset, in terms of PSNR and SSIM, with different noise levels (similar tendency can also be observed on other MSIs and with other evaluation measures). It is easy to see that the proposed ITS benefits from both of its terms. Only considering either one tends to degenerate the performance of MSI denoising.

6. Conclusion

In this paper, we have provided a new MSI denoising model under a newly designed sparsity measure which finely encodes the correlation insights under the known Tucker and CP decomponstion for tensors. We have also designed an efficient ADMM algorithm to solve the model. The experiments on simulated and real MSI denoising have substantiated the superiority of the proposed method beyond state-of-the-arts.

Acknowledgment. This research was supported by 973 Program of China with No. 3202013CB329404, and the NSFC projects with No. 61373114, 91330204, 11131006, 61503263.

References

[1] M. Aharon. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Processing*, 54(11):4311–4322, 2006.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations & Trends? in Machine Learning*, 3(1):1–122, 2011.

[3] A. Buades, B. Coll, and J. M. Morel. A non-local algorithm for image denoising. In *CVPR*, 2005.

[4] E. J. Candes, M. B. Wakin, and S. P. Boyd. Enhancing sparsity by reweighted l_1 minimization. *Journal of Fourier analysis and applications*, 14(5-6):877–905, 2008.

[5] W. Cao, Y. Wang, C. Yang, X. Chang, Z. Han, and Z. Xu. Folded-concave penalization approaches to tensor completion. *Neurocomputing*, 152:261–273, 2015.

[6] J. D. Carroll and J.-J. Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of eckart-young decomposition. *Psychometrika*, 35(3):283–319, 1970.

[7] A. A. Chen. The inpainting of hyperspectral images: A survey and adaptation to hyperspectral data. In *SPIE*, 2012.

[8] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Trans. Image Processing*, 16(8):2080–2095, 2007.

[9] D. L. Donoho. De-noising by soft-thresholding. *IEEE Trans. Information Theory*, 41(3):613–627, 1995.

[10] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Processing*, 15(12):3736–45, 2006.

[11] P. Gong, C. Zhang, Z. Lu, J. Z. Huang, and J. Ye. A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems. In *ICML*, 2013.

[12] S. Gu, L. Zhang, W. Zuo, and X. Feng. Weighted nuclear norm minimization with application to image denoising. In *CVPR*, 2014.

[13] R. Kawakami, J. Wright, Y.-W. Tai, Y. Matsushita, M. Ben-Ezra, and K. Ikeuchi. High-resolution hyperspectral imaging via matrix factorization. In *CVPR*, 2011.

[14] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009.

[15] Z. Lin, R. Liu, and Z. Su. Linearized alternating direction method with adaptive penalty for low-rank representation. *Advances in Neural Information Processing Systems*, pages 612–620, 2011.

[16] X. Liu, S. Bourennane, and C. Fossati. Denoising of hyperspectral images using the parafac model and statistical performance analysis. *IEEE Trans. Geoscience and Remote Sensing*, 50(10):3717–3724, 2012.

[17] C. Lu, C. Zhu, C. Xu, S. Yan, and Z. Lin. Generalized singular value thresholding. In *AAAI*, 2015.

[18] M. Maggioni and A. Foi. Nonlocal transform-domain denoising of volumetric data with groupwise adaptive variance estimation. In *SPIE*, 2012.

[19] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi. A nonlocal transform-domain filter for volumetric data denoising and reconstruction. *IEEE Trans. Image Processing*, 22(1):119 – 133, 2012.

[20] J. V. Manjón, P. Coupé, L. Martí-Bonmatí, D. L. Collins, and M. Robles. Adaptive non-local means denoising of mr images with spatially varying noise levels. *Journal of Magnetic Resonance Imaging*, 31(1):192–203, 2010.

[21] S. B. N. Renard and J. Blanc-Talon. Denoising and dimensionality reduction using multilinear tools for hyperspectral images. *IEEE Trans. Geoscience and Remote Sensing*, 5(2):138–142, 2008.

[22] H. V. Nguyen, A. Banerjee, and R. Chellappa. Tracking via object reflectance using a hyperspectral video camera. In *CVPR Workshop*, 2010.

[23] Y. Peng, D. Meng, Z. Xu, C. Gao, Y. Yang, and B. Zhang. Decomposable nonlocal tensor dictionary learning for multispectral image denoising. In *CVPR*, 2014.

[24] Y. Qian, Y. Shen, M. Ye, and Q. Wang. 3-D nonlocal means filter with noise estimation for hyperspectral imagery denoising. In *IGARSS*, 2012.

[25] B. Romera-Paredes and M. Pontil. A new convex relaxation for tensor completion. In *NIPS*, 2013.

[26] O. Taheri, S. Vorobyov, et al. Sparse channel estimation with l p-norm and reweighted l 1-norm penalized least mean squares. In *ICASSP*, 2011.

[27] L. R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, 1966.

[28] L. Wald. *Data Fusion: Definitions and Architectures: Fusion of Images of Different Spatial Resolutions*. Presses des lEcole MINES, 2002.

[29] H. Wang, F. Nie, and H. Huang. Low-rank tensor completion with spatio-temporal consistency. In *AAAI*, 2014.

[30] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Processing*, 13(4):600–612, 2004.

[31] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In *NIPS*, 2009.

[32] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar. Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum. *IEEE Trans. Image Processing*, 19(9):2241–2253, 2010.

[33] L. Zhang, L. Zhang, X. Mou, and D. Zhang. Fsim: a feature similarity index for image quality assessment. *IEEE Trans. Image Processing*, 20(8):2378–2386, 2011.

[34] Q. Zhao, D. Meng, X. Kong, Q. Xie, W. Cao, Y. Wang, and Z. Xu. A novel sparsity measure for tensor recovery. In *ICCV*, 2015.

[35] X. Zhao, F. Wang, T. Huang, M. K. Ng, and R. J. Plemmons. Deblurring and sparse unmixing for hyperspectral images. *IEEE Trans. Geoscience and Remote Sensing*, 51(7):4045–4058, 2013.