Kronecker-Basis-Representation Based Tensor Sparsity and Its Applications to Tensor Recovery

Qi Xie, Qian Zhao, Deyu Meng, and Zongben Xu

Abstract—As a promising way for analyzing data, sparse modeling has achieved great success throughout science and engineering. It is well known that the sparsity/low-rank of a vector/matrix can be rationally measured by nonzero-entries-number (l_0 norm)/nonzero-singular-values-number (rank), respectively. However, data from real applications are often generated by the interaction of multiple factors, which obviously cannot be sufficiently represented by a vector/matrix, while a high order tensor is expected to provide more faithful representation to deliver the intrinsic structure underlying such data ensembles. Unlike the vector/matrix case, constructing a rational high order sparsity measure for tensor is a relatively harder task. To this aim, in this paper we propose a measure for tensor sparsity, called Kronecker-basis-representation based tensor sparsity measure (KBR briefly), which encodes both sparsity insights delivered by Tucker and CANDECOMP/PARAFAC (CP) low-rank decompositions for a general tensor. Then we study the KBR regularization minimization (KBRM) problem, and design an effective ADMM algorithm for solving it, where each involved parameter can be updated with closed-form equations. Such an efficient solver makes it possible to extend KBR to various tasks like tensor completion and tensor robust principal component analysis. A series of experiments, including multispectral image (MSI) denoising, MSI completion and background subtraction, substantiate the superiority of the proposed methods beyond state-of-the-arts.

Index Terms—Tensor sparsity, tucker decomposition, CANDECOMP/PARAFAC decomposition, tensor completion, multi-spectral image restoration

1 Introduction

Sparsity is a common information representation property which means that an observation can be represented by a few atoms of an appropriately chosen dictionary. So far sparsity-based methods have achieved great success throughout science and engineering. Typical examples include face modeling [47], [60], gene categorization [54], image/video compressive sensing [3], [10], user interest prediction [20], signal restoration [47], [68], [75], etc.

It is well known that the sparsity/low-rankness of a vector/matrix can be rationally measured by the number of nonzero entries/nonzero singular values (l_0 norm/rank). Such sparsity measures, as well as their relaxation forms (e.g., l_1 norm and nuclear norm), have been shown to be helpful to finely encode the data sparsity in applications, and has inspired various sparse/low-rank models and algorithms against different practical problem [6], [9], [14], [26], [47], [50], [68], [75].

However, data from many real applications are often generated by the interaction of multiple factors. The traditional vector or matrix, which can only well address single/binary-factor variability of data, obviously is not the best way to keep the multi-factor structure of this kind of data. For example, a multispectral image (MSI) consists of a collection of images scattered over various discrete bands and thus includes three intrinsic constituent factors, i.e., spectrum and spatial width and height. Expressing MSI as a matrix will inevitably damage its three-factor structure with only two factors considered. Instead of vector or matrix, a higher-order tensor, represented as a multidimensional array, provides a more faithful representation to deliver the intrinsic structure underlying such data ensembles. The techniques on tensors have thus been attracting much attention recently and helped enhance performance of various practical tasks, such as MSI denoising [69], 3D image reconstruction [59], and higher-order web link analysis [32].

A tensor collected from real scenarios are always with an evident correlation along each of its modes. By taking the MSI shown in Fig. 1 as an example, the correlation along each of its spectral and spatial modes can be evidently observed both quantitatively and visually. This reflects the fact that the tensor along each mode resides on a low-rank subspace and the entire tensor corresponds to the affiliation of the subspaces along all tensor modes. Thus, in order to faithfully deliver such sparsity knowledge underlying tensor and enhance the performance of sparsity-based tensor recovery techniques, it is always crucial to consider a quantitative measure for assessing tensor sparsity.

Mathematically, a sparsity-based tensor recovery model can generally be expressed as follows:

$$\min_{\mathcal{X}} S(\mathcal{X}) + \gamma L(\mathcal{X}, \mathcal{Y}), \tag{1}$$

where $\mathcal{Y} \in R^{I_1 \times I_2 \times \cdots \times I_N}$ is the observation, $L(\mathcal{X}, \mathcal{Y})$ is the loss function between $\mathcal{X}$ and $\mathcal{Y}$, $S(\mathcal{X})$ defines the tensor sparsity

• *The authors are with the School of Mathematics and Statistics and Ministry of Education Key Lab of Intelligent Networks and Network Security, Xi'an Jiaotong University, Shaanxi 710049, P.R. China. E-mail: xq.liwu@stu.xjtu.edu.cn, {timmy.zhaoqian, dymeng, zbxu}@mail.xjtu.edu.cn.*

Manuscript received 18 July 2016; revised 24 June 2017; accepted 12 July 2017. Date of publication 1 Aug. 2017; date of current version 11 July 2018. (Corresponding author: Deyu Meng.)
Recommended for acceptance by R. Vidal.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TPAMI.2017.2734888

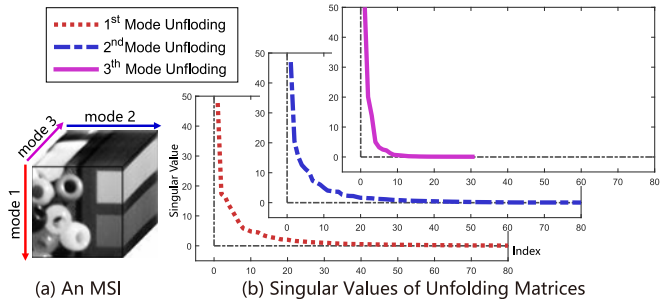

Fig. 1. Illustration of correlation priors on each mode of an MSI. (a) A real MSI of size $80 \times 80 \times 30$. (b) Singular value curves of the matrices unfolded along three tensor modes. The low-rank property of the subspace along each mode can be easily observed from the dramatic decreasing effect of these curves.

measure of $\mathcal{X}$ and γ is the compromise parameter. It is easy to see that the key problem in constructing (1) is to design an appropriate tensor sparsity measure on data. While not like vector/matrix cases, extracting a rational high-order sparsity measure for tensor is always a much harder task.

Most current works directly extended the rank of matrix to higher-order by simply summing up ranks (or its relaxations) along all tensor modes [11], [24], [40], [57]. However, different from matrix scenarios, this simple rank-summation term is generally short of a clear physical meaning for tensors. Specifically, tensor sparsity insight should be interpreted beyond the low-rank property of all its unfolded subspaces, and should more importantly consider how such subspace sparsities are affiliated over the entire tensor structure. Besides, such a simple tensor sparsity measure more or less lacks a consistent relationship with previous sparsity definitions for vectors and matrices.

To handle the aforementioned issues, this paper mainly makes three-fold contributions. First, a tensor sparsity measure is proposed, and its insight can be easily interpreted as a regularization for the number of rank-1 Kronecker bases for representing the tensor. We thus call it Kronecker-basis-representation based tensor sparsity measure (KBR) for convenience. 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 decompositions for a general tensor.

Second, we design an effective alternating direction method of multipliers (ADMM) algorithm [4], [38] for solving the KBR-based minimization (KBRM) problem (1). We further deduce the closed-form equations for updating each involved parameter, which makes the algorithm be able to be efficiently implemented. This solver facilitates a general utilization of KBR regularization to more tensor analysis tasks. Specifically, we extend KBRM to two of the most typical tensor recovery problems and propose the KBR-based tensor completion (KBR-TC) model and KBR-based robust principal component analysis (KBR-RPCA) model. Similar ADMM algorithms have been designed to solve the models by using the proposed KBRM solver.

Third, we adapt the proposed KBR-based models to multiple MSI/video tasks. In MSI denoising, we propose a new tensor-based denoising model by exploiting the proposed KBR measure to encode the inherent spatial nonlocal similarity and spectral correlation in MSI, and perform KBRM to recover the clean MSI. The model is with a concise formulation and can be easily extended to solving other MSI recovery problems. Experiments on benchmark and real MSI data show that the proposed method achieves the state-of-the-art performance on MSI denoising among comprehensive quality assessments. In MSI completion and background subtraction experiments, the proposed KBR-TC and KBR-RPCA models also achieve better performance than traditional algorithms.

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

We first introduce some necessary notions and preliminaries as follows.

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 $\mathcal{A}$. Conversely, the unfolding matrices along the nth mode can be transformed back to the tensor by $\mathcal{A} = \text{fold}_n\big(A_{(n)}\big), 1 \leq n \leq N$. 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 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 $\mathcal{C} \in \mathbb{R}^{I_1 \times \cdots \times J_n \times \cdots I_N}$, with entries

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

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 retensorization of undoing the mode-n flattening. For convenience, we define $\mathcal{A} \bar{\times}_{-n} \{B_j\}_{j=1}^N$ as

$$\mathcal{A} \times_1 B_1 \times_2 \cdots \times_{n-1} B_{n-1} \times_{n+1} B_{n+1} \cdots \times_N B_N.$$

The Frobenius norm of an tensor $\mathcal{A}$ is $\|\mathcal{A}\|_F = \big(\sum_{i_1, \cdots i_N} |a_{i_1, \cdots i_N}|^2\big)^{1/2}$.

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

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

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

3 Related Works on Tensor Sparsity

In this section, we first review some sparsity-based tensor recovery methods proposed in previous literatures, and

then briefly review two particular tensor decompositions, both containing insightful understanding of tensor sparsity.

3.1 Previous Work on Tensor Recovery

The simplest way to deal with tensor recovery problem is unfolding the objective tensor in a reasonable mode to transform it to a 2-order matrix and then perform matrix recovery on it. Most of the matrix recovery methods are constructed based on low rank matrix approximation (LRMA), which aims to recover the underlying low rank matrix from its degraded observation. The LRMA problem can be categorized as: the low rank matrix factorization (LRMF), aiming to factorize the objective matrix into two flat matrices, and rank minimization, aiming to reconstruct the data matrix through imposing an additional rank constraint upon the estimated matrix. Both of these two categories have a variety of extensions as proposed in [5], [6], [19], [30], [61]. Furthermore, there are many matrix recovery methods constructed by adding more prior knowledge related to data cases together with matrix low-rankness [16], [22], such as joint trace/TV based MSI recovery [23], [27] and joint low-rank/sparsity based matrix recovery [55], [56], etc.

Matrix completion (MC) and matrix-based robust principal component analysis (RPCA) are two of the most typical matrix recovery problems. The MC problem has been recently arising in multiple applications including collaborative filtering and latent semantic analysis [7], [19], [36]. Candès and Recht [8] prompted a new surge for this problem by showing that the matrix can be exactly recovered from an incomplete set of entries through solving a convex semidefinite programming. Various ameliorations have been further proposed for this task [28], [38]. The RPCA problem was initially formulated by Wright et al. [68] with the theoretical guarantee to exactly recover the ground truth tensor from grossly corrupted one under certain assumptions [6]. To speed up computation, Lin et al. proposed the accelerated proximal gradient (APG) [37] and the augmented Lagrangian multiplier (ALM) [36] methods. Bayesian RPCA approaches have also been investigated in [2], [15], [75].

Albeit easy for computation, unfolding a tensor to a matrix will inevitably destroy the intrinsic structure underlying a tensor. For example, unfolding an MSI along its spectral dimensionality will damage the spatial information of every bands in the MSI. Thus, recovering a tensor through imposing sparsity directly on tensors has being attracting increasing research interest in recent years. Different from matrix cases with natural sparsity measure (rank), the tensor case is more complicated and it is crucial to construct a rational sparsity measure to describe the intrinsic correlations across various tensor modes.

Corresponding to the MC and RPCA issues for matrix, investigations on both tensor completion (TC) and tensor-based RPCA (TRPCA) have been made in the recent decade. E.g., Liu et al. [39], [40] proposed a high-order tensor sparsity measure with the weighted sum of all unfolding matrix ranks (or relaxations) along all tensor modes, and applied it to implementing TC tasks. Goldfarb and Qin [24] further applied the similar sparsity measure to TRPCA problem. Romera-Paredes and Pontil [57] promoted this "sum of ranks" measure by relaxing it to a more tight convex form, and Cao et al. [11] further ameliorated it by using non-convex relaxation. Recently, Wang et al. [66] proposed a worst case measure, i.e., measuring the sparsity of a tensor by its largest rank of all unfolding matrices, and relaxed it with a sum of exponential forms. Designed mainly for videos, Zhang et al. [74] developed a tensor-SVD (t-SVD) based high-order sparsity measure for both TC and TRPCA problems. Very recently, Lu et al. [42] further proved the exactly-recover-property for t-SVD-based TRPCA method.

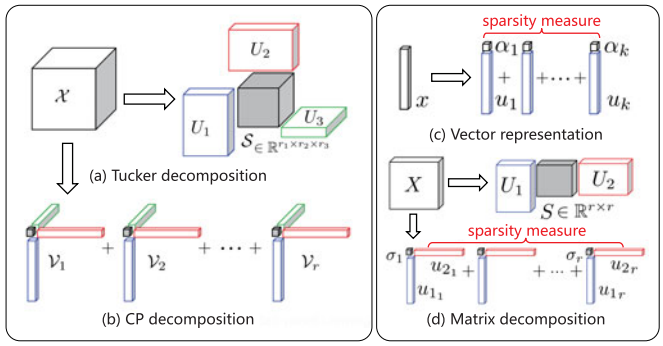

Fig. 2. Illustration of (a) Tucker decomposition and (b) CP decomposition of $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$. (c) Vector case representation. (d) Matrix case decomposition.

It is easy to see that most current high-order sparsity measures are constructed based on the weighted sum of ranks along all tensor modes. However, as we mentioned before, constructing tensor sparsity measure by this simple rank-summation term is short of a clear physical meaning for general tensors and lack of a consistent relationship with previous defined sparsity measures for vector/matrix. Besides, it is hard to decide the weight for penalizing each dimensionality ranks. A common way is using the same weights on all modes, but this is not always rational. Still taking the MSI data as an example, the rank of an MSI along its spectral dimensionality should be much lower than those along its spatial dimensionalities. We thus should more largely penalize the spectral rank than spatial ones, i.e., we should set a larger weight on the former while smaller on latter. Thus imposing the similar weights on ranks along all modes is more or less irrational in many real scenarios.

The KBR measure has been formally proposed in our previous work [69]. In this paper, we further extend the measure to TC and TRPCA problems. More comprehensive experiments on MSI completion and background subtraction have been presented to substantiate the superiority of the proposed method in wide range of applications.

3.2 Tucker Decomposition & CP Decomposition

Tucker and CP decompositions are the two most typically adopted tensor decomposition forms. Tucker decomposition decomposes a tensor as an affiliation of the orthogonal bases along all its modes integrated by a core coefficient tensor, and CP decomposes a tensor as a sum of rank-1 Kronecker bases, as visualized in Fig. 2.

Specifically, in Tucker decomposition [64], an N-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ can be written as the following form:

$$\mathcal{X} = \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 \cdots \times_N U_N, \tag{3}$$

where $\mathcal{S} \in \mathbb{R}^{R_1 \times \cdots \times R_N}$ $(r_i \leq R_i \leq I_i)$ is called the core tensor, and $U_i \in \mathbb{R}^{I_i \times R_i} (1 \leq i \leq N)$ is composed by the R_i orthogonal

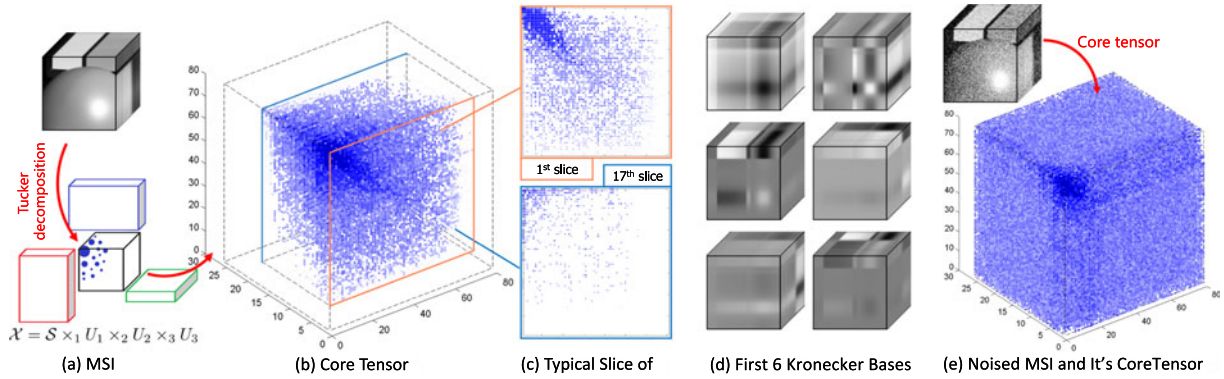

Fig. 3. (a) An MSI $\mathcal{X}_0 \in \mathbb{R}^{80\times80\times26}$ (upper) and it's Tucker decomposition. (b) Core tensor $\mathcal{S} \in \mathbb{R}^{80\times80\times26}$ of $\mathcal{X}$. Note that the size of the nonzero-block is $69 \times 71 \times 17$, and 78.4 percent of its elements are zeroes. (c) Typical slices of $\mathcal{S}$, where deeper color of the element represents a larger value of it. (d) 6 Kronecker bases of $\mathcal{X}$, which relate to the largest 6 elements in core tensor $\mathcal{S}$. (e) Noised MSI (lack of tensor sparsty) and it's core tensor (the size of the nonzero-block is $80 \times 80 \times 26$ and most of its elements are nonzero).

bases along the ith mode of $\mathcal{X}$. When the subspace bases along each mode is sorted based on their importance for tensor representation, the values of elements of the core tensor will show an approximately descending order along each of tensor modes. In this case, there will be a nonzero-block of size $r_1 \times r_2 \times \cdots \times r_N$, in the top left of the core tensor, as showed in Fig. 3, and the elements outside of this nonzero-block of core tensor will be all zeros. Under such a Tucker formulation, high-order sparsity can be quantified as the vector $(r_1, r_2, \ldots r_N)$, which is often called Tucker rank. Tucker rank considers the low-rank property of the vector subspace unfolded along each of its modes as a reflection of high-order sparsity. Such a sparsity understanding is equivalent to consider the size of nonzero-block in core tensor, which represents the coefficients affiliated from all combinations of the used subspace bases. However, Tucker decomposition delivers a freedom of $\prod_{i=1}^{N} r_i$ for assessing tensor sparsity by the volume of a core tensor, whereas sparsity within the core tensor often holds in natural data and restricts such freedom to a smaller number [58]. E.g., as depicted in Fig. 3, there can be many zero elements in the last several slices of the nonzero-block in the core tensor of a MSI, and such an elaborate information cannot be well described by the Tucker rank. The Tucker rank is thus hard to conduct a rational measure for comprehensively delivering the sparsity knowledge underlying a tensor.

CP decomposition attempts to decompose an N-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ into the linear combination of a series of Kroneker bases [31], written as

$$\mathcal{X} = \sum_{i=1}^{r} c_i \mathcal{V}_i = \sum_{i=1}^{r} c_i \mathbf{v}_{i_1} \circ \mathbf{v}_{i_2} \circ \cdots \circ \mathbf{v}_{i_N}, \tag{4}$$

where c_i denotes the coefficient imposed on the Kroneker basis $\mathcal{V}_i$. Under such a CP formulation, high-order sparsity can be easily quantified as r, which is often called CP rank. The superiority of CP rank lies on its consistence to the sparsity measure to that of a vector/matrix, which can be easily understood by seeing Fig. 2. However, CP rank cannot well encode the low-rank property of the tensor subspaces along its modes, since a low CP rank tensor can be full rank along all of its modes. In 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.

To ameliorate this issue, we attempt to propose a measure for more rationally assessing tensor sparsity.

4 KBR-Based Tensor Sparsity Measure

By employing the higher order singular value decomposition (HOSVD) [33], [53], which is a proxy of Tucker decomposition, the proposed KBR measure for a tensor $\mathcal{X}$ is of the following expression:

$$S(\mathcal{X}) = t\|\mathcal{S}\|_0 + (1-t)\prod_{i=1}^{N} \operatorname{rank}\left(X_{(i)}\right), \tag{5}$$

where $\mathcal{S} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ is the core tensor of $\mathcal{X}$ with HOSVD, and $0 < t < 1$ is a tradeoff parameter to compromise the role of the two terms.

The first term in (5) constrains the number of Kronecker bases for representing the target tensor, complying with intrinsic mechanism of the CP decomposition. This term has also been employed in many previous work to induce joint structure model on tensor decomposition, such as sparse non-negative tucker decomposition [58]. The second term can be interpreted as the volume of the nonzero-block of the core tensor. It inclines to regularize the low-rank property of the subspace spanned upon each tensor mode. Such integrative consideration in the proposed measure facilitates a tensor with both inner sparsity configurations of the core tensor and low-rank property of the tensor subspace along each mode, and thus is hopeful to alleviate the limitations in both Tucker and CP decompositions as aforementioned. Fig. 4 is a visual display of the KBR measure.

As compared with the conventional tensor sparsity measures, the proposed KBR is superior on that it has a natural physical interpretation. As shown in Fig. 3, when the rank of a

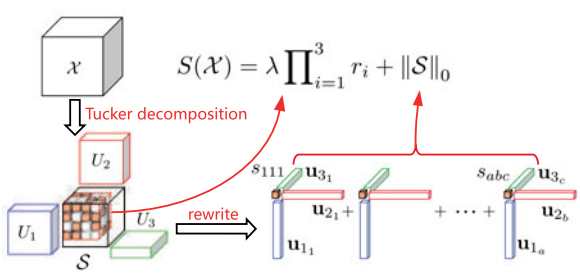

Fig. 4. A visual display of the KBR measure.

N-order tensor along its ith mode is r_i, the second term of the proposed tensor sparsity (5) corresponds to an upper bound of the number of Kronecker bases for representing this tensor, and the first term further finely rectifies the intrinsic Kronecker basis number utilized for this tensor representation. This means that the presented 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.

Besides, the proposed KBR unifies the traditional sparsity measures throughout 1-order to 2-order. The traditional 1-order sparsity of a vector is measured by the number of the bases (from a predefined 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 the Kronecker basis number required to represent the vector, which fully complies with our proposed sparsity measure and its physical meaning. The traditional 2-order sparsity of a matrix is assessed by its rank. If the rank of a matrix is r, then it implies that (i) the matrix can be represented as r Kronecker bases (each with form $\mathbf{u}\mathbf{v}^T$), and (ii) the number of dimensions of the subspace spanned upon its two modes are both r, implying the size of the nonzero-block in the core tensor of Tucker decomposition is $r \times r$. It's easy to see that (i) and (ii) is respectively consistent to the first and second term of KBR. The above analysis can be easily understood by observing Fig. 2.

4.1 Relaxation

Note that the l_0 and rank terms in (5) can only take discrete values, and lead to combinatorial optimization problem in applications which is hard to solve. We thus relax the KBR as a log-sum form to simplify computation. The effectiveness of such relaxation has been substantiated in previous research [9], [26], [43], [62].

We can then obtain the following relaxed form of KBR:

$$S^{\star}(\mathcal{X}) = tP_{ls}(\mathcal{S}) + (1-t)\prod_{j=1}^{N} P^{*}_{ls}\big(X_{(j)}\big), \tag{6}$$

where

$$P_{ls}(\mathcal{A}) = \sum_{i_1,\ldots,i_N} \big(\log\left(|a_{i_1,\ldots,i_N}| + \varepsilon\right) - \log\left(\varepsilon\right)\big)/(-\log\left(\varepsilon\right)),$$
$$P^{*}_{ls}(A) = \sum_{m} \left(\log\left(\sigma_m(A) + \varepsilon\right) - \log\left(\varepsilon\right)\right)/(-\log\left(\varepsilon\right)),$$

are both with log-sum forms (shifted to $[0, +\infty]$, and scaled by $-\log(\varepsilon)$ to better approach l_0 norm and rank, respectively), ε is a small positive number, and $\sigma_m(A)$ defines the mth singular value of A. In the later section, we will use this relaxation form of KBR to build KBR-based models.

5 KBR-Based Model & Its Solving Scheme

In this section, we first introduce the general solving scheme for the KBR-based minimization problem, and then introduce KBR-based tensor completion and KBR-based robust principal component analysis extensions.

5.1 The KBR Minimization Model

KBRM model corresponds to the fundamental KBR-based tensor recovery model aiming at restoring a tensor from its observation under KBR regularization on tensor sparsity. By using the relaxation form $S^{\star}(\cdot)$ as defined in (6), we have the following KBRM model:

$$\min_{\mathcal{X}} P_{ls}(\mathcal{S}) + \lambda\prod_{j=1}^{N} P^{*}_{ls}\big(X_{(j)}\big) + \frac{\beta}{2}\|\mathcal{Y} - \mathcal{X}\|_F^2, \tag{7}$$

where $\lambda = \frac{1-t}{t}$ and β is the compromise parameter.

The alternating direction method of multipliers [4], [38], an effective strategy for solving large scale optimization problems, is readily employed for solving (7). First, we need to introduce N auxiliary tensors $\mathcal{M}_j$ $(j = 1, 2, \ldots, N)$ and equivalently reformulate (7) as

$$\begin{aligned}\min_{\mathcal{S},\mathcal{M}_j,U_j} & P_{ls}(\mathcal{S}) + \lambda\prod_{j=1}^{N} P^{*}_{ls}\big(M_{j(j)}\big) + \frac{\beta}{2}\|\mathcal{Y} - \mathcal{S}\times_1 U_1 \cdots \times_N U_N\|_F^2 \\ & \text{s.t. } \mathcal{S}\times_1 U_1 \cdots \times_N U_N - \mathcal{M}_j = 0, \\ & U_j^T U_j = I,\ j = 1, 2, \ldots, N,\end{aligned} \tag{8}$$

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

$$\begin{aligned}& L_{\mu}(\mathcal{S}, \mathcal{M}_1, \ldots, \mathcal{M}_N, U_1, \ldots, U_N, \mathcal{P}_1, \ldots, \mathcal{P}_N) \\ & = P_{ls}(\mathcal{S}) + \lambda\prod_{j=1}^{N} P^{*}_{ls}\big(M_{j(j)}\big) + \frac{\beta}{2}\|\mathcal{Y} - \mathcal{S}\times_1 U_1 \cdots \times_N U_N\|_F^2 \\ & \quad + \sum_{j=1}^{N}\langle \mathcal{S}\times_1 U_1 \cdots \times_N U_N - \mathcal{M}_j, \mathcal{P}_j\rangle \\ & \quad + \sum_{j=1}^{N}\frac{\mu}{2}\left\|\mathcal{S}\times_1 U_1 \cdots \times_N U_N - \mathcal{M}_j\right\|_F^2,\end{aligned}$$

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

With the other parameters fixed, $\mathcal{S}$ can be updated by solving $\min_{\mathcal{S}} L_{\mu}(\mathcal{S}, \mathcal{M}_1, \ldots, \mathcal{M}_N, U_1, \ldots, U_N, \mathcal{P}_1, \ldots, \mathcal{P}_N)$, i.e.,

$$\min_{\mathcal{S}} bP_{ls}(\mathcal{S}) + \frac{1}{2}\|\mathcal{S}\times_1 U_1 \cdots \times_N U_N - \mathcal{O}\|_F^2, \tag{9}$$

where $b = \frac{1}{\beta + N\mu}$ and $\mathcal{O} = \frac{\beta\mathcal{Y} + \sum_j(\mu\mathcal{M}_j - \mathcal{P}_j)}{\beta + N\mu}$. Since for any tensor $\mathcal{D}$ we have

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

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

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

where $\mathcal{Q} = \mathcal{O}\times_1 U_1^T \cdots \times_N U_N^T$, which has been proved to have closed-form solution [25]

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

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

$$\mathrm{D}_{b,\varepsilon}(x) = \begin{cases} 0 & \text{if} \quad |x| \le 2\sqrt{c_0 b} - \varepsilon \\ \text{sign}(x)\left(\frac{c_1(x) + c_2(x)}{2}\right) & \text{if} \quad |x| > 2\sqrt{c_0 b} - \varepsilon, \end{cases} \tag{13}$$

where $c_0 = \frac{-1}{\log(\varepsilon)}$, $c_1(x) = |x| - \varepsilon$, $c_2(x) = \sqrt{(|x| + \varepsilon)^2 - 4c_0 b}$.

With $U_j(j \neq k)$ and other parameters fixed, U_k $(1 \leq k \leq N)$ can be updated by solving $\min_{U_k} L_\mu(\mathcal{S}, \mathcal{M}_1, \ldots \mathcal{M}_N, U_1, \ldots, U_N, \mathcal{P}_1, \ldots, \mathcal{P}_N)$, i.e.,

$$\min_{U_k^T U_k = I} \|\mathcal{S} \times_1 U_1 \cdots \times_N U_N - \mathcal{O}\|_F^2. \tag{14}$$

By employing (10) and the following equation:

$$\mathcal{B} = \mathcal{D} \times_n V \Longleftrightarrow \mathcal{B}_{(n)} = V\mathcal{D}_{(n)}, \tag{15}$$

we can obtain that (14) is equivalent to

$$\max_{U_k^T U_k = I} \langle A_k, U_k \rangle. \tag{16}$$

where $A_k = \mathcal{O}_{(k)}(\text{unfold}_k(\mathcal{S} \bar{\times}_{-k} \{U_i\}_{i=1}^N))^T$. By using the *von Neumann's trace inequality* [48], we can easily solve (16), and update U_k by the following formula:

$$U_k^+ = B_k C_k{}^T. \tag{17}$$

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:

$$\min_{\mathcal{M}_k} a_k P_{ls}^*\big(M_{k(k)}\big) + \frac{1}{2}\|\mathcal{L} + \frac{1}{\mu}\mathcal{P}_k - \mathcal{M}_k\|_F^2, \tag{18}$$

where $a_k = (\frac{\lambda}{\mu}\prod_{j \neq k} P_{ls}^*(M_{j(j)}))$ and $\mathcal{L} = \mathcal{S} \times_1 U_1 \cdots \times_N U_N$. By using Theorem 1 in [43], it's easy to deduce that (18) has the following closed-form solution:

$$\mathcal{M}_k^+ = \text{fold}_k\big(V_1 \Sigma_{a_k} V_2^T\big), \tag{19}$$

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

The proposed algorithm for KBRM can be summarized in Algorithm 1. Note that the convergence of this ADMM algorithm is difficult to analyze due to the non-convexity of our model. By deductions similar to the results of general ADMM algorithms [4], [36], we can obtain the following weak convergence result[1] to facilitate the construction of a rational termination condition for Algorithm 1.

Algorithm 1. Algorithm for KBRM

Input: observed tensor $\mathcal{Y}$

1: Initialize $U_1^{(0)}, \ldots, U_N^{(0)}$ and $\mathcal{S}^{(0)}$ by HO-SVD of $\mathcal{Y}$, $\mathcal{M}_k^{(0)} = \mathcal{Y}$, $\forall k = 1, 2, \ldots, N$, $l = 1$, $\rho > 1$, $\mu^{(0)} > 0$
2: **while** not convergence **do**
3: Update $\mathcal{S}^{(l)}$ by (12)
4: Update all $U_k^{(l)}$ by (17);
5: Update all $\mathcal{M}_k^{(l)}$ by (19)
6: Update multipliers by $\mathcal{P}_k^{(l)} = \mathcal{P}_k^{(l-1)} + \mu^{(l)}(\mathcal{L} - \mathcal{M}_k^{(l-1)})$
7: Let $\mu^{(l)} := \rho\mu^{(l-1)}$; $l = l + 1$
8: **end while**

Output: $\mathcal{X} = \mathcal{S}^{(l)} \times_1 U_1^{(l)} \cdots \times_N U_N^{(l)}$

Remark 1. For sequences $\{\mathcal{S}^{(l)}\}$, $\{\mathcal{M}_k^{(l)}\}$ and $\{U_k^{(l)}\}$, $k = 1, 2, \ldots, N$, generated by Algorithm 1 and $\mathcal{X}^{(l)} = \mathcal{S}^{(l)} \times_1 U_1^{(l)} \times_2 \cdots \times_N U_N^{(l)}$, $\{\mathcal{M}_k^{(l)}\}$ and $\{\mathcal{X}^{(l)}\}$ satisfy:

$$\begin{aligned} \big\|\mathcal{X}^{(l)} - \mathcal{M}_k^{(l)}\big\|_F &= O\Big(\mu^{(0)}\rho^{-l/2}\Big), \\ \big\|\mathcal{X}^{(l+1)} - \mathcal{X}^{(l)}\big\|_F &= O\Big(\mu^{(0)}\rho^{-l/2}\Big). \end{aligned} \tag{20}$$

Note that the remark implies an easy termination strategy for our algorithm: Terminate the algorithm when the variation between reconstructed tensors in two adjacent steps is smaller than a preset threshold θ. The theorem not only helps guarantee that the algorithm can stop in $T = O((\log(\mu^{(0)}/\theta))/\log(\rho))$ iterations but also implies the equality constraint can be finely approximated in iteration. Similar results can be deduced for other KBR algorithms.

5.2 KBR for Tensor Completion

Tensor completion (TC) refers to the problem of recovering a tensor from only partial observations of its entries, which arises in a number of computer vision and pattern recognition applications, such as hyperspectral data recovery and video inpainting, etc [35], [51]. It is a well known ill-posed problem which needs prior of the ground truth tensor as supplementary information for reconstruction. We can then utilize the proposed KBR regularizer to encode such prior knowledge as follows:

$$\min_{\mathcal{X}} S(\mathcal{X}) \quad \text{s.t.} \quad \mathcal{X}_\Omega = \mathcal{T}_\Omega, \tag{21}$$

where $\mathcal{X}$, $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ are the reconstructed and observed tensors, respectively, and the elements of $\mathcal{T}$ in the set Ω are given while the remaining are missing; $S(\cdot)$ is the tensor sparsity measure.

By using the relaxation $S^*(\mathcal{X})$ as defined in (6), we can get the following KBR-TC model:

$$\min_{\mathcal{X}} P_{ls}(\mathcal{S}) + \lambda \prod_{j=1}^N P_{ls}^*\big(X_{(j)}\big) \quad \text{s.t.} \quad \mathcal{X}_\Omega = \mathcal{T}_\Omega. \tag{22}$$

Similar to the KBRM problem, we apply the ADMM to solving (22). First, we need to introduce N auxiliary tensors $\mathcal{M}_j$ $(j = 1, 2, \ldots, N)$ and equivalently reformulate (22) as

$$\begin{aligned} &\min_{\mathcal{S}, \mathcal{M}_j, U_j^T U_j = I} P_{ls}(\mathcal{S}) + \lambda \prod_{j=1}^N P_{ls}^*\big(M_{j(j)}\big) \\ &\text{s.t. } X - \mathcal{S} \times_1 U_1 \cdots \times_N U_N = 0, \\ &\mathcal{X}_\Omega - \mathcal{T}_\Omega = 0, \mathcal{X} - \mathcal{M}_j = 0, j = 1, 2, \ldots, N. \end{aligned}$$

Its augmented Lagrangian function is with the form

$$\begin{aligned} &L_\mu(\mathcal{S}, \mathcal{M}_1, \ldots, \mathcal{M}_N, U_1, \ldots, U_N, \mathcal{X}, \mathcal{P}^x, \mathcal{P}^t, \mathcal{P}_1^m, \ldots, \mathcal{P}_N^m) \\ &= P_{ls}(\mathcal{S}) + \lambda \prod_{j=1}^N P_{ls}^*\big(M_{j(j)}\big) + \langle \mathcal{X} - \mathcal{S} \times_1 U_1 \cdots \times_N U_N, \mathcal{P}^x \rangle \\ &\quad + \frac{\mu}{2}\|\mathcal{X} - \mathcal{S} \times_1 U_1 \cdots \times_N U_N\|_F^2, + \sum_{j=1}^N \langle \mathcal{X} - \mathcal{M}_j, \mathcal{P}_j^m \rangle \\ &\quad + \sum_{j=1}^N \frac{\mu}{2}\big\|\mathcal{X} - \mathcal{M}_j\big\|_F^2, + \langle \mathcal{X}_\Omega - \mathcal{T}_\Omega, \mathcal{P}_\Omega^t \rangle + \frac{\mu}{2}\|\mathcal{X}_\Omega - \mathcal{T}_\Omega\|_F^2, \end{aligned}$$

where $\mathcal{P}^x$, $\mathcal{P}^t$ and $\mathcal{P}_j^m$, $\forall j = 1, 2, \ldots, N$ are the Lagrange multipliers, μ is a positive scalar and U_j must satisfy $U_j^T U_j = I$, $\forall j = 1, 2, \ldots, N$. Then we can solve the problem under the ADMM framework.

1. The proof is presented in http://dymeng.gr.xjtu.edu.cn/8.

With other parameters fixed, $\mathcal{S}$ can be updated by solving subproblem similar to (11), which has the following closed-form solution:

$$\mathcal{S}^{+} = \mathrm{D}_{b,\varepsilon}\big((\mathcal{X} + \mu^{-1}\mathcal{P}^{x}) \times U_1^T \cdots \times_N U_N^T\big), \tag{23}$$

where $b = \frac{1}{\mu}$.

When updating U_k $(k = 1, 2, \ldots, N)$ with other parameters fixed, similar to the derivation of (17), we can obtain that U_k can be updated by

$$U_k^{+} = B_k C_k^T, \tag{24}$$

where $A_k = B_k D C_k^T$ is the SVD of A_k, and

$$A_k = \mathrm{unfold}_k\big(\mathcal{X} + \mu^{-1}\mathcal{P}^{x}\big)\mathrm{unfold}_k\Big(\mathcal{S}\bar{\times}_{-k}\{U_j\}_{j=1}^{N}\Big)^T.$$

When updating $\mathcal{M}_k$ $(k = 1, 2, \ldots, N)$ with other parameters fixed, similar to the derivation of (19), we can obtain

$$\mathcal{M}_k^{+} = \mathrm{fold}_k\big(V_1 \Sigma_{a_k} V_2^T\big), \tag{25}$$

where $\Sigma_{a_k} = \mathrm{diag}\big(\mathrm{D}_{a_k,\varepsilon}(\sigma_1), \mathrm{D}_{a_k,\varepsilon}(\sigma_2), \ldots, \mathrm{D}_{a_k,\varepsilon}(\sigma_n)\big)$ and $V_1 \mathrm{diag}(\sigma_1, \ldots, \sigma_n) V_2^T$ is the SVD of $\mathrm{unfold}_k(\mathcal{X} + \mu^{-1}\mathcal{P}_k^m)$, and $a_k = (\frac{\lambda}{\mu}\prod_{j \neq k} P_{ls}^{*}(M_{j(j)}))$.

With other parameters fixed, $\mathcal{X}$ can be updated by minimizing the augmented Lagrangian function on $\mathcal{X}$, which is equivalent to solving the following problem:

$$\begin{aligned}\min_{\mathcal{X}} &\left\| \mathcal{X} - \frac{1}{(N+1)\mu}\left(\mu\mathcal{L} - \mathcal{P}^{x} + \sum_j (\mu\mathcal{M}_j - \mathcal{P}_j^m)\right)\right\|_F^2 \\ &+ \frac{1}{N+1}\left\|\mathcal{X}_\Omega + \mu^{-1}\mathcal{P}_\Omega^{t} - \mathcal{T}_\Omega\right\|_F^2,\end{aligned} \tag{26}$$

where $\mathcal{L} = \mathcal{S} \times_1 U_1 \cdots \times_N U_N$. Its closed-form solution is

$$\begin{cases} \mathcal{X}_{\Omega^\perp}^{+} = \frac{1}{(N+1)\mu}\Big(\mu\mathcal{L} - \mathcal{P}^{x} + \sum_j (\mu\mathcal{M}_j - \mathcal{P}_j^m)\Big)_{\Omega^\perp} \\ \mathcal{X}_{\Omega}^{+} = \frac{1}{(N+2)\mu}\Big(\mu\mathcal{L} - \mathcal{P}^{x} + \mu\mathcal{T} - \mathcal{P}^{t} + \sum_j (\mu\mathcal{M}_j - \mathcal{P}_j^m)\Big)_{\Omega}, \end{cases} \tag{27}$$

where $\Omega^\perp$ defines the complement of Ω.

The proposed algorithm for KBR-TC can then be summarized in Algorithm 2.

Algorithm 2. Algorithm for KBR-TC

Input: observed tensor $\mathcal{T}$

1: Initialize $\mathcal{X}_\Omega^{(0)} = \mathcal{T}_\Omega$, $\mathcal{X}_{\Omega^\perp}^{(0)} = \mathrm{mean}(\mathcal{T}_\Omega)$, initialize and $\mathcal{S}^{(0)}$ and $U_1^{(0)}, \ldots, U_N^{(0)}$ by HO-SVD of $\mathcal{X}^0$, $\mathcal{M}_k^{(0)} = \mathcal{X}^{(0)}$, $\forall k = 1, 2, \ldots, N$, $l = 1$, $\rho > 1$, $\mu^{(0)} > 0$
2: **while** not convergence **do**
3: Update $\mathcal{S}^{(l)}$ by (23)
4: Update all $U_k^{(l)}$ by (24)
5: Update all $\mathcal{M}_k^{(l)}$ by (25)
6: Update $\mathcal{X}^{(l)}$ by (27)
7: Update multipliers by $\mathcal{P}_k^{(l)} = \mathcal{P}_k^{(l-1)} + \mu^{(l)}(\mathcal{L} - \mathcal{M}_k^{(l-1)})$
8: Let $\mu^{(l)} := \rho\mu^{(l-1)}$; $l = l + 1$
9: **end while**

Output: $\mathcal{X} = \mathcal{S}^{(l)} \times_1 U_1^{(l)} \cdots \times_N U_N^{(l)}$

5.3 KBR for Tensor Robust PCA

Tensor robust PCA (TRPCA) aims to recover the tensor from grossly corrupted observations, i.e.,

$$\mathcal{T} = \mathcal{L} + \mathcal{E}, \tag{28}$$

where $\mathcal{L}$ is the main tensor and $\mathcal{E}$ corresponds the noises/outliers embedded in data. Using the proposed KBR measure, we can get the following TRPCA model:

$$\min_{\mathcal{L},\mathcal{E}} S(\mathcal{L}) + \beta\|\mathcal{E}\|_1, \ \text{s.t.}\ \mathcal{T} = \mathcal{L} + \mathcal{E}, \tag{29}$$

where β is a tuning parameter compromising the recovered tensor and noise terms.

A more general noise assumption utilized to model the RPCA problem is to consider the noise as a mixture of sparse noise and Gaussian noise [72]. Correspondingly, we can also ameliorate the observation expression as

$$\mathcal{T} = \mathcal{L} + \mathcal{E} + \mathcal{N}, \tag{30}$$

where $\mathcal{N}$ is an additional Gaussian noise embedded in tensor. We can then propose the KBR-RPCA model

$$\min_{\mathcal{L},\mathcal{E},\mathcal{N}} P_{ls}\ (\mathcal{S}) + \lambda\prod_{j=1}^{N} P_{ls}^{*}\big(L_{(j)}\big) + \beta\|\mathcal{E}\|_1 + \frac{\gamma}{2}\|\mathcal{T} - \mathcal{L} - \mathcal{E}\|_F^2, \tag{31}$$

where $\mathcal{L} = \mathcal{S} \times_1 U_1 \cdots \times_N U_N$ is the Tucker decomposition of $\mathcal{L}$.

We also apply the ADMM to solving (31). First, we need to introduce N auxiliary tensors $\mathcal{M}_j$ $(j = 1, 2, \ldots, N)$ and equivalently reformulate (31) as follows:

$$\begin{aligned}&\min_{\mathcal{S},\mathcal{M}_j,U_j^T U_j = I} P_{ls}(\mathcal{S}) + \lambda\prod_{j=1}^{N} P_{ls}^{*}\big(M_{j(j)}\big) + \beta\|\mathcal{E}\|_1 \\ &+ \frac{\gamma}{2}\|\mathcal{S} \times_1 U_1 \cdots \times_N U_N + \mathcal{E} - \mathcal{T}\|_F^2 \\ &\text{s.t. } \mathcal{S} \times_1 U_1 \cdots \times_N U_N - \mathcal{M}_j = 0, \forall j = 1, 2, \ldots, N,\end{aligned} \tag{32}$$

and its augmented Lagrangian function is with the form

$$\begin{aligned}&L_\mu(\mathcal{S}, \mathcal{M}_1, \ldots, \mathcal{M}_N, U_1, \ldots, U_N, \mathcal{E}, \mathcal{P}_1, \ldots, \mathcal{P}_N) = P_{ls}(\mathcal{S}) \\ &+ \lambda\prod_{j=1}^{N} P_{ls}^{*}\big(M_{j(j)}\big) + \beta\|\mathcal{E}\|_1 + \frac{\gamma}{2}\|\mathcal{S} \times_1 U_1 \cdots \times_N U_N + \mathcal{E} - \mathcal{T}\|_F^2 \\ &+ \sum_{j=1}^{N}\langle \mathcal{S} \times_1 U_1 \cdots \times_N U_N - \mathcal{M}_j, \mathcal{P}_j\rangle \\ &+ \sum_{j=1}^{N}\frac{\mu}{2}\left\|\mathcal{S} \times_1 U_1 \cdots \times_N U_N - \mathcal{M}_j\right\|_F^2,\end{aligned}$$

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

With other parameters fixed, $\mathcal{S}$ can be updated by solving a subproblem similar to (11), which has the following closed-form solution:

$$\mathcal{S}^{+} = \mathrm{D}_{b,\varepsilon}(\mathcal{Q}). \tag{33}$$

where $b = \frac{1}{\gamma + N\mu}$, $\mathcal{Q} = \mathcal{O} \times_1 U_1^T \cdots \times_N U_N^T$, and $\mathcal{O} = \frac{\gamma(\mathcal{T} - \mathcal{E}) + \sum_j (\mu\mathcal{M}_j - \mathcal{P}_j)}{\gamma + N\mu}$.

When updating U_k $(k = 1, 2, \ldots, N)$ with other parameters fixed, similar to the derivation of (17), we can obtain the updating equation of U_k as

$$U_k^+ = B_k C_k^T, \tag{34}$$

where $A_k = B_k D C_k^T$ is the SVD decomposition of A_k, and $A_k = \mathcal{O}_{(k)}(\text{unfold}_k(\mathcal{S}\bar{\times}_{-k}\{U_i\}_{i=1}^N))^T$.

When updating $\mathcal{M}_k$ $(k = 1, 2, \ldots, N)$ with other parameters fixed, similar to the derivation of (19), we can obtain that $\mathcal{M}_k$ can be updatedp by

$$\mathcal{M}_k^+ = \text{fold}_k\big(V_1 \Sigma_{a_k} V_2^T\big), \tag{35}$$

where $\Sigma_{a_k} = \text{diag}\big(\text{D}_{a_k,\varepsilon}(\sigma_1), \text{D}_{a_k,\varepsilon}(\sigma_2), \ldots, \text{D}_{a_k,\varepsilon}(\sigma_n)\big)$, $V_1 \text{diag}(\sigma_1, \ldots, \sigma_n) V_2^T$ is the SVD of

$$\text{unfold}_k\big(\mathcal{S} \times_1 U_1 \cdots \times_N U_N + \mu^{-1}\mathcal{P}_k\big),$$

and $a_k = \Big(\frac{\lambda}{\mu}\prod_{j\neq k} P_{ls}^*\big(M_{j(j)}\big)\Big)$.

With other parameters fixed, $\mathcal{E}$ can be updated by minimizing the augmented Lagrangian function on $\mathcal{E}$, which is equivalent to solving the following problem:

$$\min_{\mathcal{E}} \beta\|\mathcal{E}\|_1 + \frac{\gamma}{2}\|\mathcal{S} \times_1 U_1 \cdots \times_N U_N + \mathcal{E} - \mathcal{T}\|_F^2,$$

which has the following closed-form solution [63]:

$$\mathcal{E}^+ = \text{S}_{\frac{\beta}{\gamma}}(\mathcal{T} - \mathcal{S} \times_1 U_1 \cdots \times_N U_N), \tag{36}$$

where $\text{S}_\tau(\cdot)$ is the soft thresholding operator [63]:

$$\text{S}_\tau(x) = \begin{cases} 0 & \text{if} \quad |x| \leq \tau \\ \text{sign}(x)(|x| - \tau) & \text{if} \quad |x| > \tau. \end{cases} \tag{37}$$

The KBR-RPCA algorithm can then be summarized as Algorithm 3.

Algorithm 3. Algorithm for KBR-RPCA

Input: observed tensor $\mathcal{T}$

1: Initialize $U_1^{(0)}, \ldots, U_N^{(0)}$ and $\mathcal{S}^{(0)}$ by HO-SVD of $\mathcal{T}$, $\mathcal{M}_k^{(0)} = \mathcal{T}$, $\forall k = 1, 2, \ldots, N$, $\mathcal{E} = 0$, $l = 1$, $\rho > 1$, $\mu^{(0)} > 0$
2: **while** not convergence **do**
3: Update $\mathcal{S}^{(l)}$ by (33)
4: Update all $U_k^{(l)}$ by (34)
5: Update all $\mathcal{M}_k^{(l)}$ by (35)
6: Update $\mathcal{E}^{(l)}$ by (36)
7: Update multipliers by $\mathcal{P}_k^{(l)} = \mathcal{P}_k^{(l-1)} + \mu^{(l)}(\mathcal{L} - \mathcal{M}_k^{(l-1)})$
8: Let $\mu^{(l)} := \rho\mu^{(l-1)}$; $l = l + 1$
9: **end while**

Output: $\mathcal{L} = \mathcal{S}^{(l)} \times_1 U_1^{(l)} \cdots \times_N U_N^{(l)}$, $\mathcal{E}^{(l)}$ and $\mathcal{N} = \mathcal{T} - \mathcal{L} - \mathcal{E}^{(l)}$

5.4 Applying KBRM Algorithm to MSI Denoising

In this section, we briefly introduce how to apply the proposed KBRM algorithm to MSI denoising, more details can be found in our conference paper [69].

The most significant issue of recovering a clean MSI from its corruption is to fully utilize the prior structure knowledge of the to-be-reconstructed MSI. The most commonly utilized two prior structures for MSI recovery are global correlation along spectrum (GCS) and nonlocal self-similarity across space (NSS). The GCS prior indicates that an MSI contains a large amount of spectral redundancy and the images obtained across the MSI spectrum 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 prior knowledge do be possessed by wide range of natural MSIs and be very helpful for various MSI recovery issues [12], [29], [52], [76].

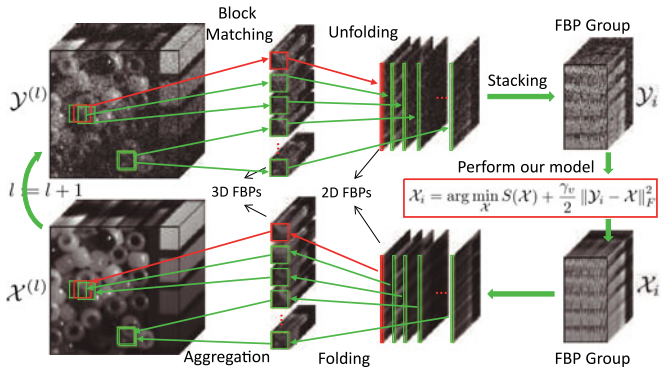

Fig. 5. Flowchart of the proposed MSI denoising method.

Albeit demonstrated to be effective to certain MSI denoising cases, most of the current methods to this task only take one such prior knowledge into full consideration. In [69], a new denoising framework as shown in Fig. 5 is proposed. Through block matching of full band patches (FBP), we can obtain a similar FBP group, and by unfolding each FBP along spectral dimensionality within the similar FBP group, we can represent each FBP group correlated to the ith key FBP as a 3-order tensor $\mathcal{X}_i$. Both GCS and NSS knowledge are well preserved and reflected by such representation, along the spectral and nonlocal-similar-patch-number modes of $\mathcal{X}_i$, respectively. Thus, high-order sparsity of all $\mathcal{X}_i$s can well encode both GCS and NSS of the whole MSI. Then, we can estimate the corresponding true nonlocal similarity FBPs $\mathcal{X}_i$ from its corruption $\mathcal{Y}_i$ by solving the following optimization problem:

$$\mathcal{X}_i = \arg\min_{\mathcal{X}} S^*(\mathcal{X}) + \frac{\gamma}{2}\|\mathcal{Y}_i - \mathcal{X}\|_F^2, \tag{38}$$

which is an KBRM problem and can be efficiently solved by Algorithm 1. By aggregating all reconstructed $\mathcal{X}_i$s we can reconstruct the estimated MSI. The entire denoising progress is summarized in Algorithm 4 and visualized in Fig. 5. We denote the algorithm as KBR-denoising for convenience.

Algorithm 4. Algorithm for KBR-Denoising

Input: Noisy MSI $\mathcal{Y}$

1: Initialize $\mathcal{X}^{(0)} = \mathcal{Y}$
2: **for** $l = 1 : L$ **do**
3: Calculate $\mathcal{Y}^{(l)} = \mathcal{X}^{(l-1)} + \delta\big(\mathcal{Y} - \mathcal{X}^{(l-1)}\big)$
4: Construct the entire FBP set $\Omega_{\mathcal{Y}^{(l)}}$
5: Construct the set of similar FBP group set $\{\mathcal{Y}_i\}_{i=1}^K$
6: **for** each FBP groups $\mathcal{Y}_i$ **do**
7: Solve the problem (38) by Algorithm 1
8: **end for**
9: Aggregate $\{\mathcal{X}_i\}_{i=1}^K$ to form the clean image $\mathcal{X}^{(l)}$
10: **end for**

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

5.5 Computational Complexity Analysis

Considering Algorithm 1 for an N-order input tensor $\mathcal{Y} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$, the main per-iteration cost lies in the update of U_j and $\mathcal{M}_j$, $j = 1, 2, \ldots, N$. Updating U_j requires computing an SVD of $I_j \times I_j$ matrix, and updating $\mathcal{M}_j$ requires computing an SVDs of $I_j \times (\prod_{i \neq j} I_i)$ matrix. Meanwhile, about $O((\log(\mu^{(0)}/\theta))/\log(\rho))$ iterations will be needed as deduced in Section 5.1. It is easy to see that Algorithms 2 and 3 also possess similar computation complexity.

Note that the per-iteration computing complexity of our algorithms is comparable to the rank-sum-based tensor recovery algorithm [39], [40], which need to perform N SVDs for matrices in the size of $I_j \times (\prod_{i \neq j} I_i)$, $j = 1, 2, \ldots, N$, in each iteration. When $N = 3$, it's also comparable to the t-SVD-based TC/TRPCA methods [42], [74], which need to compute I_3 (assuming $I_3 \gg N = 3$) SVDs for $I_1 \times I_2$ matrices each iteration.

For Algorithm 4 with an input MSI $\mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, the number of FBP groups is $K = O(I_1 I_2)$, the size of each FBP group is $p^2 \times s \times I_3$ (in our experiments, it is $36 \times 60 \times 31$), where p is the size of patch and s is number of FBPs in each similar group. The computation cost seems not very small for quite large K. However, denoising on the K FPB groups can be processed in parallel, each with relatively small computation complexity.

6 Experimental Results

In this section we will test the performance of KBRM, KBR-TC and KBR-RPCA methods, on MSI denoising, MSI completion and background subtraction tasks, respectively.[2]

6.1 MSI Denoising Experiments

We first test the KBR-denoising method on MSI denoising with synthetic noises for quantitative performance comparison of all competing methods, and then we test all methods on real MSI data.

Experimental Setting. The comparison methods include: band-wise K-SVD [1][3] and band-wise BM3D [13],[4] representing state-of-the-arts for 2D extended band-wise methods; 3D-cube K-SVD [18],[5] ANLM3D [46][6] and BM4D [44], [45],[7] representing state-of-the-arts for 2D extended 3D-cube-based methods; LRTA [49], PARAFAC [41],[8] TDL [52],[9] t-SVD [74][10] (using the same denoising framework as shown in Fig. 5) and Trace/TV[23] representing state-of-the-arts for tensor-based methods. All parameters involved in the competing methods were optimally assigned or selected as suggested in the reference papers.

Four quantitative picture quality indices (PQI) are employed for performance evaluation, including peak signal-to-noise ratio (PSNR), structure similarity (SSIM [67]), feature similarity (FSIM [73]), erreur relative globale adimensionnelle de synthèse (ERGAS [65]). 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.

TABLE 1
Performance Comparison (Mean + Variance) of 11 Competing Methods with Respect to 4 PQIs Averaged over All 32 Scenes and All Extents of Noises

	PSNR	SSIM	FSIM	ERGAS
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
Trace/TV	32.30±3.02	0.82±0.08	0.91±0.03	140.25±44.15
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
t-SVD	35.88±3.10	**0.91±0.04**	**0.96±0.02**	93.65 ± 31.68
KBR-denoising	**37.71±3.39**	**0.91±0.05**	**0.96±0.02**	**78.21±31.59**

The best result in each PQI measure is highlighted in bold.

Synthetic Noise Simulations. The Columbia MSI Database [71][11] is utilized in our simulated experiments. This dataset contains 32 real-world scenes of a 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 400 to 700 nm in 10 nm steps. In our experiments, each MSI is pre-scaled into the interval $[0, 1]$.

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 parameter λ is used to balance two parts in the same order of magnitude, and thus it should neither be too small nor too large, and we empirically find that our algorithm will achieve satisfactory performance when λ is taken in the range $[0.1, 10]$. In all of the experiments here, we just simply choose $\lambda = 10$. The parameter β is dependent on v, and we let $\beta = cv^{-1}$, where c is set as the constant 10^{-3}. The two algorithm parameters, ρ and μ, are set as 1.05 and 250, respectively.

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 the evaluation measures, e.g., our method achieves around 1.5 dB gain in

2. More experimental demonstrations are listed in http://dymeng.gr.xjtu.edu.cn/8. The related Matlab codes of our algorithms can be downloaded in http://bit.ly/2j8BVZ2.
3. http://www.cs.technion.ac.il/~elad/software
4. http://www.cs.tut.fi/~foi/GCF-BM3D/
5. http://www.cs.technion.ac.il/~elad/software
6. http://personales.upv.es/jmanjon/denoising/arnlm.html
7. http://www.cs.tut.fi/~foi/GCF-BM3D/
8. http://www.sandia.gov/ tgkolda/TensorToolbox/index-2.5.html
9. http://gr.xjtu.edu.cn/web/dymeng/2
10. http://www.ece.tufts.edu/~shuchin/software.html
11. http://www1.cs.columbia.edu/CAVE/databases/multispectral

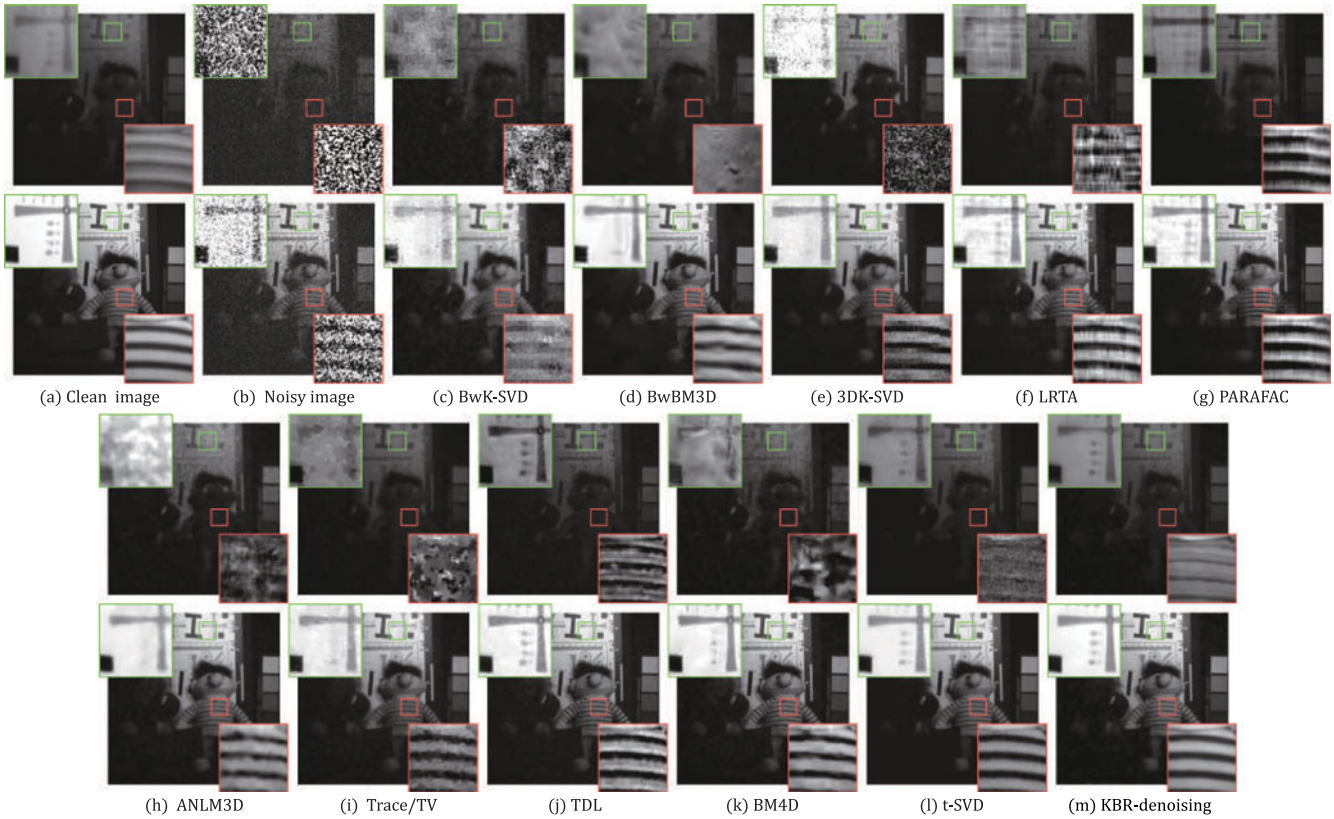

Fig. 6. (a) The images at two bands (400 and 700 nm) of *chart and staffed toy*, (b) The corresponding images corrupted by Gaussian noise with variance $v = 0.2$, (c)-(m) The restored images obtained by the 10 utilized MSI denoising methods. Two demarcated areas in each image are amplified at a 4 time larger scale and with the same degree of contrast for easy observation of details.

PSNR and 10 in ERGAS beyond the second best performed method. Such superiority can be more comprehensively observed in more detailed reports as listed in the website: http://dymeng.gr.xjtu.edu.cn/8.

To further depict the denoising performance of our method, we show in Fig. 6 two bands in *chart and stuffed toy* that centered at 400 nm (the darker one) and 700 nm (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.

Real MSI Experiments. The HYDICE MSI of natural scenes[12] is used in our experiments. The original MSI is of size $304 \times 304 \times 210$. As the bands 76, 100-115, 130-155 and 201-210 are seriously polluted by atmosphere and water absorption and provide little useful information, we remove them and leave the remaining test data with size $304 \times 304 \times 157$. We then easily pre-scaled the MSI into the interval $[0, 1]$ for all competing methods. Since the noise level is unknown for real noisy images, we use an off-the-shelf noise estimation method [17] to estimate it and utilize this knowledge to set parameter β. The parameters λ, ρ, μ are set the same as previous section.

We illustrate the experimental results at the first band of the test MSI in Fig. 7. It is seen that the band image restored by our method is capable of better removing the unexpected stripes and Gaussian noise while finely preserving the structure underlying the MSI, while the results obtained by most of other competing methods remain large amount of stripes noises. BM3D and BM4D can perform comparatively better in stripes noise removing, but their results contain evident blurry area, where our method evidently recovers more details hiding under the corrupted MSI.

12. http://www.tec.army.mil/Hypercube

6.2 MSI Completion by KBR-TC

In this section, we first test the performance of the KBR-TC method in simulations, and then validate its effectiveness in application of MSI completion.

Experimental Setting. The comparison methods including: ADMM(ALM)-based matrix completion (MC-ALM) [36], HaLRTC [40], the factorization-based TC method (TMac) [70], joint trace/TV based TC method (Trace/TV) [23], tensor-SVD-based TC method (t-SVD) [74], minmax concave plus penalty-based TC method (McpTC) [11] and the smoothly clipped absolute deviation penalty-based TC method (ScadTC) [11]. For matrix-based method MC-ALM, we applied it to the unfolded matrix along each mode of the tensor, obtaining 3 RREs, and report the best one as the final result. These competing methods cover the state-of-the-art along the research line of tensor completion. There is only one parameter λ in our model and we use the same scheme as previous section.

Synthetic Simulations. The synthetic data were generated as follows: first, the ground truth tensor was yielded from the Tucker model, i.e., $\mathcal{T} = \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 U_3$, where the

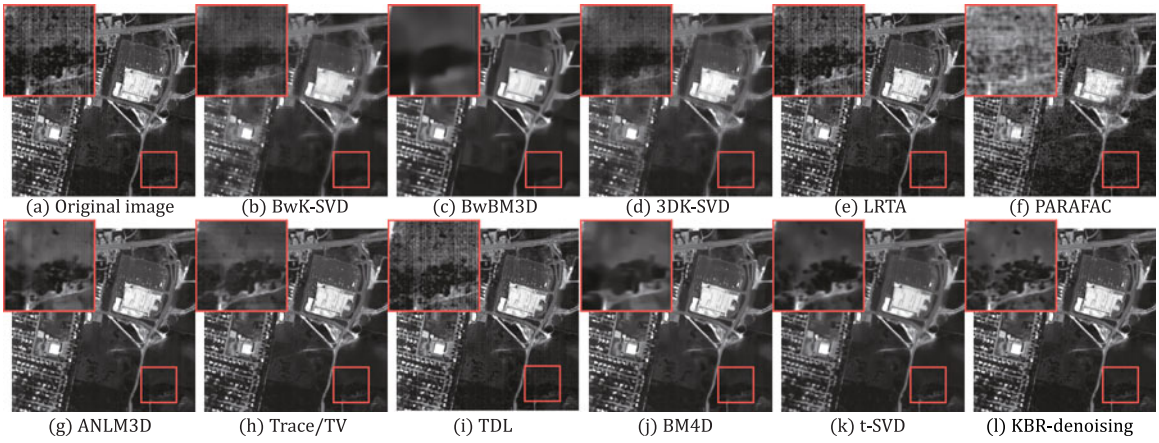

Fig. 7. (a) The Original image located at the first band in HYDICE urban data set; (b)-(l) The restored image obtained by the 11 utilized methods.

core tensor $\mathcal{S} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ was randomly generated from the standard Gaussian distribution, and each $U_i \in \mathbb{R}^{I_i \times r_i}$ was randomly generated column orthogonal matrices, i.e., $U_i^T U_i = I$ $(1 \leq i \leq 3)$; then a portion of elements was randomly sampled as observed data while the rest were left as missing components. We set $I_i (i = 1, 2, 3)$ to 50, respectively, resulting the ground truth tensor with size $50 \times 50 \times 50$. For the rank parameters r_i along each mode, we considered two settings, $(30, 30, 30)$ and $(10, 35, 40)$. The former simulates the situation that the rank is similar along each mode as most previous literatures [21], [40]; the latter simulates the case that the ranks along different modes are with evident diversity, which is always encountered in practice such as for MSI scenarios, in which the rank along the spectral mode is always much lower than those along the spatial modes. We then varied the percentage of sampled elements from 20 to 40 percent and implemented the tensor completion methods. The performance in terms of a commonly utilized measure for such simulations: *relative reconstruction error* (RRE),[13] averaged over 20 realizations, was summarized in Table 2.

It can be seen from Table 2 that, compared with other competing methods, the proposed KBR-based TC method can more accurately recover the tensor with fewer observations, and this advantage becomes more significant when the ground truth rank along each mode is with more diversity. This verifies the robustness of the proposed method in terms of rank variations along all tensor modes.

MSI Completion Experiments. We use the Columbia MSI in our simulated experiment and take the same PQI measures as introduced in the last section for performance evaluation. Each image is resized to 256×256 for all spectral bands, and rescale to $[0, 1]$. We varied the sampling rate from 5 to 20 percent, and applied TC methods to recover the images. The parameters λ, ρ, μ are set as 0.1, 1.05, 100, respectively, throughout all experiments.

For each sampling rate, all of the four PQI values for each competing MSI completion methods on all 32 scenes have been calculated and recorded. Table 3 lists the average performance (over different scenes) of all methods. From these quantitative comparison, it can be observed that our method performs evidently better than all other competing methods with respective to all evaluation measures, e.g., our method achieves around 3 dB gain in PSNR and 30 in ERGAS beyond the second best performed method. Such performance gain is obviously statistically significant.

To further visually compare the performance of all competing methods, we show in Fig. 8 the band in *fake and real lemons image* that centered at 700 nm with sampling rate 10 percent. From the figure, the superiority of the proposed method, both in the recovery of finer-grained textures and coarser-grained structures, can be easily observed.

Besides, in the figure we also show the configurations of the recovered Kronecker bases corresponding to the six largest coefficient value in the core tensor with sampling rate 10 percent (ranked based on the element value of the core tensor). It can be seen that each basis extracts certain simple while intrinsic structure underlying the input uncompleted tensor. Furthermore, the figure depicts the reconstructed tensors by only using the top-10, 100, 1,000 Kronecker bases extracted from the new algorithm. It can be observed that the reconstructed tensor gradually recovers the groudtruth ones under such a low sampling rate.

TABLE 2
Performance Comparison of 7 Competing TC Methods with Respect to RRE on Synthetic Data with Rank Setting (30,30,30) (Upper) and (10, 35, 40) (Lower)

Method	20%	25%	30%	35%	40%
AlmMC	8.23e-01	7.66e-01	7.12e-01	6.53e-01	5.93e-01
HaLRTC	8.95e-01	8.65e-01	8.37e-01	8.06e-01	7.75e-01
Tmac	1.99e-01	4.39e-03	1.41e-04	3.39e-05	2.31e-05
t-SVD	9.29e-01	8.79e-01	8.27e-01	7.69e-01	7.09e-01
McpTC	3.79e-01	9.68e-05	7.55e-09	5.12e-09	4.16e-09
ScadTC	**1.08e-01**	1.61e-04	1.22e-05	6.02e-09	4.48e-09
KBR-TC	1.49e-01	**7.09e-09**	**5.40e-09**	**4.18e-09**	**3.30e-09**
Method	20%	25%	30%	35%	40%
AlmMC	5.67e-01	4.38e-01	3.07e-01	1.75e-01	6.33e-02
HaLRTC	8.95e-01	8.66e-01	8.37e-01	8.06e-01	7.74e-01
Tmac	9.06e-01	8.63e-01	8.11e-01	7.37e-01	6.35e-01
t-SVD	8.46e-01	7.64e-01	6.69e-01	5.68e-01	4.62e-01
McpTC	5.76e-01	2.10e-01	4.84e-03	3.50e-05	1.68e-07
ScadTC	4.96e-01	4.52e-02	4.96e-04	2.15e-05	4.49e-09
KBR-TC	**2.20e-01**	**6.60e-09**	**4.69e-09**	**3.15e-09**	**2.32e-09**

13. RRE is defined as $\frac{\|\mathcal{T}-\mathcal{X}\|_F}{\|\mathcal{T}\|_F}$, where $\mathcal{T}$ and $\mathcal{X}$ denote the ground truth and reconstructed tensors, respectively.

TABLE 3
The Average Performance Comparison of 8 Competing TC Methods with Different Sampling Rates on 32 MSI

Method	5%				10%				20%				Time/s
	PSNR	SSIM	FSIM	ERGAS	PSNR	SSIM	FSIM	ERGAS	PSNR	SSIM	FSIM	ERGAS	
AlmMC	24.97±3.6	0.70±3.6	0.80±3.6	333.33±3.6	28.31±4.5	0.79±4.5	0.86±4.5	236.43±4.5	31.88±4.9	0.87±4.9	0.92±4.9	160.70±4.9	**5.48±0.7**
HaLRTC	25.54±4.8	0.74±4.8	0.83±4.8	329.58±4.8	29.76±5.3	0.84±5.3	0.89±5.3	207.74±5.3	34.30±5.6	0.92±5.6	0.94±5.6	126.57±5.6	14.89±3.0
Tmac	17.34±3.5	0.36±3.5	0.63±3.5	763.16±3.5	19.34±3.6	0.44±3.6	0.64±3.6	630.43±3.6	25.55±3.8	0.67±3.8	0.79±3.8	370.34±3.8	6.53±5.4
Trace/TV	21.71±3.9	0.70±3.9	0.81±3.9	484.26±3.9	30.07±4.5	0.88±4.5	0.92±4.5	197.30±4.5	37.43±4.4	0.96±4.4	0.97±4.4	87.02±4.4	51.66±4.0
t-SVD	30.40±4.3	0.82±4.3	0.88±4.3	186.96±4.3	34.18±4.7	0.89±4.7	0.93±4.7	124.46±4.7	38.91±4.9	0.95±4.9	0.97±4.9	74.46±4.9	658.07±108.2
McpTC	32.09±4.7	0.86±4.7	0.90±4.7	155.62±4.7	35.03±5.1	0.91±5.1	0.93±5.1	115.29±5.1	38.74±5.5	0.95±5.5	0.96±5.5	77.52±5.5	481.44±13.1
ScadTC	32.28±4.8	0.85±4.8	0.90±4.8	153.63±4.8	35.14±5.2	0.90±5.2	0.93±5.2	114.61±5.2	38.72±5.6	0.94±5.6	0.96±5.6	77.94±5.6	481.41±13.4
KBR-TC	**35.40±5.2**	**0.91±5.2**	**0.94±5.2**	**108.52±5.2**	**40.24±5.2**	**0.96±5.2**	**0.97±5.2**	**62.44±5.2**	**45.12±4.9**	**0.99±4.9**	**0.99±4.9**	**35.79±4.9**	330.82±18.9

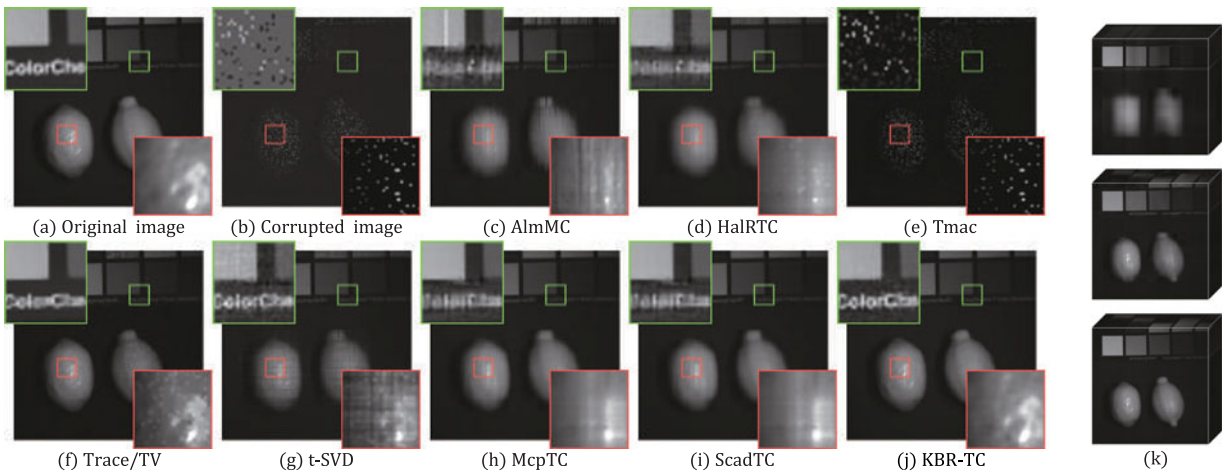

Fig. 8. (a) Original images located at the band centered at 700 nm of *fake and real lemons*; (b) The corresponding sampled images with sampling rate 10 percent; (c)-(j) Restored images obtained by 8 competing methods; (k) Combination results of the first 10, 100 and 1,000 bases obtained by KBR-TC, respectively.

6.3 Background Subtraction by KBR-RPCA

In this section, we validate the proposed KBR-RPCA model in application of background subtraction.

Experimental Setting. The comparison methods including three state-of-the-art TRPCA methods: RPCA [36], HoRPCA [24] and t-SVD [74]. For matrix-based RPCA, we only consider the low-rank property along temporal dimension, since the matrix-based method can only capture one type of correlation, and the spectral one is always the mode with the most correlation in this scenario. There are three parameters λ, β and γ required to be selection in our model. We consistently set $\lambda = 10$ throughout this experimental. Since the last term in our KBR-RPCA model (RPCA-OBJ) is used to model a slight perturbation, a large γ will thus be rational, and we set $\gamma = 100\beta$ in all our experiments empirically. As for β, we empirically found that this parameter should be set proportional to $\sqrt{\max(I_1, I_2, I_3)}$, where I_1, I_2, I_3 are the size of three video dimensionalities. Through easily tuning β around this empirical value, good performance of our method can be achieved in all our experiments.

Background Subtraction Experiments. The I2R dataset [34][14] containing 9 videos is used in our experiments. We use F-measure [10] to assess the detection performance of video foreground. The larger F-measure is, the closer the detected foreground area to the ground truth foreground area. Table 4 lists the F-measure, averaged over 20 ground-truth demarcated frames of each video in the dataset, with respect to all methods. The superiority of the proposed method throughout all videos can be observed. The visual results of representative frames in the *Shopping Mall* and *Hall* sequences are shown in Fig. 9. From these figures, we can see that KBR-RPCA method is able to deliver clear background estimation even under prominently embedded foreground moving objects. This also facilitates a more accurate foreground estimation. Comparatively, in the results estimated by other competing methods, there are more ghost shadows

14. http://perception.i2r.a-star.edu.sg/bk_model/bk_index

TABLE 4
Quantitative Performance Comparison of All Competing Methods

	RPCA	HoRPCA	t-SVD	KBR-RPCA
Hall	0.491±0.057	0.421±0.036	0.502±0.038	**0.531±0.061**
ShoppingMall	0.685±0.003	0.506±0.003	0.565±0.002	**0.704±0.003**
Campus	0.475±0.013	0.307±0.016	0.436±0.014	**0.512±0.012**
Fountain	0.586±0.034	0.319±0.012	0.520±0.018	**0.641±0.028**
Escalator	0.470±0.011	0.373±0.011	0.500±0.014	**0.532±0.013**
Curtain	0.505±0.010	0.534±0.008	0.597±0.010	**0.615±0.017**
Bootstrap	0.576±0.033	0.494±0.035	0.536±0.032	**0.586±0.037**
WaterSurface	0.276±0.031	0.291±0.021	0.423±0.019	**0.431±0.021**
Lobby	0.638±0.034	0.246±0.023	0.361±0.033	**0.653±0.039**
Time/s	**1.4 ± 1.1**	152.7±162.7	118.5 ± 93.4	42.5 ± 36.0

The results are obtained by averaging the F-measure through 20 ground-truth demarcated frames for each video provided in the dataset.

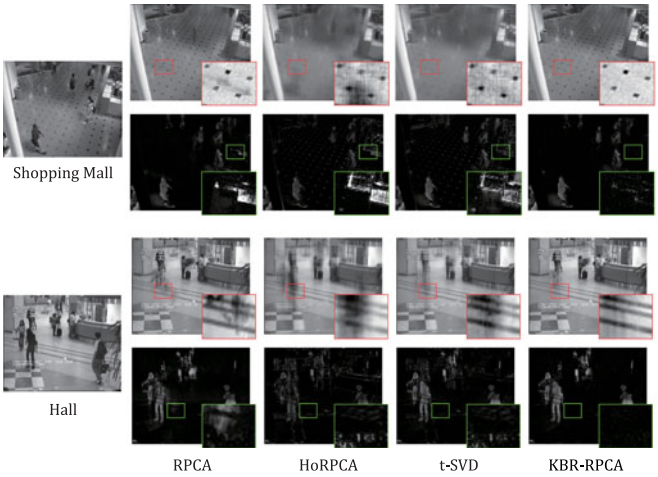

Fig. 9. From left to right: Original video frames, background and foreground extracted by all competing methods.

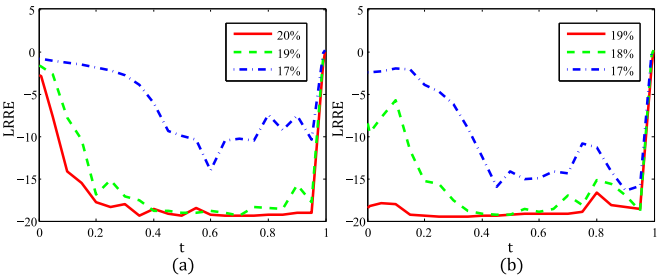

Fig. 10. Performance variation of the proposed method in terms of the LRRE on different t and different sampling rate. (a) Ranks along each mode is $(10, 30, 30)$. (b) Ranks along each mode is $(25, 25, 25)$.

in the background, and some objects belong to the background are mistakenly separated to be foreground part.

6.4 Analysis of the Compromising Parameter

Since the proposed KBR measure is composed of two terms balanced with the parameter t, it is necessary to make a discussion on properly setting t. In this section, we give some simulated TC experiments to analyze this issue. The synthetic data were generated in the same way as our TC experiments as introduced in Section 6.2. Here, the size of the generated tensor is $50 \times 50 \times 50$. We simulate two typical rank diversity situations and set the rank of each mode to $(10, 30, 30)$ and $(25, 25, 25)$.

Performance of the proposed method under different settings of t, in terms of *logarithmic relative reconstruction error* (LRRE),[15] averaged over 50 realizations, is visually shown in Fig. 10. From the figure, it is easy to observe that, just as analyzed in Section 3 of the maintext, the proposed KBR benefits from both of its sparsity components and under a wide range of t, utilizing both components performs evidently better than just using one (corresponding to those in two extreme positions). This phenomenon actually has been consistently observed in all our experiments, which empirically verifies the easy selection of this parameter.

15. Defined as $\log\left(\frac{\|\mathcal{T}-\mathcal{X}\|_F}{\|\mathcal{T}\|_F}\right)$, where $\mathcal{T}$ and $\mathcal{X}$ denote the ground truth and reconstructed tensors, respectively. Here logarithmic operator is utilized to amplify the performance variation under different parameter settings.

7 Conclusion

Inspired from the insights of both Tucker and CP decompositions for tensor, this paper has presented a tensor sparsity measure to encode intrinsic sparsity prior knowledge under a practical tensor. This measure not only has an easily interpretable physical meaning, i.e., a rectifying term to the number of rank-1 Kronecker bases for representing the tensor, but also unifies the understanding of sparsity previously designed on vector and matrix. Extensive experiments implemented on MSI denoising, MSI completion and background subtraction have substantiated the effectiveness of the proposed tensor sparsity measure. Specifically, through using this measure, the proposed methods outperform the state-of-the-art methods specifically designed on these tasks.

In our future research, we will more deeply discover the theoretical properties under this measure (e.g., under which condition the measure can help get an exact recovery from a corrupted/sampled tensor) and try to utilize it to ameliorate more tensor analysis applications. Besides, we will further investigate the efficiency speedup and better initialization issues for our algorithm.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant No. 61661166011, 61373114, 11690011, 61603292, the 973 Program of China under Grant No. 2013CB329404 and Macau Science and Technology Development Funds under Grant No. 003/2016/AFJ.

References

[1] M. Aharon, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
[2] S. D. Babacan, M. Luessi, R. Molina, and A. K. Katsaggelos, "Sparse Bayesian methods for low-rank matrix estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 3964–3977, Aug. 2012.
[3] Y. Benezeth, P.-M. Jodoin, B. Emile, H. Laurent, and C. Rosenberger, "Comparative study of background subtraction algorithms," *J. Electron. Imag.*, vol. 19, no. 3, pp. 033003–033003, 2010.
[4] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends? Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.
[5] A. M. Buchanan and A. W. Fitzgibbon, "Damped Newton algorithms for matrix factorization with missing data," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 316–322.
[6] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, 2011, Art. no. 11.
[7] E. J. Candes and Y. Plan, "Matrix completion with noise," *Proc. IEEE*, vol. 98, no. 6, pp. 925–936, Jun. 2010.
[8] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Mathematics*, vol. 9, no. 6, pp. 717–772, 2009.
[9] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5/6, pp. 877–905, 2008.
[10] W. Cao, Y. Wang, J. Sun, D. Meng, C. Yang, A. Cichocki, and Z. Xu, "Total variation regularized tensor RPCA for background subtraction from compressive measurements," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4075–4090, Jun. 2016.
[11] W. Cao, Y. Wang, C. Yang, X. Chang, Z. Han, and Z. Xu, "Folded-concave penalization approaches to tensor completion," *Neurocomputing*, vol. 152, pp. 261–273, 2015.
[12] A. A. Chen, "The inpainting of hyperspectral images: A survey and adaptation to hyperspectral data," in *Proc. SPIE*, 2012, Art. no. 85371.

[13] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[14] F. De La Torre and M. J. Black, "A framework for robust subspace learning," *Int. J. Comput. Vis.*, vol. 54, no. 1–3, pp. 117–142, 2003.

[15] X. Ding, L. He, and L. Carin, "Bayesian robust principal component analysis," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3419–3430, Dec. 2011.

[16] W. Dong, G. Shi, and X. Li, "Nonlocal image restoration with bilateral variance estimation: A low-rank approach," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 700–711, Feb. 2013.

[17] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[18] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[19] A. Eriksson and A. van Den Hengel, "Efficient computation of robust low-rank matrix approximations in the presence of missing data using the L 1 norm," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 13–18.

[20] F. Fouss, A. Pirotte, J.-M. Renders, and M. Saerens, "Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 3, pp. 355–369, Mar. 2007.

[21] S. Gandy, B. Recht, and I. Yamada, "Tensor completion and low-n-rank tensor recovery via convex optimization," *Inverse Problems*, vol. 27, no. 2, 2011, Art. no. 025010.

[22] M. Golbabaee and P. Vandergheynst, "Hyperspectral image compressed sensing via low-rank and joint-sparse matrix recovery," in *Proc. IEEE Int. Conf. Acoust.*, 2012, pp. 2741–2744.

[23] M. Golbabaee and P. Vandergheynst, "Joint trace/TV norm minimization: A new efficient approach for spectral compressive imaging," in *Proc. 19th IEEE Int. Conf. Image Process.*, 2012, pp. 933–936.

[24] D. Goldfarb and Z. Qin, "Robust low-rank tensor recovery: Models and algorithms," *SIAM J. Matrix Anal. Appl.*, vol. 35, no. 1, pp. 225–253, 2014.

[25] 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 *Proc. 30th Int. Conf. Mach. Learn.*, 2013, pp. II-37–II-45.

[26] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 2862–2869.

[27] W. He, H. Zhang, L. Zhang, and H. Shen, "Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 178–188, Jan. 2016.

[28] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He, "Fast and accurate matrix completion via truncated nuclear norm regularization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 9, pp. 2117–2130, Sep. 2013.

[29] R. Kawakami, J. Wright, Y.-W. Tai, Y. Matsushita, M. Ben-Ezra, and K. Ikeuchi, "High-resolution hyperspectral imaging via matrix factorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 2329–2336.

[30] Q. Ke and T. Kanade, "Robust L1 norm factorization in the presence of outliers and missing data by alternative convex programming," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 739–746.

[31] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.

[32] T. G. Kolda, B. W. Bader, and J. P. Kenny, "Higher-order web link analysis using multilinear algebra," in *Proc. 5th IEEE Int. Conf. Data Mining*, 2005, Art. no. 8.

[33] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, 2000.

[34] L. Li, W. Huang, I. Y.-H. Gu, and Q. Tian, "Statistical modeling of complex backgrounds for foreground object detection," *IEEE Trans. Image Process.*, vol. 13, no. 11, pp. 1459–1472, Nov. 2004.

[35] N. Li and B. Li, "Tensor completion for on-board compression of hyperspectral images," in *Proc. IEEE Int. Conf. Image Process.*, 2010, pp. 517–520.

[36] Z. Lin, M. Chen, and Y. Ma, "The augmented LaGrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.

[37] Z. Lin, A. Ganesh, J. Wright, L. Wu, M. Chen, and Y. Ma, "Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix," in *Proc. IEEE Int. Workshop Comput. Advances Multi-Sensor Adaptive Process.*, 2009, Art. no. 61.

[38] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proc. Advances Neural Inf. Process. Syst.*, 2011, pp. 612–620.

[39] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, 2009, pp. 2114–2121.

[40] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, Jan. 2013.

[41] X. Liu, S. Bourennane, and C. Fossati, "Denoising of hyperspectral images using the PARAFAC model and statistical performance analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3717–3724, Oct. 2012.

[42] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 5249–5257.

[43] C. Lu, C. Zhu, C. Xu, S. Yan, and Z. Lin, "Generalized singular value thresholding," in *Proc. 29th AAAI Conf. Artif. Intell. Mach. Learn. Appl.*, 2015, pp. 1805–1811.

[44] M. Maggioni and A. Foi, "Nonlocal transform-domain denoising of volumetric data with groupwise adaptive variance estimation," in *Proc. SPIE*, 2012, Art. no. 82960.

[45] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "A nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 119–133, Jan. 2013.

[46] 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," *J. Magn. Resonance Imag.*, vol. 31, no. 1, pp. 192–203, 2010.

[47] D. Meng and F. Torre, "Robust matrix factorization with unknown noise," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 1337–1344.

[48] L. Mirsky, "A trace inequality of John von Neumann," *Monatshefte für Mathematik*, vol. 79, no. 4, pp. 303–306, 1975.

[49] S. B. N. Renard and J. Blanc-Talon, "Denoising and dimensionality reduction using multilinear tools for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 5, no. 2, pp. 138–142, Apr. 2008.

[50] F. Parvaresh, H. Vikalo, S. Misra, and B. Hassibi, "Recovering sparse signals using sparse measurement matrices in compressed DNA microarrays," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 3, pp. 275–285, Jun. 2008.

[51] K. A. Patwardhan, G. Sapiro, and M. Bertalmío, "Video inpainting under constrained camera motion," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 545–553, Feb. 2007.

[52] Y. Peng, D. Meng, Z. Xu, C. Gao, Y. Yang, and B. Zhang, "Decomposable nonlocal tensor dictionary learning for multispectral image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 2949–2956.

[53] H. Rauhut, R. Schneider, and Ž. Stojanac, "Low rank tensor recovery via iterative hard thresholding," *Linear Algebra Appl.*, vol. 523, pp. 220–262, 2017.

[54] S. Raychaudhuri, J. M. Stuart, and R. B. Altman, "Principal components analysis to summarize microarray experiments: Application to sporulation time series," in *Proc. Pacific Symp. Biocomputing*, 2000, pp. 455–466.

[55] E. Richard, F. Bach, and J. P. Vert, "Intersecting singularities for multi-structured estimation," in *Proc. 30th Int. Conf. Mach. Learn.*, 2013, pp. 1157–1165.

[56] E. Richard, P.-A. Savalle, and N. Vayatis, "Estimation of simultaneously sparse and low rank matrices," *arXiv preprint arXiv:1206.6474*, 2012.

[57] B. Romera-Paredes and M. Pontil, "A new convex relaxation for tensor completion," in *Proc. Advances Neural Inf. Process. Syst.*, 2013, pp. 2967–2975.

[58] M. Rup, L. K. Hansen, and S. M. Arnfred, *Algorithms for Sparse Nonnegative Tucker Decompositions*. Cambridge, MA, USA: MIT Press, 2008.

[59] A. C. Sauve, A. O. Hero, W. L. Rogers, S. J. Wilderman, and N. H. Clinthorne, "3D image reconstruction for a Compton SPECT camera model," *IEEE Trans. Nucl. Sci.*, vol. 46, no. 6, pp. 2075–2084, Dec. 1999.

[60] A. Shashua, "On photometric issues in 3D visual recognition from a single 2D image," *Int. J. Comput. Vis.*, vol. 21, no. 1/2, pp. 99–122, 1997.
[61] N. Srebro and T. Jaakkola, "Weighted low-rank approximations," in *Proc. 20th Int. Conf. Mach. Learn.*, 2003, pp. 720–727.
[62] O. Taheri and S. Vorobyov, "Sparse channel estimation with L p-norm and reweighted L 1-norm penalized least mean squares," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2011, pp. 2864–2867.
[63] R. Tibshirani, "Regression shrinkage and selection via the Lasso: A retrospective," *J. Roy. Statist. Soc.*, vol. 73, no. 3, pp. 273–282, 2011.
[64] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.
[65] L. Wald, *Data Fusion: Definitions and Architectures: Fusion of Images of Different Spatial Resolutions*. Paris, France: Presses des l'Ecole MINES, 2002.
[66] H. Wang, F. Nie, and H. Huang, "Low-rank tensor completion with spatio-temporal consistency," in *Proc. 28th AAAI Conf. Artif. Intell.*, 2014, pp. 2846–2852.
[67] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[68] 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 *Proc. Advances Neural Inf. Process. Syst.*, 2009, pp. 2080–2088.
[69] Q. Xie, et al., "Multispectral images denoising by intrinsic tensor sparsity regularization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1692–1700.
[70] Y. Xu, R. Hao, W. Yin, and Z. Su, "Parallel matrix factorization for low-rank tensor completion," *Inverse Problems Imag.*, vol. 9, no. 2, pp. 601–624, 2015.
[71] 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 Process.*, vol. 19, no. 9, pp. 2241–2253, Sep. 2010.
[72] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4729–4743, Aug. 2014.
[73] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
[74] Z. Zhang, G. Ely, S. Aeron, N. Hao, and M. Kilmer, "Novel methods for multilinear data completion and de-noising based on Tensor-SVD," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 3842–3849.
[75] Q. Zhao, D. Meng, Z. Xu, W. Zuo, and L. Zhang, "Robust principal component analysis with complex noise," in *Proc. 31st Int. Conf. Mach. Learn.*, 2014, pp. 55–63.
[76] X. Zhao, F. Wang, T. Huang, M. K. Ng, and R. J. Plemmons, "Deblurring and sparse unmixing for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4045–4058, Jul. 2013.

Qi Xie received the BSc degree from Xi'an Jiaotong University, Xi'an, China, in 2013, where he is currently working toward the PhD degree. His current research interests include low-rank matrix factorization, tensor recovery, and sparse machine learning methods.

Qian Zhao received the BSc and PhD degrees from Xi'an Jiaotong University, Xi'an, China, in 2009 and 2015, respectively. He was a visiting scholar with Carnegie Mellon University, Pittsburgh, Pennsylvania, from 2013 to 2014. He is currently a lecturer in the School of Mathematics and Statistics, Xi'an Jiaotong University. His current research interests include low-rank matrix/tensor analysis, Bayesian modeling and self-paced learning.

Deyu Meng received the BSc, MSc, and PhD degrees from Xi'an Jiaotong University, Xi'an, China, in 2001, 2004, and 2008, respectively. He was a visiting scholar with Carnegie Mellon University, Pittsburgh, Pennsylvania, from 2012 to 2014. He is currently a professor in the Institute for Information and System Sciences, Xi'an Jiaotong University. His current research interests include self-paced learning, noise modeling, and tensor sparsity.

Zongben Xu received the PhD degree in mathematics from Xi'an Jiaotong University, Xi'an, China, in 1987. He currently serves as the Academician of the Chinese Academy of Sciences, the chief scientist of the National Basic Research Program of China (973 Project), and the director of the Institute for Information and System Sciences, Xi'an Jiaotong University. His current research interests include nonlinear functional analysis and intelligent information processing. He was a recipient of the National Natural Science Award of China in 2007 and the winner of the CSIAM Su Buchin Applied Mathematics Prize in 2008. He delivered a talk at the International Congress of Mathematicians in 2010.