

Low-Rank Tensor Approximation with Laplacian Scale Mixture Modeling for Multiframe Image Denoising

Weisheng Dong
Xidian University, China
wsdong@mail.xidian.edu.cn

Guangyu Li
Xidian University, China

Guangming Shi
Xidian University, China
gmshi@xidian.edu.cn

Xin Li
West Virginia University
xin.li@ieee.org

Yi Ma
ShanghaiTech University, China
mayi@shanghaitech.edu.cn

Abstract

Patch-based low-rank models have shown effective in exploiting spatial redundancy of natural images especially for the application of image denoising. However, two-dimensional low-rank model can not fully exploit the spatio-temporal correlation in larger data sets such as multispectral images and 3D MRIs. In this work, we propose a novel low-rank tensor approximation framework with Laplacian Scale Mixture (LSM) modeling for multi-frame image denoising. First, similar 3D patches are grouped to form a tensor of d -order and high-order Singular Value Decomposition (HOSVD) is applied to the grouped tensor. Then the task of multiframe image denoising is formulated as a Maximum A Posterior (MAP) estimation problem with the LSM prior for tensor coefficients. Both unknown sparse coefficients and hidden LSM parameters can be efficiently estimated by the method of alternating optimization. Specifically, we have derived closed-form solutions for both sub-problems. Experimental results on spectral and dynamic MRI images show that the proposed algorithm can better preserve the sharpness of important image structures and outperform several existing state-of-the-art multiframe denoising methods (e.g., BM4D and tensor dictionary learning).

1. Introduction

The field of image denoising has advanced rapidly in the past decades. Early attacks on denoising are based on sparse representations of images (e.g., [17, 1]); later this line of research has evolved into image denoising methods via dictionary learning [8, 13]. The combination of dictionary learning and nonlocal means [3] has led to many state-of-the-art natural image denoising methods [5, 12, 6].

Recent advances include the low-rank matrix approximation (e.g., [7]) and high-order singular value decomposition (e.g., [18]).

By contrast, the field of multi-frame image denoising has been under-researched as of today despite the abundance of volumetric data in the real world (e.g., volume rendering, remote sensing and dynamic MRI). In ANLM3D [15], the method of nonlocal-mean filtering was made spatially adaptive for the application of MRI. The well-known BM3D method was also extended into BM4D [11] for volumetric data restoration. To fully exploit the correlations among the multi-frames, the low-rank tensor approximation methods have been proposed for multi-frame image denoising. In [19], based on the Tucker decomposition, the multirank is first estimated using the alternating least square (ALS) algorithm and the hyperspectral image (HSI) is denoised by multirank truncation of the core tensor. The parallel factor analysis (PARAFAC) decomposition has also been exploited for HSI denoising [10]. However, both the two methods didn't consider the rich nonlocal self-similarity of HSI and the ALS algorithm didn't exploit the sparse nature of the core tensor. In [16], the tensor dictionary learning method combing similar patches grouping has been proposed for HSI denoising, where the Akaike information criterion (AIC)/minimum description length (MDL) technique is used to estimate the multirank. The multirank truncation is then applied to denoise the core tensor. As the AIC/MDL technique cannot exploit the sparse nature of the core tensor, the denoising results are still often unsatisfied. In [18], the high-order singular value decomposition (HOSVD) is used to decompose the grouped similar 3D patches. The resulted core tensor is then denoised by hard-thresholding followed by Wiener filtering. Though effective, the two-stage denoising method is heuristic in nature.

In this paper, we present a novel low-rank tensor approximation framework with Laplacian Scale Mixture (LSM)

model, which is analogous to Gaussian Scale Mixture (GSM) model, for multi-frame image denoising. We first group similar 3D patches to form a tensor of d -order and apply HOSVD to the grouped tensor. Unlike [18], we propose to formulate the problem of multiframe image denoising in SVD domain as a Maximum A Posterior (MAP) estimation problem with the LSM prior for tensor coefficients. Such formulation allows us to solve the multi-frame image denoising in a more principled manner than those of [19, 16, 18] where multirank truncation and ad-hoc two-stage filtering methods were used. Compared with the classic i.i.d Laplacian model for ℓ_1 -norm sparse estimation, the LSM prior allows to jointly estimate both the unknown hidden LSM parameters and sparse coefficients from the observed data via alternating optimization. In fact, we have derived closed-form solutions for both subproblems admitting computationally efficient implementations. Experimental results have shown that the proposed multiframe image denoising method outperforms current state-of-the-art volumetric data denoising methods including tensor dictionary learning [16], BM4D [11], and the two-stage denoising method [18] significantly for multispectral images ($> 1.5dB$) and convincingly for 3D MRIs (up to $1.04dB$).

2. Related Works

In this section, we briefly review three classes of existing image denoising methods and their extensions into multi-frame scenario.

Sparse methods are built upon the fact that small patches in a natural image can often be well approximated by the linear combination of a small number of atoms from a dictionary. Instead of using fixed dictionaries, it has been shown that dictionary learning [8, 13] can substantially improve the denoising performance. The combination of the dictionary learning and nonlocal self-similarity modeling [3] leads to even more effective denoising performance [5, 12, 6]. For volumetric images, a straightforward approach of exploiting temporal redundancy is to replace 2D patches by their 3D counterpart (e.g., BM4D [11]). However, as dimensionality increases, the effectiveness of dictionary learning degrades rapidly (e.g., insufficient similar patches can be found due to the notorious curse of dimensionality).

Low-rank methods recover a clean image from its corrupted observation by low-rank matrix approximation [7]. When similar patches are grouped to form a data matrix \mathbf{Y} , the rank of \mathbf{Y} is usually low. Therefore, the task of noise removal can be achieved by recovering a low-rank matrix from its noisy version; more specifically, singular value thresholding has been a standard tool for the class of matrix completion problems [18]. By designing appropriate shrinkage/thresholding operators, state-of-the-art image denoising performance can be achieved [7]. For volumetric

data, high-order SVD becomes a natural extension for handling the grouped multi-dimensional data arrays. However, only ad-hoc thresholding or Wiener filtering techniques were considered in previous work [18].

Tensor methods have been specifically proposed for volumetric data denoising. In [19, 10], a low-rank tensor approximation method was developed for multispectral image denoising. Most recently, Peng *et al.* [16] proposed an effective multispectral image denoising method using non-local tensor dictionary learning. To exploit nonlocal spatio-temporal redundancy, 3D image patches are clustered into groups via k-means clustering. Then each set of grouped similar 3D patches are linearly approximated by low-rank tensor approximation. More specifically, the AIC/MDL criterion [21] can be used to determine the rank for each tensor model. However, the AIC/MDL criterion doesn't fully exploit the sparse nature of the core tensor, leading to inaccurate estimation of the multirank.

3. Low-rank Tensor Approximation with Laplacian Scale Mixture Modeling

In this paper, we present a novel low-rank tensor approximation framework with Laplacian Scale Mixture (LSM) modeling for multi-frame image denoising. We first introduce the nonlocal low-rank tensor approximation method, then connect it with Laplacian Scale Mixture Modeling and finally derive efficient optimization algorithms.

3.1. Nonlocal low-rank tensor approximation

Low-rank tensor approximation consists of two steps: *patch grouping* and *low-rank approximation*. First, we extract 3D patches from a noisy 3D image of size $H \times W \times L$; for each exemplar patch \mathcal{P}_i (sized by $\sqrt{n} \times \sqrt{n} \times L$ and located at spatial position i), we search for similar patches within a local window (e.g., 40×40). Specifically, we group the similar patches by k -nearest neighbor (k -NN) method. After patch grouping, we can group similar 3D patches into a 3^{rd} -order tensor by reshaping each matrix slice of 3D patches into 1D vectors, i.e., $\mathcal{Y}_i \in \mathbb{R}^{n \times m \times L}$ ¹.

Second, low-rank approximation is applied to the formed tensor \mathcal{Y}_i . Given a noisy tensor \mathcal{Y}_i , its HOSVD is given as follows [9, 2],

$$\begin{aligned} \mathcal{Y}_i &= \sum_{r=1}^n \sum_{c=1}^m \sum_{l=1}^L \tilde{\mathcal{S}}_i(r, c, l) \mathbf{u}_{i,r} \times \mathbf{v}_{i,c} \times \mathbf{w}_{i,l} \\ &= \tilde{\mathcal{S}}_i \times_1 \mathbf{U}_i \times_2 \mathbf{V}_i \times_3 \mathbf{W}_i, \end{aligned} \quad (1)$$

where $\mathbf{U}_i = [\mathbf{u}_{i,1}, \dots, \mathbf{u}_{i,n}] \in \mathbb{R}^{n \times n}$, $\mathbf{V}_i = [\mathbf{v}_{i,1}, \dots, \mathbf{v}_{i,m}] \in \mathbb{R}^{m \times m}$ and $\mathbf{W}_i = [\mathbf{w}_{i,1}, \dots, \mathbf{w}_{i,L}] \in$

¹Instead of forming a 4^{th} order tensor for the set of similar 3D patches, we found that combining them into a 3^{rd} order tensor leads to slightly better denoising performance.

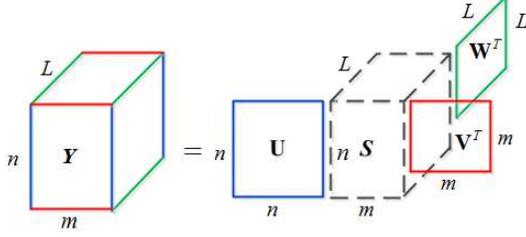


Figure 1. The HOSVD expansion of a low-rank Tensor \mathcal{Y} .

$\mathbb{R}^{L \times L}$ are orthogonal matrices, $\tilde{\mathcal{S}}_i \in \mathbb{R}^{n \times m \times L}$ is the 3D coefficient array (also called ‘‘core tensor’’), $\tilde{\mathcal{S}}_i(r, c, l)$ are the components of $\tilde{\mathcal{S}}_i$, \times denotes the tensor product (i.e., $\mathbf{x} \times \mathbf{y} = \mathbf{x}\mathbf{y}^\top$), and \times_j denotes the j -th model tensor product. Three orthogonal matrixes \mathbf{U}_i , \mathbf{V}_i and \mathbf{W}_i can be computed from the SVD of model- j ($j = 1, 2, 3$) flattening of \mathcal{Y}_i respectively. The HOSVD decomposition is illustrated in Fig.3.1. Thanks for the grouping of similar patches, \mathcal{Y}_i can be approximated by a low-rank tensor, i.e.,

$$\begin{aligned} \hat{\mathcal{X}}_i &= \sum_{r=1}^{r_1} \sum_{c=1}^{r_2} \sum_{l=1}^{r_3} \hat{\mathcal{S}}_i(r, c, l) \mathbf{u}_{i,r} \times \mathbf{v}_{i,c} \times \mathbf{w}_{i,l} \\ &= \hat{\mathcal{S}}_i \times_1 \hat{\mathbf{U}}_i \times_2 \hat{\mathbf{V}}_i \times_3 \hat{\mathbf{W}}_i, \end{aligned} \quad (2)$$

where $\hat{\mathbf{U}}_i = [\mathbf{u}_{i,1}, \dots, \mathbf{u}_{i,r_1}] \in \mathbb{R}^{n \times r_1}$, $\hat{\mathbf{V}}_i = [\mathbf{v}_{i,1}, \dots, \mathbf{v}_{i,r_2}] \in \mathbb{R}^{m \times r_2}$ and $\hat{\mathbf{W}}_i = [\mathbf{w}_{i,1}, \dots, \mathbf{w}_{i,r_3}] \in \mathbb{R}^{L \times r_3}$ are the thin matrices associated with \mathbf{U}_i , \mathbf{V}_i and \mathbf{W}_i respectively, $r_1 \leq n$, $r_2 \leq m$ and $r_3 \leq L$, and $\hat{\mathcal{S}}_i \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ denotes the smaller core tensor. The triple (r_1, r_2, r_3) is often called the *multirank* of \mathcal{Y}_i . To estimate the multirank of a tensor, the AIC/MDL method [21] can be used for different modes flattening of the tensor. With the estimated rank parameters (r_1, r_2, r_3) , low-rank tensor approximation can be easily obtained by setting the last $n - r_1$, $m - r_2$ and $L - r_3$ slices along different modes in $\tilde{\mathcal{S}}_i$ to be zero matrices.

Instead of explicitly estimating multirank parameters, we can also obtain the low-rank tensor approximation by inducing the sparsity on the array of tensor coefficients - namely

$$\begin{aligned} \hat{\mathcal{S}}_i &= \operatorname{argmin}_{\mathcal{S}_i} \psi(\mathcal{S}_i), \\ \text{s.t. } \|\mathcal{Y}_i - \mathcal{S}_i \times_1 \mathbf{U}_i \times_2 \mathbf{V}_i \times_3 \mathbf{W}_i\|_F^2 &\leq \sigma_w^2, \end{aligned} \quad (3)$$

where $\psi(\cdot)$ is a sparse regularization function inducing the sparsity to the components of \mathcal{S}_i ; and $(\mathbf{U}_i, \mathbf{V}_i, \mathbf{W}_i)$ are orthogonal matrices obtained by HOSVD of \mathcal{Y}_i . Due to the orthogonality of these matrices, Eq. (3) can be rewritten into

$$\hat{\mathcal{S}}_i = \operatorname{argmin}_{\mathcal{S}_i} \psi(\mathcal{S}_i), \text{ s.t. }, \|\tilde{\mathcal{S}}_i - \mathcal{S}_i\|_F^2 \leq \sigma_w^2, \quad (4)$$

where $\tilde{\mathcal{S}}_i = \mathcal{Y}_i \times_1 \mathbf{U}_i^\top \times_2 \mathbf{V}_i^\top \times_3 \mathbf{W}_i^\top$. The above problem

is often formulated in the Lagrangian form,

$$\hat{\mathcal{S}}_i = \operatorname{argmin}_{\mathcal{S}_i} \|\tilde{\mathcal{S}}_i - \mathcal{S}_i\|_F^2 + \lambda \psi(\mathcal{S}_i). \quad (5)$$

Popular choices of $\psi(\cdot)$ include the pseudo-norm ℓ_0 and the ℓ_1 norm, which exactly lead to the hard and soft thresholding of tensor coefficient array $\tilde{\mathcal{S}}_i$ respectively. Generally, the selection of threshold λ is non-trivial; therefore, a heuristic two-stage method similar to BM3D has been adopted in [18].

3.2. Laplacian scale mixture modeling for low-rank tensor approximation

From Eq. (5), one can see that the selection of sparsity regularization function $\psi(\cdot)$ is a critical step in low-rank tensor approximation. In this subsection, we propose a Maximum a Posterior (MAP) method for estimating \mathcal{S}_i from $\tilde{\mathcal{S}}_i$. For simplicity, we will drop the subscript index i and let $\tilde{\mathbf{s}} \in \mathbb{R}^{n \cdot m \cdot L}$ and $\mathbf{s} \in \mathbb{R}^{n \cdot m \cdot L}$ denote the one-dimensional representations of $\tilde{\mathcal{S}}$ and \mathcal{S} respectively. If \mathbf{s} denotes the noiseless version of $\tilde{\mathbf{s}}$, i.e., $\tilde{\mathbf{s}} = \mathbf{s} + \mathbf{n}$, where $\mathbf{n} \in \mathbb{R}^{n \cdot m \cdot L}$ denotes additive Gaussian noise, the MAP estimation of \mathbf{s} from $\tilde{\mathbf{s}}$ can be formulated as

$$\mathbf{s} = \operatorname{argmin}_{\mathbf{s}} \{-\log P(\tilde{\mathbf{s}}|\mathbf{s}) - \log P(\mathbf{s})\}, \quad (6)$$

where $\log P(\tilde{\mathbf{s}}|\mathbf{s})$ is given by the Gaussian distribution of noise, i.e.,

$$P(\tilde{\mathbf{s}}|\mathbf{s}) \propto \exp\left(-\frac{1}{2\sigma_w^2} \|\tilde{\mathbf{s}} - \mathbf{s}\|_2^2\right), \quad (7)$$

and *a priori* distribution of \mathbf{s} is given by

$$P(\mathbf{s}) \propto \prod_j \exp\left(-\frac{\psi(s_j)}{\theta_j}\right). \quad (8)$$

It is easy to verify that the above MAP estimation leads to the following weighted ℓ_1 norm minimization problem when $P(\mathbf{s})$ is chosen to be an independent and identically distributed Laplacian,

$$\mathbf{s} = \operatorname{argmin}_{\mathbf{s}} \|\tilde{\mathbf{s}} - \mathbf{s}\|_2^2 + 2\sqrt{2}\sigma_w^2 \sum_j \frac{1}{\theta_j} |s_j|, \quad (9)$$

where θ_j denotes the standard derivation of s_j . It has been shown that a *weighted* ℓ_1 norm is more effective than the original ℓ_1 norm in sparse estimation [4]. Now the question is how to estimate variance parameters θ_j from noisy observation $\tilde{\mathbf{s}}$.

In this paper, we propose a Laplacian Scale Mixture (LSM) prior for \mathbf{s} . With the LSM prior, one can decompose \mathbf{s} into point-wise product of a Laplacian vector α and a positive hidden scalar multiplier θ with probability $P(\theta_j)$ -

i.e., $s_j = \theta_j \alpha_j$. Note that this decomposition is analogous to the one used in Gaussian Scale Mixture model [17]. Conditioned on θ_j , s_j is Laplacian with standard deviation θ_j . Assuming that θ_j and α_j are independent, we can write the LSM prior of \mathbf{s} as

$$P(\mathbf{s}) = \prod_i P(s_i), P(s_j) = \int_0^\infty P(s_j|\theta_j)P(\theta_j)d\theta_j. \quad (10)$$

It should be noted that for many choices of $P(\theta_j)$ there is no analytic expression for $P(\mathbf{s})$. Consequently, it is usually difficult to compute the MAP estimation of \mathbf{s} with the LSM prior. However, such difficulty can be overcome by using a joint prior model $P(\mathbf{s}, \boldsymbol{\theta})$. By substituting $P(\mathbf{s}, \boldsymbol{\theta})$ into the MAP estimation of Eq.(6), we obtain

$$(\mathbf{s}, \boldsymbol{\theta}) = \underset{\mathbf{s}, \boldsymbol{\theta}}{\operatorname{argmin}} \{-\log P(\tilde{\mathbf{s}}|\mathbf{s}) - \log P(\mathbf{s}|\boldsymbol{\theta}) - \log P(\boldsymbol{\theta})\}. \quad (11)$$

Here we adopt a factorial distribution for the multipliers - specifically the noninformative Jeffrey's prior, i.e., $P(\theta_j) = \frac{1}{\theta_j}$. With Jeffrey's prior, Eq. (11) can be written into

$$(\mathbf{s}, \boldsymbol{\theta}) = \underset{\mathbf{s}, \boldsymbol{\theta}}{\operatorname{argmin}} \|\tilde{\mathbf{s}} - \mathbf{s}\|_2^2 + 2\sqrt{2}\sigma_w^2 \sum_j \frac{|s_j|}{\theta_j} + 4\sigma_w^2 \sum_j \log \theta_j. \quad (12)$$

Note that in LSM we have $\mathbf{s} = \mathbf{\Lambda}\boldsymbol{\alpha}$, where $\mathbf{\Lambda} = \operatorname{diag}(\theta_j) \in \mathbb{R}^{n \cdot m \cdot L \times n \cdot m \cdot L}$. Then Eq.(12) can be rewritten as

$$(\boldsymbol{\alpha}, \boldsymbol{\theta}) = \underset{\boldsymbol{\alpha}, \boldsymbol{\theta}}{\operatorname{argmin}} \|\tilde{\mathbf{s}} - \mathbf{\Lambda}\boldsymbol{\alpha}\|_2^2 + 2\sqrt{2}\sigma_w^2 \sum_j |\alpha_j| + 4\sigma_w^2 \sum_j \log(\theta_j + \epsilon), \quad (13)$$

where ϵ is a small positive constant for numerical stability. From Eq. (13), one can see that with the LSM prior, the sparse estimation of \mathbf{s} has been translated into the joint estimation of $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$.

3.3. Alternative Optimization

A straightforward approach of solving Eq. (13) is to adopt the method of alternating optimization. Surprisingly, both sub-problems in our formulation admit closed-form solutions, which is highly desirable from an implementation point of view.

First, given an initial estimate of $\boldsymbol{\alpha}$, one can solve for $\boldsymbol{\theta}$ by optimizing

$$\boldsymbol{\theta} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \|\tilde{\mathbf{s}} - \mathbf{A}\boldsymbol{\theta}\|_2^2 + 4\sigma_w^2 \sum_j \log(\theta_j + \epsilon), \quad (14)$$

where $\mathbf{A} = \operatorname{diag}(\boldsymbol{\alpha})$ and $\mathbf{\Lambda}\boldsymbol{\alpha} = \mathbf{A}\boldsymbol{\theta}$. Equivalently, Eq. (14) can also be rewritten as

$$\boldsymbol{\theta} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_j \{a_j \theta_j^2 + b_j \theta_j + c \log(\theta_j + \epsilon)\}, \quad (15)$$

where $a_j = \alpha_j^2$, $b_j = 2\alpha_j \tilde{s}_j$ and $c = 4\sigma_w^2$. Thus, Eq. (15) boils down to solving a sequence of scalar minimization problems

$$\theta_j = \underset{\theta_j}{\operatorname{argmin}} a_j \theta_j^2 + b_j \theta_j + c \log(\theta_j + \epsilon), \quad (16)$$

which can be solved by taking $\frac{df(\theta_j)}{d\theta_j} = 0$, where $f(\theta)$ denotes the right hand side of Eq. (16). By taking $\frac{df(\theta_j)}{d\theta_j} = 0$, two stationary points can be obtained - i.e.,

$$\theta_{j,1} = -\frac{b_j}{4a_j} + \sqrt{\frac{b_j^2}{16} - \frac{c}{2a_j}}, \theta_{j,2} = -\frac{b_j}{4a_j} - \sqrt{\frac{b_j^2}{16} - \frac{c}{2a_j}} \quad (17)$$

Case I) $b_j^2/(16a_j^2) - c/(2a_j) \geq 0$. In this case, the global minimization of Eq. (16) can be obtained by comparing $f(0)$, $f(\theta_{j,1})$ and $f(\theta_{j,2})$;

Case II) $b_j^2/(16a_j^2) - c/(2a_j) < 0$. Note that there are no stationary points in the range of $[0, \infty)$. Since ϵ is a small positive constant, $g(0) = b_j + c/\epsilon$ is always positive, where $g(\theta) = \frac{f(\theta)}{d\theta}$. Therefore, $f(0)$ is the global minimum for this case. The solution to Eq. (16) can then be written as

$$\theta_j = \begin{cases} 0, & \text{if } b_j^2/(16a_j^2) - c/(2a_j) < 0, \\ t_j, & \text{otherwise} \end{cases} \quad (18)$$

where $t_j = \underset{\theta_j}{\operatorname{argmin}} \{f(0), f(\theta_{j,1}), f(\theta_{j,2})\}$.

Second, for a fixed $\boldsymbol{\theta}$, $\boldsymbol{\alpha}$ can be updated by solving

$$\boldsymbol{\alpha} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \|\tilde{\mathbf{s}} - \mathbf{\Lambda}\boldsymbol{\alpha}\|_2^2 + 2\sqrt{2}\sigma_w^2 \sum_j |\alpha_j|, \quad (19)$$

which also admits a closed-form solution namely

$$\alpha_j = \mathcal{S}_{\tau_j} \left(\frac{\tilde{s}_j}{\theta_j} \right), \quad (20)$$

wherein $\mathcal{S}_{\tau_j}(\cdot)$ denotes the soft-thresholding function with a threshold $\tau_j = \frac{\sqrt{2}\sigma_w^2}{\theta_j^2}$.

By alternately solving the sub-problems of Eqs. (14) and (19), sparse coefficients \mathbf{s} can be estimated as $\hat{\mathbf{s}} = \hat{\mathbf{\Lambda}}\hat{\boldsymbol{\alpha}}$, wherein $\hat{\mathbf{\Lambda}}$ and $\hat{\boldsymbol{\alpha}}$ denote the estimates of $\mathbf{\Lambda}$ and $\boldsymbol{\alpha}$ respectively. Finally, the reconstructed tensor can be obtained by

$$\hat{\mathcal{X}} = \hat{\mathcal{S}} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W}, \quad (21)$$

where $\hat{\mathcal{S}}$ is the coefficient array corresponding to $\hat{\mathbf{s}}$.

4. Multiframe Image Denoising with Low-rank Tensor Approximation

Suppose noisy observation data are denoted by $\mathcal{Y} = \mathcal{X} + \mathcal{N}$, where $\mathcal{X} \in \mathbb{R}^{H \times W \times L}$ and $\mathcal{N} \in \mathbb{R}^{H \times W \times L}$ correspond to unknown image and additive noise respectively. Let $\mathcal{Y}_i =$

$\tilde{\mathcal{R}}_i \mathcal{Y}$ denote the 3^{rd} tensor formed by similar 3D patches, where $\tilde{\mathcal{R}}_i$ denotes an operator grouping similar 3D patches into a 3^{rd} tensor. It follows that the problem of denoising volumetric images can be formulated as

$$\begin{aligned} (\mathcal{X}, \{\mathcal{S}_i\}) = \operatorname{argmin}_{\mathcal{X}, \{\mathcal{S}_i\}} & \|\mathcal{Y} - \mathcal{X}\|_F^2 \\ & + \eta \sum_i \|\tilde{\mathcal{R}}_i \mathcal{X} - \mathcal{S}_i \times_1 \mathbf{U}_i \times_2 \mathbf{V}_i \times_3 \mathbf{W}_i\|_F^2 \\ & + 2\sqrt{2}\sigma_w^2 \sum_i \|\mathbf{\Lambda}_i \mathbf{s}_i\|_1 + 4\sigma_w^2 \sum_i \log \theta_i, \end{aligned} \quad (22)$$

where $\mathbf{U}_i, \mathbf{V}_i, \mathbf{W}_i$ denotes the three orthogonal matrices calculated by HOSVD. The above global minimization problem can be decomposed into the following two sub-problems and solved by the method of alternating optimization again.

4.1. Solving for whole image

Let $\hat{\mathcal{X}}_i = \hat{\mathcal{S}}_i \times_1 \mathbf{U}_i \times_2 \mathbf{V}_i \times_3 \mathbf{W}_i$ denote the reconstructed low-rank tensor with an initial estimate of \mathcal{S}_i . Then, for a fixed $\{\hat{\mathcal{S}}_i\}$, the whole image \mathcal{X} can be recovered by solving the following ℓ_2 -minimization problem

$$\mathcal{X} = \operatorname{argmin}_{\mathcal{X}} \|\mathcal{Y} - \mathcal{X}\|_F^2 + \eta \sum_{i=1}^N \|\tilde{\mathcal{R}}_i \mathcal{X} - \hat{\mathcal{X}}_i\|_F^2, \quad (23)$$

which is equivalent to the following equation (after reshaping tensors into long vectors)

$$\mathbf{x} = \operatorname{argmin}_{\mathbf{x}} \|\mathbf{y} - \mathbf{x}\|_2^2 + \eta \sum_{i=1}^N \|\tilde{\mathbf{R}}_i \mathbf{x} - \hat{\mathbf{x}}_i\|_2^2, \quad (24)$$

where $\mathbf{y} \in \mathbb{R}^{H \cdot W \cdot L}$, $\mathbf{x} \in \mathbb{R}^{H \cdot W \cdot L}$, $\hat{\mathbf{x}}_i \in \mathbb{R}^{\sqrt{n} \cdot \sqrt{n} \cdot L}$ correspond to the vector representations of tensors $\mathcal{Y}, \mathcal{X}, \hat{\mathcal{X}}_i$ respectively; and $\tilde{\mathbf{R}}_i \doteq [\tilde{\mathbf{R}}_{i0}, \tilde{\mathbf{R}}_{i1}, \dots, \tilde{\mathbf{R}}_{im-1}]$ the operator extracting similar patches. Eq.(24) can be solved in a closed-form by

$$\mathbf{x} = (\mathbf{I} + \eta \sum_{i=1}^N \tilde{\mathbf{R}}_i^T \tilde{\mathbf{R}}_i)^{-1} (\mathbf{y} + \eta \sum_{i=1}^N \tilde{\mathbf{R}}_i^T \hat{\mathbf{x}}_i), \quad (25)$$

where the matrix to be inverted is diagonal and can be easily calculated. Similar to K-SVD [8], Eq. (25) can be computed by averaging each set of reconstructed 3D patches $\hat{\mathcal{X}}_i$.

4.2. Solving for $\{\mathbf{s}_i\}$ and $\{\theta_i\}$

For a fixed \mathcal{X} , Eq. (22) reduces to a sequence of low-rank tensor approximation problems - i.e., for each exemplar 3D patch,

$$(\mathbf{s}_i, \theta_i) = \operatorname{argmin}_{\mathbf{s}_i, \theta_i} \|\tilde{\mathbf{s}}_i - \mathbf{s}_i\|_2^2 + 2\sqrt{2} \frac{\sigma_w^2}{\eta} \|\mathbf{\Lambda}_i \mathbf{s}_i\|_1 + 4 \frac{\sigma_w^2}{\eta} \log \theta_i, \quad (26)$$

where we have used $\tilde{\mathcal{S}}_i = \mathcal{X}_i \times_1 \mathbf{U}_i^T \times_2 \mathbf{V}_i^T \times_3 \mathbf{W}_i^T$. This is exactly the problem studied in the previous section.

Putting things together, the proposed volumetric image denoising algorithm based on nonlocal low-rank tensor approximation with Laplacian Scale Mixture (NLTA-LSM) is summarized in **Algorithm 1**. We have found that the inner iteration often converges in just few iterations ($J = 2$ in our implementation). In **Algorithm 1**, we have adopted the strategy of iterative regularization [20] where noise fed back to the denoised image is controlled by a small positive parameter δ .

Algorithm 1 NLTA-LSM based MDI denoising

• **Initialization:**

- (a) Set the initial estimate $\hat{\mathcal{X}} = \mathcal{Y}$ and the parameter η ;
- (b) Obtain the set of tensors $\{\mathcal{X}_i\}$ from $\hat{\mathcal{X}}$ via k -NN search for each exemplar patch.

• **Outer loop:** for $k = 1, 2, \dots, K_{max}$ **do**

(a) Tensor dataset \mathcal{X}_i construction: grouping a set of similar 3D patches into a 3^{rd} tensor for each exemplar patch;

(b) **Inner loop** (Low-rank tensor approximation by solving Eq. (26)): for $j = 1, 2, \dots, J$ **do**

(I) Compute θ_i for fixed α_i via Eq.(18);

(II) Compute α_i for fixed θ_i via Eq.(20);

(III) Output $\mathbf{s}_i = \operatorname{diag}(\theta_i) \alpha_i$ if $j = J$.

End for

(c) Reconstruct $\{\mathcal{X}_i\}$ from $\{\mathcal{S}_i\}$ via Eq.(21).

(d) Reconstruct the whole image $\hat{\mathcal{X}}^{(k+1)}$ from $\{\mathcal{X}_i\}$ by solving Eq.(25).

(e) If $k < K_{max}$ set $\hat{\mathcal{X}}^{(k+1)} = \hat{\mathcal{X}}^{(k+1)} + \delta(\mathcal{Y} - \hat{\mathcal{X}}^{(k+1)})$

End for

• **Output** $\hat{\mathcal{X}}^{(k+1)}$

5. Experimental results

We have implemented the proposed algorithm under MATLAB. Both multispectral images and MR image sequences are used to verify the performance of the proposed algorithm. To verify the performance of the propose LSM prior, we also implemented a reweighted ℓ_1 nonlocal low-rank tensor approximation (NLTA-reweighted ℓ_1) algorithm by iteratively computing θ_i in Eq. (9) as $1/(|\hat{s}_j^{(k)}| + \epsilon)$ [4], where $\hat{s}_j^{(k)}$ is the estimate of the k -th iteration and ϵ is a small constant. The basic parameters are set as follows: block size $5 \times 5 \times L$ (L denotes the number of frames), the number of similar patches $m = 100$, the regularization parameter $\delta = 0.12$, and iteration numbers $K_{max} = 7$ and $J = 2$.

5.1. Multispectral image denoising

The whole CAVE database consisting of 32 hyperspectral images is used as the test set. The images of size $512 \times 512 \times 31$ are captured with the wavelengths in the range of 400 – 700 nm at a interval of 10 nm. Two sets of experiments are conducted. In the first experiment, additive Gaussian noise with different standard derivations is added to the hyperspectral images; in the second experiment, mixed noise of additive Gaussian and Poisson is added as done in [16]. In the setting of mixed noise, standard derivations of Gaussian noise vary from 10 to 100, and the variance of Poisson noise is fixed at $y/2^k$, wherein $k = 5$. We have compared the proposed method against several recently developed multispectral image denoising methods - including tensor dictionary learning (TensorDL) method [16], BM4D method [11], PARAFAC method [10], low-rank tensor approximation (LATA) method [19], ANLM3D method [15] and band-wise BM3D method [5], the two-stage HOSVD denoising method [18].² Similar to [16], for the case of mixed noise we have applied the variance-stabilizing transformation (VST) [14] and its inverse to noisy spectral images before and after applying a test denoising method.

Average PSNR results at different noise levels are reported in Table 1. From Table 1, it can be seen that the proposed method consistently outperforms all other competing methods. The average PSNR improvements over BM4D, HOSVD [18], and NLTA-reweighted ℓ_1 methods are larger than $2dB$, $1.2dB$ and $0.8dB$, respectively. In Fig. 2 we have compared the portions of denoised images at the 410nm band of *Toy* and *Cloth* with Gaussian noise of $\sigma_w = 30$. It can be seen that the other test methods all produce visually annoying artifacts; by contrast, the denoised image by our proposed method contain much fewer artifacts. Surprisingly, we can also see that the reconstructed images by the proposed method are even sharper than the original images. This is due to the effective collaborative filtering across both the spectral and the spatial dimensions by the proposed method.

5.2. 3D MRIs denoising

We have also applied the proposed method to 3D MRIs denoising. The T_1 -weighted 3D MRIs are obtained from the Brainweb database³. The 3D MRIs is of size $217 \times 181 \times 10$ with $1 \times 1 \times 1mm^3$ resolution. Additive Gaussian noise with different noise levels σ_w is added to simulate noisy 3D MRIs⁴. The proposed method is compared with some recently developed 3D MRIs denoising methods, including the ANLM3D method [15], the band-

wise BM3D method [5], and the BM4D method [11]. For the reason of completeness, LATA [19], TensorDL [16], HOSVD [18] and NLTA-Reweighted ℓ_1 developed for spectral image denoising are also included here.

Table 2 shows the PSNR result comparison among competing methods. From Table 2 we observe that the BM4D method [11] performs much better than the TensorDL method [16]. The reason is that the correlations between the slices are not strong and smaller 3D patches (i.e., $4 \times 4 \times 4$) used in BM4D can better exploit the local correlations. Even though the full slices 3D patches (i.e., $5 \times 5 \times 10$) are used, the proposed method still outperforms the BM4D [11] for all noise levels. The PSNR gain over BM4D method can be up to 1.04 dB. To facilitate visual comparison, we have shown portions of the reconstructed MRI by different methods in Fig. 3. It can be seen that the recovered image by our method is subjectively superior to those by other methods.

6. Conclusions

In this paper we have proposed a low-rank tensor approximation approach for multiframe/volumetric image denoising. To fully exploit the spatio-temporal dependency, we group similar 3D image patches into 3^{rd} tensors, which lend themselves to be approximated by low-rank tensors. For the purpose of making low-rank tensor approximation spatio-temporally adaptive, we propose a new regularization term for sparse coefficients using Laplacian scale mixture model. LSM modeling translates low-rank tensor approximation into an optimization problem with sparse coefficients and hidden scalar variables. We have adopted the method of alternating optimization and shown that both subproblems can be solved in closed-form. Experimental results on both hyperspectral images and MRI volumetric data show that the proposed method performs significantly better than existing methods.

7. Acknowledgements

This work was supported in part by the Major State Basic Research Development Program of China (973 Program) under Grant 2013CB329402, in part by the Natural Science Foundation (NSF) of China under Grant 61471281, Grant 61227004 and Grant 61390512, in part by the Program for New Scientific and Technological Star of Shaanxi Province under Grant 2014KJXX-46, in part by the Fundamental Research Funds of the Central Universities of China under Grant BDY081424, in part by NSF under Award ECCS-0968730 and in part by the support of startup fund from ShanghaiTech.

References

- [1] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring

²We thank the authors of [16, 11, 10, 19, 15, 5] for providing their source codes in their websites.

³<http://brainweb.bic.mni.mcgill.ca/brainweb/>

⁴Our method can also be used for Rician noise by using the VST [14].

Table 1. Average PSNR results of the competing methods for different noise levels on the set of test hyperspectral images.

method	Gaussian noise				
	$\sigma_w = 10$	$\sigma_w = 20$	$\sigma_w = 30$	$\sigma_w = 50$	$\sigma_w = 100$
PARAFAC [10]	33.35	33.13	32.83	30.80	26.62
ANLM3D [15]	39.83	36.74	35.40	33.10	30.08
LRTA [19]	41.01	37.66	35.65	33.04	29.48
BwBM3D [5]	41.85	38.16	35.95	33.40	29.74
BM4D [11]	44.35	40.63	38.32	35.32	31.35
TensorDL [16]	44.13	40.67	38.66	36.03	32.43
HOSVD [18]	46.04	42.54	40.29	37.27	32.89
NLTA-Reweighted ℓ_1	45.52	42.41	40.48	37.78	34.27
Proposed NLTA-LSM	46.78	43.49	41.47	38.79	34.76
method	Mixture of Poisson and Gaussian noise with fixed $k = 4$				
	$\sigma_w = 10$	$\sigma_w = 20$	$\sigma_w = 30$	$\sigma_w = 50$	$\sigma_w = 100$
PARAFAC [10]	33.19	30.42	29.07	25.47	21.35
ANLM3D [15]	34.68	33.91	33.43	32.27	30.05
LRTA [19]	35.08	34.70	33.65	32.27	29.43
BwBM3D [5]	35.69	34.54	33.28	31.67	28.45
BM4D [11]	38.42	37.25	35.98	34.31	31.24
TensorDL [16]	38.28	37.17	36.33	34.67	31.96
HOSVD [18]	40.17	38.95	37.79	35.98	32.68
NLTA-Reweighted ℓ_1	40.17	39.37	38.46	36.79	33.82
Proposed NLTA-LSM	41.13	39.96	38.90	37.35	34.37

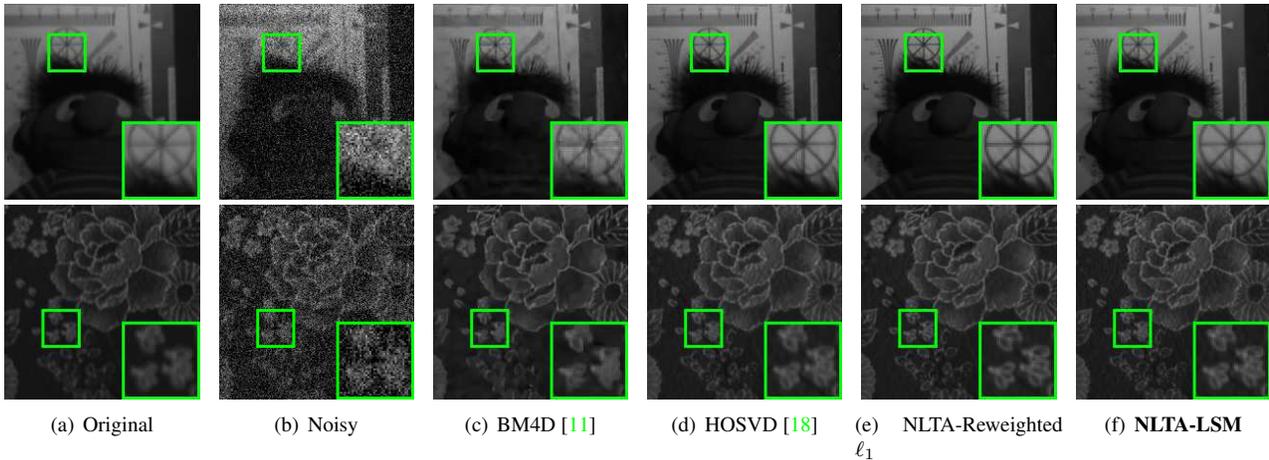


Figure 2. (a)Original images at 410nm band of *Toy* and *Cloth* in CAVE dataset [22]; (b) The images corrupted by Gaussian noise of $\sigma_w = 30$; (c) BM4D [11] (PSNR=37.49dB,32.24dB); (d) HOSVD [18] (PSNR=39.62dB,33.79dB); (e) NLTA-Reweighted ℓ_1 (PSNR=39.41dB,32.64dB); (f)**Proposed NLTA-LSM** (PSNR=**40.60dB,34.49dB**).

problems. *IEEE Trans. Image Process.*, 18(11):2419–2434, Nov. 2009. 1

[2] G. Bergqvist and E. G. Larsson. The higher-order singular value decomposition: Theory and an application. *IEEE Signal Process. Magazine*, pages 151–154, May 2010. 2

[3] A. Buades, B. Coll, and J. M. Morel. A non-local algorithm for image denoising. In *Proc. of the IEEE CVPR*, volume 2, pages 60–65, 2005. 1, 2

[4] E. Candes, M. Wakin, and S. Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fou. Ana. and App.*, 14(5):877–905, 2008. 3, 5

[5] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Trans. Image Process.*, 16(8):2080–2095, Aug. 2007. 1, 2, 6, 7, 8

[6] W. Dong, X. Li, L. Zhang, and G. Shi. Sparsity-based image denoising via dictionary learning and structural clustering. In *Proc. of the IEEE CVPR*, pages 457–464, 2011. 1, 2

[7] W. Dong, G. Shi, and X. Li. Nonlocal image restoration with bilateral variance estimation: a low-rank approach. *IEEE Trans. Image Process.*, 22(2):700–711, Feb. 2013. 1, 2

[8] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.*, 15(12):3736–3745, Dec. 2006. 1, 2,

Table 2. The PSNR results of the test methods for additive Gaussian noise on the 3D MRIs.

method	3D MRIs				
	$\sigma_w = 10$	$\sigma_w = 20$	$\sigma_w = 30$	$\sigma_w = 50$	$\sigma_w = 100$
ANLM3D [15]	32.85	29.11	27.30	25.49	23.59
LRTA [19]	32.79	28.72	27.08	24.50	21.59
BwBM3D [5]	35.42	31.81	29.78	27.33	23.94
BM4D [11]	36.37	32.85	30.89	28.49	25.39
TensorDL [16]	33.63	30.03	28.48	26.18	23.66
HOSVD [18]	36.46	33.17	31.21	28.75	25.45
NLTA-Reweighted ℓ_1	35.18	31.79	30.28	27.88	25.45
Proposed NLTA-LSM	37.06	33.62	31.83	29.53	26.41

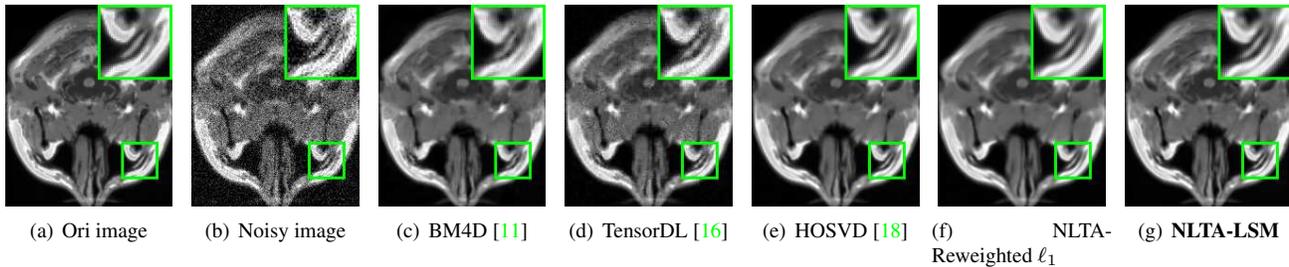


Figure 3. (a) The original MRI (the 3^{rd} slice); (b) The noisy MRI ($\sigma_w = 30$, PSNR=18.58dB); (c) BM4D [11] (PSNR=30.89dB); (d) TensorDL [16] (PSNR=28.48dB); (e) HOSVD [18] (PSNR=31.21dB); (f) NLTA-Reweighted ℓ_1 (PSNR=30.28dB); (g) **Proposed NLTA-LSM** (PSNR= 31.83 dB).

5

- [9] L. D. Lathauwer, B. D. Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM J. Matrix Anal. Applicat.*, 21(4):1253–1278, 2000. 2
- [10] X. Liu, S. Bourennane, and C. Fossati. Denoising of hyperspectral images using the parafac model and statistical performance analysis. *IEEE Trans. Geos. and Remote Sensing*, 50(10):3717C–3724, Oct. 2012. 1, 2, 6, 7
- [11] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi. Non-local transform-domain filter for volumetric data denoising and reconstruction. *IEEE Trans. Image Process.*, 22(1):119–133, Jan. 2013. 1, 2, 6, 7, 8
- [12] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. In *Proc. of the IEEE ICCV*, Tokyo, Japan, 2009. 1, 2
- [13] J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *IEEE Trans. Image Process.*, 17(1):53–69, Jan. 2008. 1, 2
- [14] M. Makitalo and A. Foi. Optimal inversion of the generalized anscombe transformation for poisson-gaussian noise. *IEEE Trans. Image Process.*, 22(1):91–103, 2013. 6
- [15] J. V. Manjon, P. Coupe, L. Marti-Bonmati, L. Collins, and M. Robles. Adaptive non-local means denoising of mr images with spatially varying noise levels. *Journal of Mag. Res. Imaging*, 31(1):192–203, 2010. 1, 6, 7, 8
- [16] Y. Peng, D. Meng, Z. Xu, C. Gao, Y. Yang, and B. Zhang. Decomposable nonlocal tensor dictionary learning for multi-spectral image denoising. In *Proc. of the IEEE CVPR*, pages 4321–4328, 2014. 1, 2, 6, 7, 8
- [17] J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli. Image denoising using scale mixtures of gaussian-s in the wavelet domain. *IEEE Trans. Signal Process.*, 12(11):1338C–1351, Nov. 2003. 1, 4
- [18] A. Rajwade, A. Rangarajan, and A. Banerjee. Image denoising using the higher order singular value decomposition. *IEEE Trans. PAMI*, 35(4):849C–862, 2013. 1, 2, 3, 6, 7, 8
- [19] N. Renard, S. Bourennane, and J. Blanc-Talon. Denoising and dimensionality reduction using multilinear tools for hyperspectral images. *IEEE GRS Letters*, 5(2):138–142, 2008. 1, 2, 6, 7, 8
- [20] D. G. J. X. S. Osher, M. Burger and W. Yin. An iterative regularization method for total variation-based image restoration. *Mult. Modeling and Sim.*, 4(2):460–489, 2005. 5
- [21] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Trans. Aco., Speech and Sig. Process.*, 33(2):387–392, 1985. 2, 3
- [22] 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.*, 19(9):2241–2253, 2010. 7