Efficient Multi-Dimensional Tensor
Sparse Coding Using t-Linear Combination

Fei Jiang,[1] **Xiao-Yang Liu,**[1,2] **Hongtao Lu,**[1] **Ruimin Shen**[1]

[1]Department of Computer Science and Engineering,
Shanghai Jiao Tong University, Shanghai, 200240, China
[2]Department of Electrical Engineering,
Columbia University, New York, NY10027, USA
Emails: jiangf@alumni.sjtu.edu.cn; xiaoyang@ee.columbia.edu; htlu@sjtu.edu.cn; rmshen@sjtu.edu.cn

Abstract

In this paper, we propose two novel multi-dimensional tensor sparse coding (MDTSC) schemes using the t-linear combination. Based on the t-linear combination, the shifted versions of the bases are used for the data approximation, but without need to store them. Therefore, the dictionaries of the proposed schemes are more concise and the coefficients have richer physical explanations. Moreover, we propose an efficient alternating minimization algorithm, including the tensor coefficient learning and the tensor dictionary learning, to solve the proposed problems. For the tensor coefficient learning, we design a tensor-based fast iterative shrinkage algorithm. For the tensor dictionary learning, we first divide the problem into several nearly-independent subproblems in the frequency domain, and then utilize the Lagrange dual to further reduce the number of optimization variables. Experimental results on multi-dimensional signals denoising and reconstruction (3DTSC, 4DTSC, 5DTSC) show that the proposed algorithms are more efficient and outperform the state-of-the-art tensor-based sparse coding models.

Introduction

As a classical unsupervised feature extraction technique, sparse coding has been successfully applied to numerous fields across computer vision (Wright et al. 2010) and pattern recognition (Lu et al. 2015; Liu et al. 2017). Traditional sparse coding represents a vector-valued signal as a linear combination of a few bases with large coefficients in an overcomplete dictionary. When dealing with multi-dimensional (MD) signals (e.g., images and videos), the traditional SC first converts the MD signals into a vector space, and then utilize the vector-based operations. Two major drawbacks of this preprocessing are: (i) the vectorization preprocess will break the intrinsic structures of the MD data and reduce the reliability of further processing; (ii) such vectorization will incur high computational costs since a vectorized MD signal will be quite long and requires the large size dictionary.

Recent research has demonstrated the superiorities of tensor-based sparse coding schemes for MD signal analysis. By maintaining the original form of MD signals, tensor-based sparse coding schemes preserve the intrinsic structures. Existing tensor-based sparse coding algorithms can be

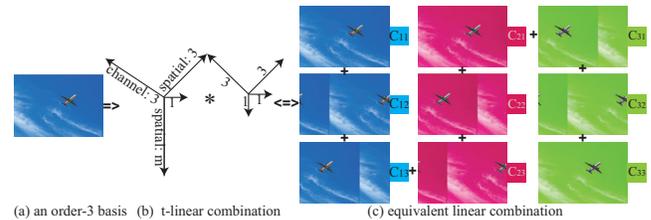(a) an order-3 basis (b) t-linear combination   (c) equivalent linear combination

Figure 1: Equivalence between the t-linear combination (one basis) and the linear combination (9 bases). The bases on the right are shifted versions of the left basis, including spatial shifting and the channel shifting.

divided into two main categories. The first category is based on Tucker decomposition (Qi et al. 2016; Xie et al. 2017). In this category, a series of separable dictionaries are utilized to approximate the structures in each dimension, so that correlations among the dimensions are not explicitly taken into consideration. The second category is based on circular convolution operations, such as CSC (Wohlberg 2014; Heide, Heidrich, and Wetzstein 2015), TCSC (Bibi and Ghanem 2017), K-TSVD (Zhang and Aeron 2016), and GTSC (Jiang et al. 2017). TCSC is an extension of CSC to high-order tensors. Though TCSC can achieve similar results as standard CSC techniques with much fewer parameters, TCSC is computationally more expensive. K-TSVD and GTSC are based on the t-linear combination (Kilmer et al. 2013), where the dictionaries can capture the inherent patterns of the MD inputs. However, K-TSVD and GTSC are proposed for dealing with order-3 tensors, not for arbitrary order tensors. Moreover, the algorithm of K-TSVD is of high computational complexity and hard to extend to large-scale high-order tensor applications (Zhang and Aeron 2016).

The basic idea that motivates us to address the above-mentioned challenges of the traditional sparse coding lies in two aspects: (i) tensor representation can preserve the intrinsic structures of the MD data; (ii) we exploit the t-linear combination under the circulant algebraic framework, and find several distinctive properties compared with the linear combination, including the small-size dictionary, shifting invariance and rich physical explanations of the tensor coefficients. Figure 1 shows the equivalence between the t-linear combination with one tensor basis and the linear combina-

tion with 9 vector bases, which means that a basis in the t-linear combination works actually as a group of bases in the linear combination. Thus, the dictionary can be much more concise under the t-linear combination. Moreover, the memory and the computational complexity are significantly reduced, especially for the high-order tensors.

To handle the aforementioned challenges, this paper proposes two novel multi-dimensional tensor sparse coding schemes using the t-linear combination under different sparsity measurements. These two schemes can simultaneously preserve the intrinsic structures of multi-dimensional data by tensor representations and reduce the computational complexity due to small-size dictionaries under the t-linear combination. The main contributions of the paper can be summarized as follows:

- We exploit the distinctive properties induced by the t-linear combination, including the small-size dictionary, shifting invariance and rich physical explanations of the tensor coefficients. All the properties support the proposed t-linear combination based sparse coding schemes to utilize more concise dictionaries for the representations of high-order tensors, leading to significantly reductions of memory and computational costs.

- We propose two novel multi-dimensional tensor sparse coding (MDTSC) schemes using t-linear combination, which extend the traditional SC to arbitrary order tensors.

- We propose an efficient and effective alternative minimization algorithm for the MDTSC with relatively low memory and computational complexity. For the tensor coefficient learning, we design a new tensor-based fast iterative shrinkage algorithm with a provable convergence rate. For the tensor dictionary learning, we divide the whole problem into several nearly-independent subproblems in the frequency domain. After that, we adopt the Lagrange dual algorithm to further reduce the number of optimization variables.

- We demonstrate that it is possible to extend our schemes to an arbitrary order tensor with experiments on 3D hyperspectral images (3DTSC, 4DTSC) and 4D colored video dataset (5DTSC).

Notations and Preliminaries

We briefly introduce the notations and preliminaries throughout the paper. We use calligraphy letters for tensors, e.g. $\mathcal{X}$, boldface capital letters for matrices, e.g. $\mathbf{X}$, boldface lowercase letters for vectors, e.g., $\mathbf{x}$, and lowercase for scalars, e.g. x. We use $[n]$ to denote the index set $\{1, 2, \cdots, n\}$.

For an order-p tensor $\mathcal{X}$ of size $n_1 \times n_2 \times \cdots \times n_p$, $\mathcal{X}^{(\ell)} \equiv \mathcal{X}(:,:,\cdots,\ell)$ is the ℓth frontal tensor, $\overrightarrow{\mathcal{X}}_j \equiv \mathcal{X}(:,j,:,\cdots,:)$ is the jth lateral tensor, $\mathcal{X}_{ij} = \mathcal{X}(i,j,:,\cdots,:)$, and $\mathcal{X}_{i_1 i_2 \cdots i_p}$ represents the $i_1 i_2 \cdots i_p$-th entry of $\mathcal{X}$. By reshaping $\mathcal{X}$ into an order-3 tensor of size $n_1 \times n_2 \times n_3 n_4 \cdots n_p$, the ℓth frontal matrix slice is denoted as $\mathcal{X}(:,:,\ell) \in \mathbb{R}^{n_1 \times n_2}$. $\widehat{\mathcal{X}}$ is the frequency domain representation of $\mathcal{X}$ that is obtained by performing Fourier transform along the dimensions $(3, \cdots, p)$, i.e., $\widehat{\mathcal{X}} = \text{fft}(\widehat{\mathcal{X}}, [\,], i)$, for $i = 3 : p$.

Some norms and operators of matrix and tensor are used. We denote the ℓ_1-norm as $\|\mathcal{X}\|_1 = \sum |\mathcal{X}_{i_1 i_2 \cdots i_p}|$, and the Frobenius norm as $\|\mathcal{X}\|_F = \sqrt{\sum \mathcal{X}_{i_1 i_2 \cdots i_p}^2}$. The spectral norm is denoted as $\|\mathbf{X}\| = \max_i \sigma_i(\mathbf{X})$, where σ_i is the ith singular value of $\mathbf{X}$. The inner product operator $\langle \mathcal{X}, \mathcal{Y} \rangle = \sum \mathcal{X}_{i_1 i_2 \cdots i_p} \mathcal{Y}_{i_1 i_2 \cdots i_p}$. The trace operator $\text{Tr}(\mathbf{X}) = \sum \mathbf{X}(i,i)$. The operators $\cdot$, $\odot$ and $\otimes$ denote the standard matrix multiplication, element-wise multiplication and the Kronecker product, respectively. The superscripts ^{T}, ^{H} and † denote the transpose, the conjugate transpose of a matrix, and the transpose of a tensor, respectively.

To construct our MDTSC schemes, it is necessary to introduce three block-based operators (Martin, Shafer, and LaRue 2013), i.e., circ($\cdot$), unfold($\cdot$), and fold($\cdot$), in advance. For $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \cdots \times n_p}$, the $\mathcal{X}^{(\ell)}$ is used to form the block circulant pattern

$$\text{circ}(\mathcal{X}) := \begin{bmatrix} \mathcal{X}^{(1)} & \mathcal{X}^{(n_p)} & \cdots & \mathcal{X}^{(2)} \\ \mathcal{X}^{(2)} & \mathcal{X}^{(1)} & \cdots & \mathcal{X}^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{X}^{(n_p)} & \mathcal{X}^{(n_p-1)} & \cdots & \mathcal{X}^{(1)} \end{bmatrix}, \quad (1)$$

and the unfolding and folding operations

$$\text{unfold}(\mathcal{X}) := \begin{bmatrix} \mathcal{X}^{(1)} \\ \mathcal{X}^{(2)} \\ \vdots \\ \mathcal{X}^{(n_p)} \end{bmatrix}, \quad \text{fold}(\text{unfold}(\mathcal{X})) = \mathcal{X}, \quad (2)$$

where $\text{circ}(\mathcal{X}) \in \mathbb{R}^{n_1 n_p \times n_2 n_p \times n_3 \times \cdots \times n_{p-1}}$, and $\text{unfold}(\mathcal{X}) \in \mathbb{R}^{n_1 n_p \times n_2 \times \cdots \times n_{p-1}}$. Two induced matrix formulations are denoted as $\mathcal{X}^c = \underbrace{\text{circ}(\cdots(\text{circ}(\mathcal{X})\cdots))}_{p-2} \in \mathbb{R}^{n_1 n_3 \cdots n_p \times n_2 n_3 \cdots n_p}$, and

$$\mathcal{X}^u = \underbrace{\text{unfold}(\cdots(\text{unfold}(\mathcal{X})\cdots))}_{p-2} \in \mathbb{R}^{n_1 n_3 \cdots n_p \times n_2}.$$

Other useful definitions from (Kilmer et al. 2013; Martin, Shafer, and LaRue 2013; Liu and Wang 2017) are given below:

Definition 1 *(Order-3 t-product) (Kilmer et al. 2013) Let $\mathcal{A}$ be $n_1 \times r \times n_3$ and $\mathcal{B}$ be $r \times n_2 \times n_3$, the t-product $\mathcal{A} * \mathcal{B}$ is the order-3 tensor of size $n_1 \times n_2 \times n_3$, where*

$$\mathcal{A} * \mathcal{B} = \text{fold}(\text{circ}(\mathcal{A}) \cdot \text{unfold}(\mathcal{B})). \quad (3)$$

Note that (3) only involves standard matrix multiplication. The following definition extends the t-product to tensors whose orders are higher than three.

Definition 2 *(Order-p t-product) (Martin, Shafer, and LaRue 2013) Let $\mathcal{A}$ be $n_1 \times r \times n_3 \times \cdots \times n_p$ and $\mathcal{B}$ be $r \times n_2 \times n_3 \times \cdots \times n_p$, the t-product $\mathcal{A} * \mathcal{B}$ is the order-p tensor of size $n_1 \times n_2 \times n_3 \times \cdots \times n_p$, where*

$$\mathcal{A} * \mathcal{B} = \text{fold}(\text{circ}(\mathcal{A}) * \text{unfold}(\mathcal{B})). \quad (4)$$

Definition 2 is recursive since (4) involves the t-product between two order-$(p-1)$ tensors.

The t-product (4) can be converted to the matrix multiplication in the frequency domain as

$$\widehat{\mathcal{C}}(:,:,\ell) = \widehat{\mathcal{A}}(:,:,\ell) \cdot \widehat{\mathcal{B}}(:,:,\ell), \ \ell \in [n_3 n_4 \cdots n_p]. \quad (5)$$
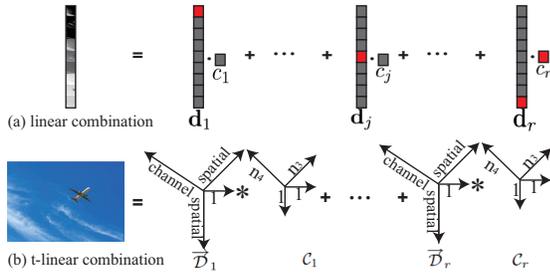
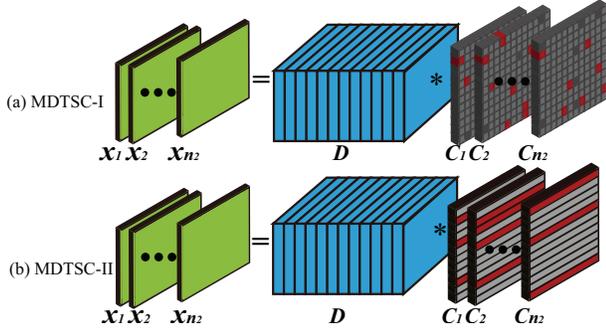Figure 2: Linear combination vs. t-linear combination.

Figure 3: MDTSC-I vs. MDTSC-II on order-3 tensors. MDTSC-II can be regarded as a group sparse formulation of MDTSC-I, where the elements of the coefficient corresponding to one basis will be activated simultaneously. Red blocks represent nonzero elements.

Definition 3 *(Tensor Transpose) Let $\mathcal{X}$ be $n_1 \times n_2 \times n_3 \times \cdots \times n_p$, its transpose $\mathcal{X}^\dagger \in \mathbb{R}^{n_2 \times n_1 \times n_3 \times \cdots \times n_p}$ is given by*

$$\mathcal{X}^{\dagger(1)} = \mathcal{X}^{(1)\dagger}, \ \mathcal{X}^{\dagger(p+2-\ell)} = \mathcal{X}^{(\ell)\dagger}, \ 2 \le \ell \le p. \quad (6)$$

Definition 4 *(t-linear Combination) Given r order-p coefficients $\{\mathcal{C}_j\}_{j=1}^r \subset \mathbb{R}^{1 \times 1 \times n_3 \times \cdots \times n_p}$, a t-linear combination of $\{\mathcal{D}_j\}_{j=1}^r \subset \mathbb{R}^{n_1 \times 1 \times n_3 \times \cdots \times n_p}$ is defined as*

$$\mathcal{D}_1 * \mathcal{C}_1 + \mathcal{D}_2 * \mathcal{C}_2 + \cdots + \mathcal{D}_r * \mathcal{C}_r. \quad (7)$$

Definition 5 *(Tensor Spectral Norm) The tensor spectral norm of $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3 \times \cdots \times n_p}$, denoted as $\|\mathcal{X}\|$, is defined as:*

$$\|\mathcal{X}\| = \max_{\ell \in [n_3 n_4 \cdots n_p]} \|\widehat{\mathcal{X}}(:,:,\ell)\|. \quad (8)$$

Multi-dimensional Tensor Sparse Coding

We first describe the motivation of using t-linear for MD data approximations. Then, we propose two novel multi-dimensional tensor sparse coding (MDTSC) schemes based on different sparse measurements. Finally, we show the advantages of MDTSC over the traditional SC, including the small-size dictionary, shifting invariance, and rich physical meanings of coefficients.

From Linear Combination to t-linear Combination

From Definition 4, t-linear combination (7) has similar expression with linear combination, but with t-product replacing the standard matrix multiplication, as shown in Figure 2. The importance of a basis for the data reconstruction is determined by the activation of the corresponding tensor coefficient.

In the following, we exploit the distinctive properties of t-linear combination. Firstly, t-linear combination is based on tensor representation, which preserves the intrinsic structure information of high-order tensors. Secondly, t-linear combination based on t-product, which can be converted into linear combination, as shown in Lemma 1.

Lemma 1 *The t-linear combination of r order-p tensor bases $\mathcal{D} = \{\overrightarrow{\mathcal{D}}_1, \cdots, \overrightarrow{\mathcal{D}}_r\} \in \mathbb{R}^{n_1 \times r \times n_3 \times \cdots \times n_p}$ is equivalent to the linear combination of $\mathbf{D} = \{\mathbf{D}_1, \cdots, \mathbf{D}_r\} \in \mathbb{R}^{n_1 n_3 \cdots n_p \times r n_3 \cdots n_p}$, i.e., $\forall \mathcal{C} \in \mathbb{R}^{r \times 1 \times n_3 \times \cdots \times n_p}$*

$$\left\{ \mathcal{X} | \mathcal{X} = \mathcal{D} * \mathcal{C} = \overrightarrow{\mathcal{D}}_1 * \mathcal{C}_1 + \cdots + \overrightarrow{\mathcal{D}}_r * \mathcal{C}_r \right\} \quad (9)$$
$$\Leftrightarrow \{ \mathbf{x} | \mathbf{x} = \mathbf{D}\mathbf{c} = \mathbf{D}_1 \mathbf{c}_1 + \cdots + \mathbf{D}_r \mathbf{c}_r \},$$

where $\mathcal{C}_j = \mathcal{C}(j, 1, :, \cdots, :) \in \mathbb{R}^{1 \times 1 \times n_3 \times \cdots \times n_p}$, $\mathbf{D}_j = \overrightarrow{\mathcal{D}}_j^c \in \mathbb{R}^{n_1 n_3 \cdots n_p \times n_3 \cdots n_p}$, and $\mathbf{c}_j = \mathcal{C}_j^u \in \mathbb{R}^{n_3 \cdots n_p}$.

Lemma 1 can be easily verified by expanding the t-product to the standard matrix multiplication with $\text{circ}(\cdot)$ and $\text{unfold}(\cdot)$ operations. From Lemma 1, we can see that one tensor basis in the t-linear combination actually represents a group of bases in the linear combination. It indicates that much fewer bases under the t-linear combination are required for a high-order tensor approximation than those under linear combination. Fewer bases usually indicates a smaller size dictionary for the data generation, which will significantly reduce the computational complexity.

Moreover, it is easy to see that the first basis of $\mathbf{D}_j$, i.e., $\mathbf{D}_j(:, 1)$, is the vectorization of the tensor basis $\overrightarrow{\mathcal{D}}_j$, and the other bases in $\mathbf{D}_j$ are the shifted versions of $\mathbf{D}_j(:, 1)$. Figure 1 explicitly shows the shifted bases for an order-3 tensor basis. It implies that under the t-linear combination, the data can be generated by the basis and its shifted versions, i.e., shifted invariance induced by the t-linear combination.

Based on the above analysis, the challenges of the traditional SC can be solved by a new data generation using t-linear combination.

Multi-dimensional Tensor Sparse Coding

To approximate MD signal $\mathcal{X}_\ell$ of size $n_1 \times n_3 \times \cdots \times n_p$ using t-linear combination, we first reshape $\mathcal{X}_\ell$ into an order-p tensor as $\overrightarrow{\mathcal{X}}_\ell \in \mathbb{R}^{n_1 \times 1 \times n_3 \times \cdots \times n_p}$. Given an overcomplete tensor dictionary $\mathcal{D} \in \mathbb{R}^{n_1 \times r \times n_3 \times \cdots \times n_p}$, $r > n_1$, $\overrightarrow{\mathcal{X}}_\ell$ can be approximated as follows:

$$\overrightarrow{\mathcal{X}}_\ell = \mathcal{D} * \mathcal{C} = \overrightarrow{\mathcal{D}}_1 * \mathcal{C}_1 + \cdots + \overrightarrow{\mathcal{D}}_r * \mathcal{C}_r, \quad (10)$$

where $\mathcal{C} \in \mathbb{R}^{r \times 1 \times n_3 \times \cdots \times n_p}$ is the tensor coefficient, and $\mathcal{C}_j = \mathcal{C}(j, 1, :, \cdots, :)$.

In the following, we propose two multi-dimensional tensor sparse coding schemes based on different sparsity measurements. On the one hand, due to the equivalence between

t-linear combination and linear combination (Lemma 1), we extend the ℓ_0 norm of the vector coefficients to the tensor coefficients, $\|\mathcal{C}\|_0$, as the first tensor coefficient sparse measurement, i.e., the number of the non-zero entries in $\mathcal{C}$. With $\|\mathcal{C}\|_0$, we can select the bases related to the input from the original tensor bases and their shifted versions. The convex relaxation $\|\mathcal{C}\|_1$ is used for easy computation. The corresponding problem of MDTSC scheme is defined as follows:

MDTSC-I Representing n_2 MD inputs as an order-p tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3 \times \cdots \times n_p}$, the first MDTSC is proposed based on the ℓ_1 norm of the coefficient, and the problem is defined as follows:

$$\min_{\mathcal{D},\mathcal{C}} \quad \frac{1}{2}\|\mathcal{X} - \mathcal{D} * \mathcal{C}\|_F^2 + \beta\|\mathcal{C}\|_1$$
$$\text{s.t.} \quad \|\overrightarrow{\mathcal{D}}_j\|_F^2 \leq 1, \; j \in [r], \qquad (11)$$

where the size of $\mathcal{D}$ is $n_1 \times r \times n_3 \times \cdots \times n_p$, $r > n_1$.

On the other hand, the tensor coefficient $\mathcal{C}_j$ has the same role as c_j in the traditional SC, as shown in Figure 2. By selecting the tensor bases related to the input, we will restrict the active numbers of the tensor coefficients. Thus, we propose the second sparsity measurement, named group sparsity, as below:

Definition 6 *Given an order-p tensor $\mathcal{C}$ of size $r \times n_2 \times n_3 \times \cdots \times n_p$, its group sparsity is defined as the number of non-zero entries of $\mathcal{C}_{ij}$, $i \in [r]$ and $j \in [n_2]$.*

Since the combinational nature of group sparsity leads to NP-hardness, we use the following norm as a convex relaxation of the group sparsity:

$$\|\mathcal{C}\|_{1,1,2} = \sum_{i=1}^{r} \sum_{j=1}^{n_2} \|\mathcal{C}_{ij}\|_F. \qquad (12)$$

Note that when $p = 3$, the group sparsity reduces to the tensor tubal sparsity in (Zhang and Aeron 2016). The corresponding problem of MDTSC scheme is defined as follows:

MDTSC-II The second MDTSC scheme is based on the group sparsity measurement, and the problem is defined as follows:

$$\min_{\mathcal{D},\mathcal{C}} \quad \frac{1}{2}\|\mathcal{X} - \mathcal{D} * \mathcal{C}\|_F^2 + \beta\|\mathcal{C}\|_{1,1,2}$$
$$\text{s.t.} \quad \|\overrightarrow{\mathcal{D}}_j\|_F^2 \leq 1, \; j \in [r]. \qquad (13)$$

Figure 3 shows the difference between MDTSC-I and MDTSC-II. MDTSC-II can be regarded as a group sparsity formulation of the MDTSC-I. In MDTSC-II, the basis $\overrightarrow{\mathcal{D}}_j$ and its shifted versions form a group, which are activated together, while in MDTSC-I, the basis and its shifted versions are separated, their activations are independent.

Advantages of MDTSC

The proposed MDTSC are not a trivial extension of the traditional sparse coding to MD data, which have novel properties, including the small-size dictionary, shifting invariance, and the rich physical explanations of the tensor coefficients.

Algorithm 1 Algorithms for the MDTSC.

Input: Input tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3 \times \cdots n_p}$, sparsity parameter β, maximum iteration: T,
1: **Initialization:** tensor dictionary $\mathcal{D} \in \mathbb{R}^{n_1 \times r \times n_3 \times \cdots n_p}$, coefficients $\mathcal{C} \in \mathbb{R}^{r \times n_2 \times n_3 \times \cdots n_p}$, and Lagrange dual variables $\lambda \in \mathbb{R}^r$,
2: **for** $t = 1$ to T **do**
3:     **//Tensor Coefficient Learning**
4:     Solving $\mathcal{C}$ by Algorithm 2,
5:     **//Tensor Dictionary Learning**
6:     Making Fourier transform for the MD data to obtain $\widehat{\mathcal{X}}$ and $\widehat{\mathcal{C}}$
7:     Solving (28) for λ by Newton's method,
8:     Calculate $\widehat{\mathcal{D}}^{(\ell)}$ from (27), $\ell \in [n_3 n_4 \cdots n_p]$,
9:     Making inverse Fourier transform of $\widehat{\mathcal{D}}$ to obtain $\mathcal{D}$.
10: **end for**
Output: $\mathcal{D}$ and $\mathcal{C}$.

Small-size Dictionary Traditional vector-based SC cannot be sufficiently handled MD inputs due to the curse of dimensionality. For a MD input of size $n_1 \times 1 \times n_3 \times \cdots \times n_p$, traditional SC requires the size of the dictionary is $(n_1 n_3 \cdots n_p) \times r_1$, $r_1 > n_1 n_3 \cdots n_p$, which increased dramatically along with the increase in dimensionality. However, using the t-linear combination, the size of the dictionary is required to be $n_1 \times r_2 \times n_3 \times \cdots \times n_p$ where $r_2 > n_1$. That is to say, for the same MD inputs, the sizes of the dictionary in the traditional SC can be $(n_3 \cdots n_p)$ times larger than that in the MDTSC. The above conclusion is obtained by the following Lemma:

Lemma 2 *The tensor space $\mathbb{R}^{n_1 \times 1 \times n_3 \times \cdots \times n_p}$ can be spanned by n_1 tensor orthogonal bases based on the t-linear combination (Kilmer et al. 2013).*

Shifting Invariance Though the size of the dictionary in the MDTSC is small, the shifted versions of the bases are involved in the t-product, which can be used for the data reconstruction without explicitly storing them, as shown in Figure 1.

On the one hand, such shifted invariance can remedy the limitations that the bases learned from the traditional SC are translated versions of each other. On the other hand, under the shifted invariance, the input and its shifted versions can be represented by the same bases, which can automatically cluster these shifted versions of the input into one group. While, for the traditional SC, the shifted versions of the input are often represented with different bases. Usually, the shifted versions of the input and the original one belong to the same class. Therefore, the MDTSC can model the shifting invariance for the classification task.

Physical Explanations In the traditional SC, the activated elements in the coefficient show the importance of the corresponding bases for the data reconstruction. Due to the equivalence between the t-linear combination and the linear combination, the activated elements in the tensor coefficient indicate the importance of the corresponding bases and their

Algorithm 2 Tensor-based Fast Iterative Shrinkage Thresholding Algorithm (TFISTA)

1: **Input:** Input tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3 \times \cdots \times n_p}$, dictionary $\mathcal{D} \in \mathbb{R}^{n_1 \times r \times n_3 \times \cdots \times n_p}$, sparsity parameter β, maximum iteration: T,
2: **Initialization:** tensor coefficients: $\mathcal{C}_0$, set $\mathcal{B}_1 = \mathcal{C}_0$, $d_1 = 1$, and Lipschitz constant $L = \|\mathcal{D}^\dagger * \mathcal{D}\|$
3: **for** $t = 1$ to T **do**
4:     Compute $\nabla f(\mathcal{B}_t)$ via (18),
5:     Compute $\mathcal{C}_t$ in the MDTSC-I or MDTSC-II via (22) and (23), respectively
6:     Update $d_{t+1} = \frac{1+\sqrt{1+4d_t^2}}{2}$,
7:     Update $\mathcal{B}_{t+1} = \mathcal{C}_t + \frac{d_t-1}{d_{t+1}}(\mathcal{C}_t - \mathcal{C}_{t-1})$,
8: **end for**
9: **Output:** Sparse coefficients $\mathcal{C} = \mathcal{C}_t$.

---

shifted versions for the data reconstruction, simultaneously.

Alternating Minimization Algorithm

Solving problems (11) and (13) are quite challenging due to the non-convex objective functions and the t-product operator. We propose an efficient algorithm by alternately optimizing $\mathcal{C}$ and $\mathcal{D}$ while fixing the other variable, as shown in Algorithm 1.

Tensor Coefficient Learning

Given the dictionary $\mathcal{D}$, find the sparse coefficient $\mathcal{C}$ via solving

$$\min_{\mathcal{C}} \frac{1}{2}\|\mathcal{X} - \mathcal{D} * \mathcal{C}\|_F^2 + \beta \|\mathcal{C}\|_1. \quad (14)$$

First, we show that (14) is a special case of the following convex minimization problem:

$$\min_{\mathcal{C}} F(\mathcal{C}) := f(\mathcal{C}) + \beta g(\mathcal{C}), \quad (15)$$

Where

- $f(\mathcal{C}) = \frac{1}{2}\|\mathcal{X} - \mathcal{D} * \mathcal{C}\|_F^2$ is Lipschitz continuous and convex, as shown in Lemma 3.

- $g(\mathcal{C})$ is 'simple', i.e., the proximal operation of $g(\mathcal{C})$ is easy to calculate (Parikh and Stephen 2014).

Then, we design a tensor-based fast iterative shrinkage thresholding algorithm (TFISTA) for solving (14) with $O(1/t^2)$ convergence rate where t is the iteration count, shown in Algorithm 2. Similar to FISTA (Beck and Teboulle 2009), at each iteration $t = 0, 1, 2, \cdots$,

$$\mathcal{C}^{t+1} = \arg\min\{f(\mathcal{C}^t) + \langle \nabla f(\mathcal{C}^t), \mathcal{C} - \mathcal{C}^t \rangle +$$
$$\frac{L}{2}\|\mathcal{C} - \mathcal{C}^t\|_F^2 + \beta g(\mathcal{C})\} \quad (16)$$
$$= \text{prox}_{r_t,g}\left(\mathcal{C}^t - r_t \nabla f(\mathcal{C}_t)\right),$$

and

$$\mathcal{C}^{t+1} = \mathcal{C}^{t+1} + \frac{d_t - 1}{d_{t+1}}(\mathcal{C}^{t+1} - \mathcal{C}^t). \quad (17)$$

where L is the Lipschitz constant, $r_t = \beta/L$, $\text{prox}_{r_t,g}(\cdot)$ is the proximal operator (Parikh and Stephen 2014), $d_1 = 1$ and $d_{t+1} = \frac{1+\sqrt{1+4d_t^2}}{2}$.

The gradient $\nabla f(\mathcal{C})$ can be calculated as follows:

$$\nabla f(\mathcal{C}) = \mathcal{D}^\dagger * \mathcal{D} * \mathcal{C} - \mathcal{D}^\dagger * \mathcal{X}, \quad (18)$$

and L can be calculated by Lemma 3:

Lemma 3 $f(\mathcal{C}) = \frac{1}{2}\|\mathcal{X} - \mathcal{D} * \mathcal{C}\|_F^2$ is Lipschitz continuous with the Lipschitz constant $L = \|\mathcal{D}^\dagger * \mathcal{D}\|$.

We prove Lemma 3 by the definition of the Lipschitz constant. For $\forall \mathcal{C}$ and $\mathcal{B} \in \mathbb{R}^{r \times 1 \times n_3 \times \cdots \times n_p}$,

$$\|\nabla f(\mathcal{C}) - \nabla f(\mathcal{B})\|_F = \|\mathcal{D}^\dagger * \mathcal{D} * (\mathcal{C} - \mathcal{B})\|_F$$
$$= \|(\mathcal{D}^\dagger \mathcal{D})^c \cdot (\mathcal{C}^u - \mathcal{B}^u)\|_2 \quad (19)$$
$$\leq \|(\mathcal{D}^\dagger \mathcal{D})^c\|_2 \|\mathcal{C}^u - \mathcal{B}^u\|_2 = \|\mathcal{D}^\dagger * \mathcal{D}\|_2 \|\mathcal{C} - \mathcal{B}\|_F.$$

Since the block circulant matrix $(\mathcal{D}^\dagger \mathcal{D})^c$ can be diagonalized by using the Fourier transform, let $\mathcal{A} = \mathcal{D}^\dagger * \mathcal{D}$, there is

$$\text{diag}(\widehat{\mathcal{A}}) = (\mathbf{F}_{n_p} \otimes \mathbf{F}_{n_{p-1}}) \cdots \otimes \mathbf{F}_r \otimes \mathbf{I}_r) \cdot$$
$$\mathcal{A}^c \cdot (\mathbf{F}_{n_p} \otimes \mathbf{F}_{n_{p-1}}) \cdots \otimes \mathbf{F}_r \otimes \mathbf{I}_r)^H, \quad (20)$$

where $\mathbf{F}_{n_i}$ is the $n_i \times n_i$ discrete Fourier transform matrix, and $\mathbf{I}_r$ is an $r \times r$ identity matrix, and the operator $\text{diag}(\cdot)$ diagonalizes the blocks of $\widehat{\mathcal{A}}(:,:,\ell)$, $\ell \in [n_3 n_4 \cdots n_p]$.

Due to the unitary invariance of the spectral norm, we obtain the Lipschitz constant as follows:

$$L = \|\mathcal{A}^c\| = \|\text{diag}(\widehat{\mathcal{A}})\| = \|\mathcal{D}^\dagger * \mathcal{D}\|. \quad (21)$$

Finally, we present the solution of (16) with different sparsity regularization $g(\mathcal{C})$. Let $\mathcal{T}$ be $\mathcal{C}^t - r_t \nabla f(\mathcal{C}^t)$. If $g(\mathcal{C}) = \|\mathcal{C}\|_1$, $\mathcal{C}^{t+1}$ can be obtained by the soft-thresholding operator

$$\mathcal{C}^{t+1} = \text{sign}(\mathcal{T}) \odot \max\{|\mathcal{T}| - r_t, 0\}, \quad (22)$$

if $g(\mathcal{C}) = \|\mathcal{C}\|_{1,1,2}$, $\mathcal{C}^{t+1}$ can be updated by

$$\mathcal{C}_{ij}^{t+1} = \mathcal{T}_{ij} \odot \max\left\{1 - \frac{r_t}{\|\mathcal{T}_{ij}\|_F}\right\}. \quad (23)$$

Tensor Dictionary Learning

For learning the dictionary $\mathcal{D}$ while fixing $\mathcal{C}$, the optimization problem is:

$$\min_{\mathcal{D}} \quad \|\mathcal{X} - \mathcal{D} * \mathcal{C}\|_F^2$$
$$\text{s.t.} \quad \|\mathcal{D}_j\|_F^2 \leq 1, \ j \in [r], \quad (24)$$

where bases are coupled together. Therefore, we firstly decompose (24) into k, $k = n_3 n_4 \cdots n_p$, nearly-independent problems in the frequency domain as follows:

$$\min_{\widehat{\mathcal{D}}^{(\ell)}, \ell \in [k]} \quad \sum_{\ell=1}^{k} \|\widehat{\mathcal{X}}^{(\ell)} - \widehat{\mathcal{D}}^{(\ell)}\widehat{\mathcal{C}}^{(\ell)}\|_F^2$$
$$\text{s.t.} \quad \sum_{\ell=1}^{k} \|\widehat{\mathcal{D}}^{(\ell)}(:,j)\|_F^2 \leq k, j \in [r]. \quad (25)$$

Table 1: Complexity analysis of the tensor coefficient learning (TCL) and the tensor dictionary learning (TDL) for the TenSR, TCSC, K-TSVD and the proposed MDTSC.

	TenSR	TCSC	MDTSC	KTSVD	3DTSC-II
TCL	$O(r_1^2 r_2 r_3 + r_1 r_2^2 r_3 + r_1 r_2 r_3^2)$	$O(n_2 n_3 r^3) + O(n_1 n_2 n_3 \log n_2 n_3)$	$O(n_1 n_2 n_3 r + n_1 n_2 n_3 \log n_2 n_3)$	$O(r^3 n_2 n_3 + n_3 \log n_3)$	$O(r n_1 n_2 n_3 + n_3 \log n_3)$
TDL	$O(\sum_{\ell=1}^{3} \sum_{k \neq \ell, k=1}^{3} \prod_{i=1}^{k} r_i \prod_{j=k}^{3} n_j n_\ell)$	$O(r^3 n_2 n_3 + n_1 n_2 n_3 \log n_2 n_3)$	$O(r^2 n_2 n_3 + n_1 n_2 n_3 \log n_2 n_3)$	$O(r^2 n_1 n_2 n_3 + r n_1^3 n_2 n_3 + r n_3 \log n_3)$	$O(r^2 n_2 n_3 + n_3 \log n_3)$

Table 2: Average performance of 6 competing methods with respect to PSNR and SSIM measurements. The results are obtained by averaging through the 32 scenes of the Columbia MSI database.

| Method	PSNR (dB)					SSIM				
	$\sigma = 5$	$\sigma = 10$	$\sigma = 20$	$\sigma = 30$	$\sigma = 50$	$\sigma = 5$	$\sigma = 10$	$\sigma = 20$	$\sigma = 30$	$\sigma = 50$
Noisy Image	34.15	28.13	22.11	18.58	14.15	0.8932	0.7599	0.6117	0.5324	0.4462
BwK-SVD	37.79	34.11	30.99	29.34	27.35	0.8873	0.7854	0.6571	0.5727	0.4614
3DK-SVD	39.47	36.33	33.47	31.08	29.63	0.9199	0.8612	0.7927	0.7457	0.6761
LRTA	44.25	41.07	37.75	35.79	33.2	0.9711	0.9483	0.9089	0.8762	0.8186
PARAFAC	32.77	32.72	32.48	32.15	30.22	0.8368	0.8344	0.8235	0.8052	0.7051
TenSR	43.74	39.05	35.01	33.31	31.38	0.975	0.9264	0.8473	0.7837	0.7778
4DTSC-I (ours)	**44.91**	**41.74**	**38.35**	**36.20**	**33.58**	**0.9838**	**0.9683**	**0.9343**	**0.8850**	**0.8220**
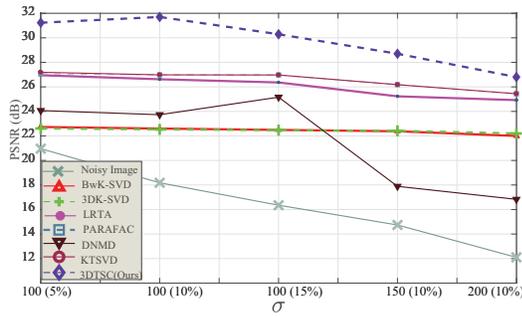
Figure 4: Denoising performances (PSNR) of the 3DTSC-II on chart and stuffed toy images in the Columbia MSI databases, where $100(5\%)$ denotes $\sigma = 100$ with the sparsity of the noise pixels 5%.

Then, we adopt the Lagrange dual (Lee et al. 2007) to solve the dual variables by Newton's algorithm. Another advantage of Lagrange dual is that the number of optimization variables is r, which is much smaller than $n_1 n_3 \cdots n_p r$ of the primal problem (24). Introducing Lagrange variable $\lambda \in \mathbb{R}^r \geq 0$, we first consider the Lagrangian of (25):

$$\mathcal{L}_{\text{prime}}(\widehat{\mathcal{D}}, \lambda) = \sum_{\ell=1}^{k} \|\widehat{\mathcal{X}}^{(\ell)} - \widehat{\mathcal{D}}^{(\ell)} \widehat{\mathcal{C}}^{(\ell)}\|_F^2 +$$
$$\sum_{j=1}^{r} \lambda_j \left(\sum_{\ell=1}^{k} \|\widehat{\mathcal{D}}^{(\ell)}(:,j)\|^2 - k \right) \quad (26)$$

Secondly, minimizing over $\widehat{\mathcal{D}}$ analytically, we obtain the optimal solution of $\widehat{\mathcal{D}}^{(\ell)}$, $\ell \in [k]$:

$$\widehat{\mathcal{D}}^{(\ell)} = \left(\widehat{\mathcal{X}}^{(\ell)} \widehat{\mathcal{C}}^{(\ell)H} \right) \left(\widehat{\mathcal{C}}^{(\ell)} \widehat{\mathcal{C}}^{(\ell)H} + \text{diag}(\lambda) \right)^{-1}. \quad (27)$$

Substituting (27) back into the Lagrangian primary function $\mathcal{L}_{\text{prime}}(\widehat{\mathcal{D}}, \lambda)$, we obtain the Lagrange dual function $\mathcal{L}_{\text{dual}}(\lambda)$:

$$\mathcal{L}_{\text{dual}}(\lambda) = -\sum_{\ell=1}^{k} \text{Tr} \left(\widehat{\mathcal{D}}^{(\ell)H} \widehat{\mathcal{X}}^{(\ell)} \widehat{\mathcal{C}}^{(\ell)H} \right) - k \sum_{j=1}^{r} \lambda_j. \quad (28)$$

The optimal solution of (28) can be obtained by Newton's method or conjugate gradient. Once getting λ, the dictionary can be recovered by (27).

Complexity Analysis

We analyze the computational complexities of our algorithms compared with the recently proposed state-of-the-art tensor-based sparse coding methods, including TCSC (Bibi and Ghanem 2017), TenSR (Qi et al. 2016) and K-TSVD (Zhang and Aeron 2016), where TCSC, TenSR and our MDTSC can be applied to arbitrary order tensors, while K-TSVD only for order-3 tensors.

To compare with the TCSC and TenSR, we consider the order-4 sparse coding, where the input is an order-3 tensor of size $n_1 \times n_2 \times n_3$. The size of the dictionary in TCSC and our MDTSC is set to $n_1 \times r \times n_2 \times n_3$, where $r > n_1$. Three dictionaries are used in the TenSR, and the sizes are set to $n_i \times r_i$, where $r_i > n_i$, $i \in [3]$. To compare with K-TSVD, we consider the order-3 sparse coding, where the input is $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and the dictionary $\mathcal{D} \in \mathbb{R}^{n_1 \times r \times n_3}$, $r > n_1$.

The computational complexities are shown in Table 1. As can be seen that the computational complexity of the proposed algorithm is significantly reduced. With fixed the input size, the time complexity of the proposed tensor coefficient learning algorithm (TFISTA) is linear in the number of bases, while others are all polynomial time complexities.

Experimental Results

In this section, we demonstrate the effectiveness and efficiency of the proposed MDTSC schemes on three different experiments, two on MSI denoising (3DTSC and 4DTSC) and one on colored video reconstruction (5DTSC).

Multi-spectral Image Denoising

we evaluate the performance of the proposed MDTSC schemes using 3D MSI images in Columbia MSI database (Wang et al. 2004)[1] by adding two kinds of commonly existing noises in MSIs. The MSI dataset contains 32 scenes at

---

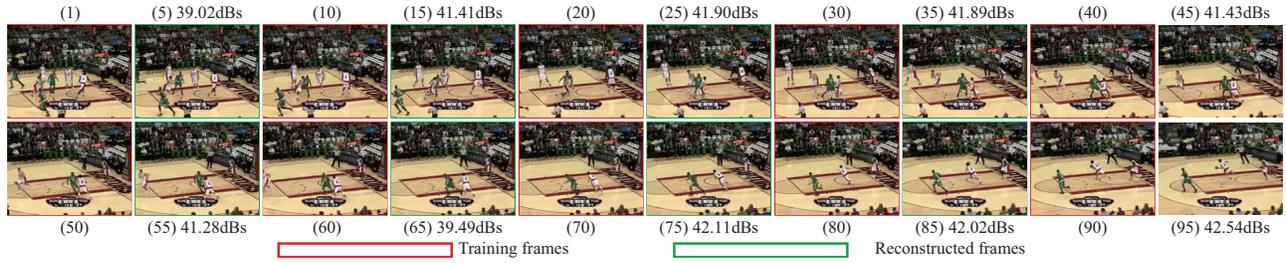[1] http://www1.cs.columbia.edu/CAVE/databases/multispectral/

Figure 5: Color-coded frames of the basketball video for training, along with the reconstructed frames. The number in brackets indicates the frame ID. We also report the reconstruction PSNR for each of the reconstructed frames in green.
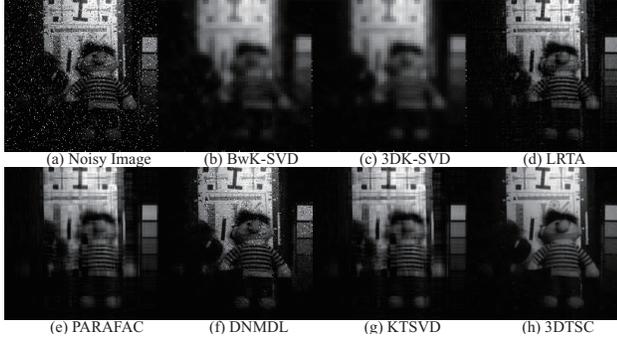

Figure 6: Visual comparison of denoising results at 610nm band of chart and stuffed toy. The sparsity of the noisy pixels is 15% and the noise with $\sigma = 100$.

a spatial resolution of 512×512 and a spectral resolution of 31 ranging from 400nm to 700nm.

We compare our MDTSC schemes with state-of-the-art MSI denoising methods, including band-wise K-SVD (BwK-SVD) (Aharon, Elad, and Bruckstein 2006), 3D-cube K-SVD (Elad and Aharon 2006), LRTA (Renard, Bourennane, and Blanc-Talon 2008), PARAFAC (Liu, Bourennane, and Fossati 2012), DNMDL (Peng et al. 2014), TenSR (Qi et al. 2016), and K-TSVD (Zhang and Aeron 2016). PSNR and SSIM are two metrics for denoising, which evaluate the similarity between the denoised images and the reference ones based on MSE and structural consistency, respectively.

3DSC Denoising In the first experiment, we consider the fixed-location defects without knowing the noisy positions as (Zhang and Aeron 2016). Let Ω indicates the set of the noisy pixel locations, then for each $(i, j) \in \Omega$ and $\ell \in [n_3]$, $\mathcal{X}(i, j, \ell) = \mathcal{X}(i, j, \ell) + \mathcal{W}(i, j, \ell)$, where $\mathcal{X}$ is the clean tensor and $\mathcal{W}(i, j, \ell) \sim \mathcal{N}(0, \sigma)$ is the additive Gaussian noise with standard deviation σ. We randomly extract $10,000$ overlapping patches of size $8 \times 8 \times 10$ from the noisy MSIs. The size of the dictionary is set to $64 \times 256 \times 10$, the same as (Zhang and Aeron 2016).

Figure 4 shows the quantitative comparison results in terms of PSNR, where the proposed 3DSC-II significantly outperforms the other competing algorithms for high-level noises ($\sigma = 100, 150$, and 200). The visual denoising result is shown in Figure 6. Compared with other algorithms, the

3DSC-II exhibits clearer details in texture regions or edges, meanwhile produces cleaner results in smooth regions with higher PSNR value.

4DSC Denoising In the second experiment, we add Guassian white noise at different noise levels $\sigma = 5, 10, 20, 30, 50$. We extract $5 \times 5 \times 5$ cubes with overlap of 3 pixels between adjacent cubes. The size of the dictionary is set to $5 \times 10 \times 5 \times 5$.

Table 2 show the comparison results in terms of average PSNR and SSIM among the 32 scenes of the Columbia MSI dataset. We can see that the proposed 4DSC-I achieves the best denoising performance compared with the competing algorithms.

Colored Video Reconstruction

In the third experiment, we show how the proposed scheme handles an arbitrary order tensor on colored video reconstruction. We use the basketball video from the OTB50 dataset (Wu, Lim, and Yang 2013). From this video, we select 10 frames ($i \in [1, 10, 20, \cdots, 90]$), saving as an order-4 tensor of size $432 \times 576 \times 3 \times 10$, for the MD dictionary learning and reconstruct 10 intermediate testing frames at $i \in [5, 15, \cdots, 95]$. We randomly extract overlapping patches of size $8 \times 8 \times 3 \times 10$. The size of the dictionary is set to $8 \times 16 \times 8 \times 3 \times 10$.

Figure 5 shows example reconstructions of test frames after training on other frames in the same video. As can be seen that the dictionary learned from the 5DSC-I can capture the correlations across frames and the three color channels simultaneously. We achieve a better reconstruction results (41.31dBs) on average among the 10 frames with the TCSC (40.54dBs). While the dictionary size of the TCSC (Bibi and Ghanem 2017), $3 \times 100 \times 100 \times 100 \times 10$, is almost 1000 times larger than ours.

Conclusion

In this paper, we have proposed two novel multi-dimensional tensor sparse coding schemes based on the t-linear combinations to preserve the inherent structures and capture the features of the MD data. We explore the advantages of the proposed MDTSC over the traditional SC, including the small-size dictionary, shifting invariance, and rich physical explanations of the tensor coefficients. Moreover, an efficient alternating minimization algorithm is proposed. The effective-

ness and efficiency of the proposed MDTSC are demonstrated by MD signals denoising and reconstruction problems.

Acknowledgements

References

Aharon, M.; Elad, M.; and Bruckstein, A. 2006. rmk-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing* 54(11):4311–4322.

Beck, A., and Teboulle, M. 2009. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences* 2(1):183–202.

Bibi, A., and Ghanem, B. 2017. High order tensor formulation for convolutional sparse coding. In *Proceedings of the IEEE International Conference on Computer Vision*.

Elad, M., and Aharon, M. 2006. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing* 15(12):3736–3745.

Heide, F.; Heidrich, W.; and Wetzstein, G. 2015. Fast and flexible convolutional sparse coding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5135–5143.

Jiang, F.; Liu, X.; Lu, H.; and Shen, R. 2017. Graph regularized tensor sparse coding for image representation. In *Proceedings of the IEEE Conference on Multimedia and Expo*, 67–72.

Kilmer, M.; Braman, K.; Hao, N.; and Hoover, R. 2013. Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. *SIAM Journal on Matrix Analysis and Applications* 34(1):148–172.

Lee, H.; Battle, A.; Raina, R.; and Ng, A. Y. 2007. Efficient sparse coding algorithms. In *Advances in Neural Information Processing Systems*, 801–808.

Liu, X., and Wang, X. 2017. Fourth-order tensors with multidimensional discrete transforms. *arXiv preprint arXiv:1705.01576*.

Liu, F.; Wang, S.; Rosenberger, J.; Su, J.; and Liu, H. 2017. A sparse dictionary learning framework to discover discriminative source activations in eeg brain mapping. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, 1431–1437.

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

Lu, Z.; Gao, X.; Wang, L.; Wen, J.; and Huang, S. 2015. Noise-robust semi-supervised learning by large-scale sparse coding. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, 2828–2834.

Martin, C. D.; Shafer, R.; and LaRue, B. 2013. An order-p tensor factorization with applications in imaging. *SIAM Journal on Scientific Computing* 35(1):A474–A490.

Parikh, N., and Stephen, B. 2014. Proximal algorithms. *Foundations and Trends® in Optimization* 1(3):127–239.

Peng, Y.; Meng, D.; Xu, Z.; Gao, C.; Yang, Y.; and Zhang, B. 2014. Decomposable nonlocal tensor dictionary learning for multispectral image denoising. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2949–2956.

Qi, N.; Shi, Y.; Sun, X.; and Yin, B. 2016. TenSR: Multi-dimensional tensor sparse representation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5916–5925.

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

Wang, Z.; Bovik, A.; Sheikh, H.; and Simoncelli, E. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* 13(4):600–612.

Wohlberg, B. 2014. Efficient convolutional sparse coding. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing*, 7173–7177. IEEE.

Wright, J.; Ma, Y.; Mairal, J.; Sapiro, G.; Huang, T.; and Yan, S. 2010. Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE* 98(6):1031–1044.

Wu, Y.; Lim, J.; and Yang, M.-H. 2013. Online object tracking: A benchmark. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2411–2418.

Xie, Q.; Zhao, Q.; Meng, D.; and Xu, Z. 2017. Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Zhang, Z., and Aeron, S. 2016. Denoising and completion of 3d data via multidimensional dictionary learning. In *International Joint Conference on Artificial Intelligence*.