Multi-modal Core Tensor Factorization based Low-Rankness and Its Applications to Tensor Completion

Haijin Zeng, Jize Xue, Hięp Q. Luong and Wilfried Philips

Abstract—Low-rank tensor completion has been widely used in computer vision and machine learning. This paper develops a novel multi-modal core tensor factorization (MCTF) method combined with a tensor low-rankness measure and a better nonconvex relaxation form of this measure (NC-MCTF). The proposed models encode low-rank insights for general tensors provided by Tucker and T-SVD, and thus are expected to simultaneously model spectral low-rankness in multiple orientations and accurately restore the data of intrinsic low-rank structure based on few observed entries. Furthermore, we study the MCTF and NC-MCTF regularization minimization problem, and design an effective block successive upper-bound minimization (BSUM) algorithm to solve them. This efficient solver can extend MCTF to various tasks, such as tensor completion. A series of experiments, including hyperspectral image (HSI), video and MRI completion, confirm the superior performance of the proposed method.

Index Terms—Tensor, Low-rankness, Tensor factorization, Nonconvex optimization.

I. INTRODUCTION

Low-rankness is a common attribute of many data sources. To date, methods based on low-rankness have reported empirical and theoretical success on a large variety of scientific and engineering applications: face modeling [1], gene categorization [2], camera image processing [3], compressive imaging [4], user interest prediction [5], etc.

A promising method that measures the low-rankness of a matrix is to account for the number of non-zero singular values [6]. This low-rankness metric and its relaxations (for example, the $L_1$ norm and nuclear norm) have been proven useful as regularisation terms in applications, and have inspired various low-rank models and algorithms to cope with different tasks. On the other hand, a large amount of data generated by modern sensors is naturally represented by high-order tensors, whereas the SVD is restricted to 2D data.

Early high-dimensional data analysis methods reformatted high-dimensional data tensors artificially into 2D matrices and resorted to methods developed for classic two-dimensional analysis methods. However, this flattening strategy and the strict assumptions inherent in two-dimensional analysis do not always match the high-dimensional data well. For example, hyperspectral image (HSI) is the imaging result of different spectral bands from the same spatial scene, which indicates that there is a high correlation in the spectral dimensions [7]; a video contains multiple frames, which contain a high correlation in the temporal dimension, especially for adjacent frames [8]. Therefore, converting these high-dimensional tensors data artificially to 2-D matrices spreads this “local correlation” (e.g., between adjacent frequencies) over large strides in the 2D matrix, complicating analysis. Thanks to the inherent high-dimensional structure of the data, high-order tensor decomposition allows capturing correlation in a more local fashion along each dimension [9]. In other words, only when analyzing the existing inherent multi-dimensional patterns, we are able to discover the hidden components in the high-dimensional data, so as to model the data more accurately.

The tensor is the generalization of the matrix and vector concept: a vector is a first-order or one-way tensor, and a matrix is a second-order tensor. To measure the low-rankness of tensors, a lot of the current work is to decompose the tensor into a combination of several factors to explore its low-rank structure through preliminary tensor decomposition, or to unfold the target tensor into matrices according to the modal, and then directly applies the rank of matrix or the sparsity of vector to the resulting matrices. Popular decompositions include Tucker [10], Canonical Polyadic (CP) [11] and tensor SVD (t-SVD) [12]. In addition, there are also some models that further improve the above decomposition or approximation methods. For example, some works impose specific prior constraints on the factors obtained by these decompositions [6, 13].

According to the well-established theory of rank function in matrix case, it seems natural to directly extend matrix completion methods to the tensor completion problem. However, it has been proven that calculating such a tensor rank (whether it is based on Tucker, CP or T-SVD decomposition) is an NP-hard problem [14]. It is difficult to determine or even limit the rank of arbitrary tensors compared to matrix rank, due to tensor low-rankness insight should be explained beyond the low-rank properties of all its expanded subspaces, and more importantly, how these subspace low-rank properties are related to the entire tensor structure should also be considered [9]. Some of current works directly extend the rank of a matrix to higher-order by simply summing up ranks (or its relaxations) along all tensor modes [15–17], or directly consider the global low-rankness of underlying tensors. Different from the matrix scenarios, the simple rank summation term is generally short...
of a clear physical meaning for tensors [6]. Furthermore, when the sampling rate is very low, it is also not sufficient to explore only global low-rankness. As shown in Fig. 1, after exploring the global low-rank prior through tensor decomposition, for the factor obtained by the decomposition, instead of being identical global low-rankness, the low-rankness in different modes or orientations are evidently existed and different. Actually, from Fig. 1-(B-I), one can see that most singular values of the factor matrices are very close to zero, and much smaller than the first several larger singular values. Moreover, instead of being independent, as shown in Fig. 1-(B-III), there are apparent correlations across different slices of the each mode of the factor tensor and matrix.

In this paper, for the tensor completion problem, we propose a novel tensor low-rankness measure to effectively model multi-modal low-rankness of high-order tensors. Similar to Tucker, our method is also based upon the tensor and matrix decomposition definitions. But instead of using the tensor Tucker decomposition directly, and require the components to be orthogonal, firstly, a novel tensor decomposition is proposed, in which a high-order low-rank decomposition is introduced into each mode of the underlying tensor, but the factors is not required to be orthogonal, as shown in the Fig. 1(A)-(B). Instead of utilizing only one mode low-rankness of the underlying tensor as tensor nuclear norm based on T-SVD, this decomposition utilize all mode low-ranknesses of the tensor to give much better performance. Compared with Tucker decomposition, our method do not require the components to be orthogonal, thus, no need to use the SVD in our decomposing algorithms, which is computationally much cheaper than Tucker and T-SVD.

Secondly, a tensor low-rankness measure based on the proposed decomposition is proposed, which combines both the low-rank prior of the global tensor and the local factors obtained by the proposed tensor decomposition method. Its insight can be easily interpreted as a regularization for the factor tensor and matrix derived from the non-orthogonal Tucker decomposition. Furthermore, an alternative convex relaxation of the proposed low-rankness measure is presented. Such measure not only unifies the traditional understanding of low-rankness from matrix to tensor, but also encodes both sparsity insights delivered by common Tucker, SVD and T-SVD low-rank decompositions for a general tensor.

Thirdly, we applied the proposed low-rank measures to high-dimensional tensor completion tasks, e.g., video, hyperspectral image, MRI completion, and designed a block successive upper-bound minimization (BSUM) method to efficiently solve the resulting models. The validity of the proposed models are evaluated on a series of experiments, including video, MRI and hyperspectral image completion.

II. NOTIONS AND PRELIMINARIES

In this section, we summarized some notations, tensor operations and operators used in this paper.

Table I summarizes the common notations throughout this paper. Following [18, 19], a fiber of tensor \( \mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3} \) is defined as a vector obtained by fixing all indices of \( \mathcal{X} \) except one, and a slice of \( \mathcal{X} \) is defined as a matrix by fixing all indices of \( \mathcal{X} \) except two. \( \mathcal{X} \) denotes the result of discrete Fourier transformation (FFT) of \( \mathcal{X} \) along the 3-rd dimension, and \( \mathcal{X} \) can be computed from \( \hat{\mathcal{X}} \) via the inverse FFT. Then the multi-rank of \( \mathcal{X} \) is defined as the array \( \text{rank}(\mathcal{X}) = (\text{rank}(\mathcal{X}(1)), \ldots, \text{rank}(\mathcal{X}(N))) \), where \( \mathcal{X}(N) \) denotes the rank of the \( N \)-th frontal slice of \( \mathcal{X} \). Specially, for a 3-way tensor...
A ∈ \mathbb{C}^{n_1 \times n_2 \times n_3}, its (i, j, k)-th entry is denoted as a_{ijk} and 
\mathcal{A}(i,\cdot,\cdot), \mathcal{A}(\cdot, i, \cdot) and \mathcal{A}(\cdot, \cdot, i) represent the 
i-th horizontal, lateral and frontal slice, respectively.

Based on these common notations of tensor, one can define inner product, t-product and n-mode product.

Definition 1 (inner product [20]): For \(X, Y \in \mathbb{R}^{I_1 \times \cdots \times I_N}\), their inner product is defined as

\[
\langle X, Y \rangle = \sum_{i_1=1}^{I_1} \cdots \sum_{i_N=1}^{I_N} x_{i_1, \cdots, i_N} y_{i_1, \cdots, i_N},
\]

and \(\|X\|_F = \sqrt{\langle X, X \rangle}\) denotes the Frobenius norm of \(X\).

Definition 2 (T-product [21]): Given \(X \in \mathbb{R}^{d_1 \times d_2 \times d_3}\) and 
\(Y \in \mathbb{R}^{d_4 \times d_5 \times d_6}\), their t-product \(\mathcal{T} = X \ast Y \in \mathbb{R}^{d_3 \times d_4 \times d_5 \times d_6}\) is a tensor whose \((i,j)\)-th fiber \(\mathcal{T}(i,j,:) = \sum_{r=1}^{d_2} X(i,r,:) \odot Y(j,r,:),\) where \(\odot\) is the circular convolution.

Definition 3 (n-mode (matrix) product [20]): The n-mode product of tensor \(X \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}\) with a matrix \(U \in \mathbb{R}^{J \times I_n}\), denoted by \(X \times_n U \in \mathbb{R}^{I_1 \times \cdots \times I_{n-1} \times J \times I_{n+1} \times \cdots \times I_N}\), is defined as

\[
(X \times_n U)_{i_1 \cdots i_{n-1} j i_{n+1} \cdots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 i_2 \cdots i_N} u_{j i_n},
\]

Definition 4 (Conjugate transpose [22]): The conjugate transpose of tensor \(A \in \mathbb{C}^{n_1 \times n_2 \times n_3}\) is the tensor \(A^* \in \mathbb{C}^{n_2 \times n_1 \times n_3}\) obtained by conjugate transposing each of the frontal slices and then reversing the order of transposed frontal slices 2 through 3.

### III. RELATED WORKS ON TENSOR LOW-RANKNESS

Tensor completion is to fill in the missing values of a tensor \(Y \in \mathbb{R}^{n_1 \times n_2 \times n_3}\) under a given subset \(\Omega\) of its entries \(\{y_{ijk} | (i,j,k) \in \Omega\}\). Since tensor data of high dimensional are usually underlying low-rank [23], the formulation of tensor completion can be written as

\[
\min_{\mathcal{C}} \text{rank}_k(\mathcal{X}), \quad \text{s.t.} \quad P_{\Omega}(\mathcal{X} - \mathcal{Y}) = 0
\]

where \(\text{rank}_k(\mathcal{X})\) denotes the tubal rank of \(\mathbb{C}\) and \(P_{\Omega}\) is the linear operator that extracts entries in \(\Omega\) and fills the entries not in \(\Omega\) with zeros, i.e.,

\[
(P_{\Omega}(Y))_{i_1 \cdots i_N} = \begin{cases} y_{i_1 \cdots i_N}, & (i_1, \cdots, i_N) \in \Omega \\ 0, & \text{otherwise} \end{cases}
\]

A tensor is a high-dimensional extension of a two-dimensional matrix, and on the other hand, it can also be reordered into a two-dimensional matrix. Therefore a natural tensor filling technique is to unfold the tensor into a matrix and use matrix-based filling methods to achieve its filling and then fold it back to the original high-dimensional tensor. There are two main categories of such methods: low-rank matrix decomposition methods (LRMF) and rank minimization techniques. The principle of low-rank matrix decomposition is to decompose the target matrix into two planar matrices to achieve the inscription of low-rank prior, while rank minimization achieves this by directly imposing an additional rank constraint on the matrix to be estimated [24–27].

Although the method of reordering the tensor into a matrix is computationally efficient, this method of dimensionality reduction inevitably destroys the intrinsic structure of the tensor. For example, unfolding hyperspectral or multispectral images along the spectral dimension, and those unfolding videos along the temporal dimension, will destroy the spatial information of each frequency/time band/frame of these data [6]. Therefore, in the past ten years, a lot of work has focused on completing the task of tensor completion by directly imposing low-rank and sparse constraints on the target tensor. Motivated by the great success of matrix nuclear norms and decomposition, their promotion in the form of tensors has aroused more and more research interests, and there have been many results, e.g., tubal nuclear norm (TNN) [28] and partial sum of the tubal nuclear norm (PSTNN) [29], Tucker rank based on Tucker decomposition, CANDECOMP/PARAFAC (CP) rank based on CP decomposition, framelet based TNN (FTNN) [30]. Among existing tensor nuclear norms and tensor decomposition, Tucker decomposition, TNN and their extensions has shown superior performance in various applications, such as image/video inpainting/denoising [31–33], clustering [34] and WiFi fingerprint-based indoor localization [35].

In Tucker decomposition [10], an \(N\)-order tensor \(X \in \mathbb{R}^{I_1 \times \cdots \times I_N}\) can be written as the following form:

\[
X = S \times_1 U_1 \times_2 U_2 \times_3 \cdots \times_N U_N
\]

where \(S \in \mathbb{R}^{R_1 \times \cdots \times R_N}\) 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 bases along the \(i\)-th mode of \(X\). With this Tucker formula, high-order low-rankness can be quantified as a vector \((r_1, r_2, \cdots, r_N)\), i.e., Tucker rank. The degree of freedom of the above-mentioned Tucker decomposition is \(\prod_{i=1}^{N} r_i\), which uses the volume of the core tensor to evaluate the low-rankess of the underlying tensor. However, the core tensor obtained by the decomposition of natural data usually has a low-rank structure, which causes this degree of freedom to be further restricted to a smaller number [6, 36]. Therefore, it is difficult for Tucker rank to take reasonable measures to fully describe the inherent low-rank priors of tensors.

TNN is induced by the T-SVD [12] which attempts to decompose an third-order tensor as a tensor product of three factor tensors. Specifically, let \(A \in \mathbb{R}^{n_1 \times n_2 \times n_3}\), then, by using T-SVD, it can be factorized as

\[
A = U \ast S \ast V^* \tag{3}
\]

where \(U \in \mathbb{R}^{n_1 \times n_2 \times n_3}\), \(V \in \mathbb{R}^{n_2 \times n_2 \times n_3}\) are orthogonal, and \(S \in \mathbb{R}^{n_1 \times n_2 \times n_3}\) is a f-diagonal tensor, which is defined as a tensor whose frontal slices is a diagonal matrix. \(U \ast S \) uses a different convention than the one in [20].
and $\mathcal{V}$ are the T-product of $\mathcal{U}, \mathcal{S}$, conjugate transpose of $\mathcal{V}$, respectively, which are defined in Section II.

TNN induced method reported success on various applications in recent years [37–40]. It utilizes the low rank of the tensor spectrum, which can well capture the spatial-temporally smooth. However, by calculating the nuclear norm of the frontal slice after 1-D DFT on the mode-3 fiber, it is sensitive to the orientation of mode and cannot capture the complex intra-mode and inter-mode correlation of tensors in multiple directions.

To reduce the sensitivity to direction and simultaneously capture the complex intra- and inter-modal correlations of high-order tensors in multiple directions, improve the limited representation ability and flexibility of the tensor decomposition model in multi-origin correlation modeling, we first propose an omnidirectional tensor decomposition strategy, called the multi-modal core tensor factorization (MCTF) model, by employing multilinear techniques.

**Definition 2.1 (MCTF).** Given an $N$-way tensor $\mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, as Tucker our MCTF decomposition decomposes the input tensor as follows:

$$
\mathcal{Y} = \sum_{n=1}^{N} w_n (\mathcal{G}_n \times_N X_n),
$$

(4)

however, the main difference is that we do not require the components $X_n$ to be orthogonal, and where $X_n \in \mathbb{R}^{I_n \times r_n}$ is the $n$-th ($n = 1, 2, \cdots, N$) factor matrix which reflecting the connections (or links) between the latent components and factor matrices, $\mathcal{G}_n \in \mathbb{R}^{I_1 \times \cdots \times I_{n-1} \times r_n \times I_{n+1} \times \cdots \times I_N}$ is a tensor reflecting the joint connections between the latent components in each mode, $w_n(n = 1, 2, \cdots, N)$ are positive weights satisfying $\sum_{n=1}^{N} w_n = 1$. Tucker decomposition imposes the condition of all-orthogonality, instead of diagonality, on tensor $\mathcal{G}_n$, implies that the Tucker is always defined. As a matter of fact, $\mathcal{G}$ cannot be diagonal in general, which means that the Tucker does not necessarily reveal the rank of $\mathcal{Y}$; in the cases where $\mathcal{G}$ is diagonal, the orthogonality of the matrices of mode- $n$ singular vectors implies that $\mathcal{Y} = \sum_{i=1}^{r_n} g_{n,1} \cdots g_{n,-1} X_{i}^{(1)} \cdots X_{i}^{(N)}$ is a decomposition in a minimal number of rank-1 terms. On the other hand, the number of non-zero (significant) mode-$n$ singular values corresponds to the mode-$n$ rank (in a numerical sense) of $\mathcal{Y}$ [41].

MCTF encourages a low-rank structure, which means low-rankness in spectral domain of all orientations. It models a data tensor as simultaneously low tubal rank in all orientations (See Fig. 5 and 4). It differs from TNN which only considers low tubal rank of one spectral orientation. On the other hand, the proposed MCTF can be regarded as a generalization of matrix factorization in the form of tensor. When the factor tensor $\mathcal{G}_n$ degenerates into a matrix, MCTF degenerates into a classic matrix factorization. In addition, MCTF is also a generalized form of Tucker decomposition, if we additionally require the factor matrix $X_n$ to be orthogonal, then MCTF is equivalent to Tucker-1 decomposition.

For ease of reading, while facilitating the algorithm implementation, we further simplified the form of MCTF. Firstly,

1http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes
we defined a "tensor permutation" operation $\mathcal{R}^k$ to rearrange the dimensions of an tensor. Specifically, for a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\mathcal{R}^k$ is defined as the tensor whose $i$th mode-3 slice is the $i$-th mode-$k$ slice of $\mathcal{X}$, i.e., $\mathcal{X}(i, j, s) = \mathcal{R}^1(j, s, i) = \mathcal{R}^2(s, i, j) = \mathcal{R}^3(i, j, s)$. Then, by using this permutation, the modal-$n$ product of the tensor and the matrix can be uniformly transformed into the modal-3 product, and so the MCTF can be rewritten as follows:

$$\mathcal{Y} = \sum_{n=1}^{N} w_n (G_n^3 \times X_n) = \sum_{n=1}^{N} w_n (G_n^3 \times X_n) \label{eqn:mctf}$$

B. MCTF based Tensor Low-rankness Measure

Existing methods use either factorization or approximation schemes to recover the missing components. However, as the number of missing entries increases, factorization schemes may overfit the model because of incorrectly predefined ranks, while approximation schemes may fail to obtain easy to interpret model factors. Taking the video "Suzie" as an example, as shown in the Fig. 5, it can be seen that when the sampling rate is large, all methods can restore a clear image. But when the sampling rate is very low, that is, when there are few known entries, both the TMac and TNN failed to recover the main information of the image.

Fortunately, there are extra priors that we can utilize, i.e., the model structure is implicitly included in the low-rank factorization model, according to the factor priors, which are usually known a priori in real-world tensor objects. Fig. 3 is an example of low rank factors. From the figure, one can observe that the factor tensor obtained by the proposed MCTF decomposition is close to low rank along all the mode slices. Meanwhile, similar to the factor tensor, the factor matrix is also low-rank. This fact inspires us to further explore the structure of the factors obtained by the proposed decomposition, and then better describe the low-rank nature of the original tensor.

To this end, based on the proposed MCTF, we are motivated to we designed the multiple transform domains based tensor nuclear norm regularization for the factor tensor obtained by the decomposition, together with classic matrix nuclear norm for the factor matrix, to represent the underlying joint-manifold drawn from the model factors, and finally proposed a low-rankness measure for tensor $\mathcal{Y}$ based the proposed MCTF, i.e.,

$$S(\mathcal{Y}) = \sum_{n=1}^{N} (\tau_n \|X_n\|_s + \lambda_n \|G_n\|_{\Lambda_n, s}), \quad \label{eqn:lowrank}$$

where $X_n$ and $G_n$ are the factor matrix and tensor of $\mathcal{Y}$ with MCTF, respectively; $\alpha_n$ and $\tau_n$ is a parameter to tradeoff the two terms; $\|X\|_s$ is matrix nuclear norm, $\|X\|_{\Lambda_n, s}$ denotes the TNN of $\mathcal{X}$ based on transform domain $\Lambda_n$. Applying transform domain $\Lambda_n$ to $\mathcal{X}$ is equivalent to perform the DFT along
each mode-$n$ fiber of $\mathbf{A}'$ (as shown in the A and A-I column of Fig. 1), which can enhance the flexibility for handling different correlations along different modes and reduce the sensitivity to direction. An illustration of the proposed low-rankness measure can be found in Fig. 6. As shown in the Fig. 6, the first operation in the proposed (6) is the MCTF decomposition of the underlying tensor, which serves as a complexity measure in the original domain for all orientations. It is based on the matrix decomposition theory to extend the Tucker decomposition, and conforms to the internal mechanism of these two decompositions. As shown in the Fig. 4, the second term can be interpreted as the number of non-zero singular values of the factor matrix, and the third term models the TNN of $G_n$ in all orientations, which measures low-rankness in the Fourier domain. They tend to normalize the low-rank attributes across the subspace of each tensor mode. This comprehensive consideration in the proposed measures is conducive to explore the internal low-rank construction of the factor tensor and the low-rank nature of the tensor quantum space along each mode.

C. MCTF based Tensor Nonconvex Low-rankness Measure

Although the proposed multiple transform domains based low-rankness measure (6) can provide an efficient numerical solution and report success on low rank completion experiments (the detail performance about MCTF can be found in Section VI), we have to admit that it also has two shortcomings. Firstly, the TNN in MCTF is essentially the nuclear norm of each spectral slice in the Fourier domain, measuring the $L_{\infty}$ norm of non-zero singular values, which is not an ideal approximation of the tensor tubal rank. Secondly, both TNN and NN treat each singular value equally, so the main information may not be well preserved. Due to larger singular values usually correspond to main information, such as contours, sharp edges and smooth areas, while smaller singular values are mainly composed of noise or outliers [29, 42–44]. This means that singular values with different numerical values should be treated differently, that is, the punishment for singular values with large numerical values should be reduced, and the punishment for singular values with small numerical values should be increased.

To overcome the above two shortcomings, we designed a novel tensor log-norm and matrix log-norm to perform non-convex relaxation of TNN and NN respectively to more accurately describe the low-rank structure of factor tensor and factor matrix,

$$\min \sum_{n=1}^{3} \| G_n \times_n X_n \|^2_F + \tau_n \| X_n \|_{\log} + \lambda_n \| G_n \|_{\Lambda_n, \log},$$

where

$$\|G_n\|_{\Lambda_n, \log} = \|n^q G_n\|_{\log} = \| \text{blockdiag} (n^q G_n) \|_{\log} = \frac{1}{p} \sum_{i=1}^{p} \| n^q G_n^{(i)} \|_{\log}$$

and $\|X\|_{\log} = \sum_{i=1}^{\min\{m,n\}} (\log (|\sigma_i (X)| + \epsilon)), \text{ for } X \in \mathbb{R}^{m \times n}$, $\sigma_i (X)$ is the i-th singular value of the matrix $X$, $\epsilon > 0$ is a constant. $\alpha_n, n = 1, 2, 3$, are positive weights satisfying $\sum_{n=1}^{3} \alpha_n = 1$. For $\mathcal{L} \in \mathbb{R}^{m \times n \times p}$, $\Lambda \mathcal{L} \in \mathbb{C}^{m \times n \times p}$ denotes the result of Discrete Fourier Transformation (DFT) on $\mathcal{L} \in \mathbb{R}^{m \times n \times p}$ along the $n$-th dimension, i.e., $\Lambda \mathcal{L} = F_n \mathcal{L}$, where $F_n$ is the DFT matrix defined as $F_n = [f_1, \cdots, f_i, \cdots, f_{n-1}] \in \mathbb{R}^{n \times n}$. $\Lambda \mathcal{L}$ denotes the block-diagonal matrix of the tensor $\mathcal{L}$ in the Fourier domain, i.e. $\Lambda \mathcal{L} = \text{blockdiag} (\hat{\mathcal{L}}) = \text{diag}(\hat{\mathcal{L}}^{(1)}, \hat{\mathcal{L}}^{(2)}, \cdots, \hat{\mathcal{L}}^{(p)}) \in \mathbb{C}^{mp \times np}$, where $\hat{\mathcal{L}}^{(i)}$ denotes the i-th frontal slices of $\hat{\mathcal{L}}, i = 1, 2, \cdots, p$.

V. MCTF LOW-RANKNESS MEASURE-BASED MODEL & ITS SOLVING SCHEME

Here we introduce the optimization of the proposed two models and analyze their convergence.

A. The MCTF and NC-MCTF minimization models

Before giving the optimization of MCTF and NC-MCTF, we first introduce two lemmas as follows:

**Lemma 1 (singular value shrinkage operator, SVT):** For $M \in \mathbb{R}^{m \times n \times n_2}$, $\mathbf{P} \in \mathbb{C}^{m \times n_1}$ denote the singular value decomposition (SVD) of matrix $\mathbf{M}$ with rank $r$, where $\mathbf{E}_r = \text{diag}(\{\sigma_i\}_{1 \leq i \leq r})$, $\sigma_i$ is the i-th largest singular value of $\mathbf{M}$. Then, the following properties hold,

$$D_\delta (\mathbf{M}) = \arg \min_{\text{rank}(X) \leq r} \delta \| X \|_* + \frac{1}{2} \| X - W \|_F^2,$$

where $D_\delta (\mathbf{W}) = \mathbf{P} \text{ diag}(\{\sigma_i - \delta, 0\}) \mathbf{Q}^\dagger$, and $\| . \|_*$ is the matrix nuclear norm.

**Lemma 2 (weighted nuclear norm minimization, WNNM) [42]:** For any $\gamma > 0$, $Y \in \mathbb{R}^{m \times n}$ and $0 \leq d_1 \leq d_2 \leq \cdots \leq d_r (r = \min\{m,n\})$, a global optimal solution to the following problem

$$\min_X \sum_{j=1}^{r} \gamma d_j \sigma_j (X) + \frac{1}{2} \| Y - X \|_F^2$$

is given by the following singular value thresholding

$$X^* = W_{\gamma,d}(Y) = US_{\gamma,d}(\Sigma) V^T$$

where $\Sigma = U \Sigma V^T$ is the SVD of $Y$, $\sigma_j (X)$ denotes the j-th singular value of $X$ and $S_{\gamma,d}(\Sigma)_{ij} = \max (\Sigma_{ij} - \gamma d, 0)$. 

We then analyzing the optimization of the proposed models, the objective function of the proposed MCTF and NC-MCTF are listed as follows:

\[
f(X, G, Y) = \sum_{n=1}^{N} \frac{\alpha_n}{2} \| Y - G_n \times_n X_n \|_F^2 + \tau_n \| X_n \|_*\text{,} \quad (11)
\]

The minimization of the proposed models are two complicated optimization problems, which are difficult to solve directly. Here, we adopt the block successive upper-bound minimization (BSUM) [45] to solve them.

According to the proximal operator [46], for the \(k\)-th iteration, the update can be written as follows:

\[
\text{Prox}_f(S, S^k) = \arg \min_{S} f(S) + \frac{\rho}{2} \| S - S^k \|_F^2, \quad (13)
\]

where \(\rho > 0\) is the proximal parameter, \(S = (X, G, Y)\) and \(S^k = (X^k, G^k, Y^k)\).

Let \(S^0 = (X^0, G^0, Y^0)\), \(S^k = (X^{k+1}, G^k, Y^k)\), \(S_{S^k} = (X^{k+1}, G^k, Y^k)\). By BSUM, (13) can be rewritten as follows:

\[
X^{k+1} = \text{Prox}_f(G, S_1^k) = \arg \min_{X} f(X, G, J^k) \quad + \frac{\rho}{2} \| X - X^k \|_F^2, \quad (14)
\]

\[
G^{k+1} = \text{Prox}_f (G, S_2^k) = \arg \min_{G} f(X^{k+1}, G, Y^k) \quad + \frac{\rho}{2} \| G - G^k \|_F^2, \quad (15)
\]

\[
Y^{k+1} = \text{Prox}_f (Y, S_3^k) = \arg \min_{Y} f(X^{k+1}, G^{k+1}, Y) \quad + \frac{\rho}{2} \| Y - Y^k \|_F^2. \quad (16)
\]

1) **Update \(X_n\) with fixing others:** By introducing one auxiliary variable \(Z_n\), the \(X_n\)-subproblem in (14) can be rewritten as

\[
\begin{align*}
\text{arg min}_{X_n, Z_n} & \sum_{n=1}^{N} \frac{\alpha_n}{2} \| Y - G_n \times_n X_n \|_F^2 + \tau_n \| Z_n \|_* \text{,} \\
& + \frac{\rho_n}{2} \| X_n - X_n \|_F^2, \quad \text{s.t., } X_n = Z_n. 
\end{align*} \quad (17)
\]

Based on the augmented Lagrange multiplier (ALM) method, the above minimization problem (17) can be transformed into non-contrived problem, and be solved by SVT (V-A) and WNNM operator (10):

\[
Z_n^{k+1} = D_{\frac{\alpha_n}{2}} \left( X_n^k + \frac{G_n}{\rho_n} \right), \quad n = 1, 2, \ldots, N; \quad (16)
\]

\[
Z_n^{k+1} = W_{\frac{\alpha_n}{2}} \left( X_n^k + \frac{G_n}{\rho_n} \right), \quad n = 1, 2, \ldots, N; \quad (17)
\]

\[
X_n^{k+1} = (\alpha_n G_n^T G_n + 2 \rho_n I_n)^{-1} \left[ \alpha_n G_n^T Y^{(n)} + Z_n^{k+1} - \frac{\gamma_i}{\rho_n} + \frac{X_n^k}{\rho_n} \right], \quad (18)
\]

where (16) for MCTF, (17) for NC-MCTF. Based on the ALM method, the multipliers are updated by the following equations:

\[
\Gamma_n = \Gamma_n^k + X_n - Z_n. \quad (19)
\]

2) **Update \(G_n\) with fixing others:** By introducing an auxiliary variable, the \(G_n\)-subproblem can be rewritten as

\[
\begin{align*}
\text{arg min}_{G_n} & \sum_{n=1}^{N} \frac{\alpha_n}{2} \| Y - G_n \times_n X_n \|_F^2 + \tau_n \| G_n \|_* \text{,} \\
& + \frac{\rho_n}{2} \| G_n - G_n \|_F^2, \quad \text{s.t., } G_n = G_n. 
\end{align*} \quad (20)
\]

By using the ALM and SVT operator (V-A), one can also obtain the solutions:

\[
n_{j+1} = D_{\frac{\alpha_n}{2}} \left( \frac{\mu_n}{\tau_n} \right), \quad J^k = 1, 2, \ldots, p. \quad (21)
\]

Then, the \((k+1)\)-th updating of \(J^{k+1}\) can be obtained via inverse Fourier transform

\[
J^{k+1} = \text{iift} \left( \frac{\mu_n}{\tau_n} \right), \quad (22)
\]

Similarly, the \(J_n\) related subproblem can be solved by the WNNM operator (10), i.e.,

\[
n_{j+1} = W_{\frac{\alpha_n}{2}} \left( \frac{\mu_n}{\tau_n} \right), \quad J^k = 1, 2, \ldots, p. \quad (23)
\]

Then, the \((k+1)\)-th updating of \(J^{k+1}\) can be obtained via inverse Fourier transform

\[
J^{k+1} = \text{iift} \left( \frac{\mu_n}{\tau_n} \right), \quad (24)
\]

With other variables fixed, the minimization subproblem for \(G_n\) is also convex and has the following closed-form solution

\[
G_n^{k+1} = \text{fold} \left( \begin{align*}
& \left( Y_n^{(n)} \right) \left( X_n^{k+1} \right)^T + 2 \rho_n \left( \frac{\gamma_n^{k+1}-\gamma_n^{k+n}}{2} + \frac{G_n^k}{2} \right), \\
& \left( X_n^{k+1}\right)^T + 2 \rho_n I_n \right)^{-1} \
\end{align*} \right), \quad n = 1, 2, \ldots, N. \quad (25)
\]

Finally, the Lagrangian multiplier can be updated by the following equations

\[
\Gamma_n^{k+1} = \Gamma_n^k + G_n - J_n. \quad (26)
\]

3) **Update \(Y\) with fixing others:** The update of \(Y^{k+1}\) can be written explicitly as

\[
Y^{k+1} = P_{\Omega} \left[ \sum_{n=1}^{N} \alpha_n \text{fold} \left( \frac{G_n^{k+1} X_n^{k+1} + \rho_n Y_n^{(n)}}{1 + \rho_n} \right) \right] + F. \quad (27)
\]

where \(F\) is the observed data; \(P_{\Omega}\) is an operator defined in subsection II.
Algorithm 1: Algorithm for the proposed MCTF and NC-MCTF based tensor low-rankness measure.

1: **Input:** The observed tensor $F$; The set of index of observed entries $\Omega$; The given $n$-rank, $r = (r_1, r_2, r_3)$; stopping criterion $\varepsilon$.

2: **Output:** the completed tensor.

3: Initialize: $X_0 = Z_0 = 0, \varphi^0_n = J_n^0 = 0, \Gamma_n^0 = 0, n = 1, 2, \ldots, N$; $\rho_{\text{max}} = 10^6$, $\rho = 1.5$, $Y = \mathcal{P}_\Omega(F)$, and $k = 0$.

4: Repeat until convergence:

5: Update $X$, $Z$, $G$, $J$, $\gamma$, $\Gamma^X$, $\Gamma^G$ via

   1st step: Update $Z_n$ of MCTF via (16) or $Z_n$ of NC-MCTF via (17)

   2nd step: Update $X_n$ via (18)

   3rd step: Update $G_n$ via (25)

   4th step: Update $J_n$ of MCTF via (22) or $J_n$ of NC-MCTF via (24)

   5th step: Update $\gamma$ via (27)

   6th step: Update the parameter via (19), (26)

6: Check the convergence condition: $\frac{\|x^{k+1} - x^k\|_F}{\|x^k\|_F} < \varepsilon$.

B. Complexity Analysis

In this subsection, the proposed algorithm for the proposed MCTF and NC-MCTF are summarized as Algorithm 1. Further, we discuss the complexity of the proposed algorithms. Complexity Analysis: The cost of computing $X_n$ is $O(I_n r_n^2 + I_n r_n s_n + r_n^2 s_n)$; calculating $Z_n$ has a complexity of $O(\prod_{x} I_x \times r_n^2)$; the complexity of updating $J_n$ is $O(I_n r_n^2)$; calculating $G_n$, $n = 1, 2, 3$, in both MCTF-based solver and NC-MCTF-based solver, have a complexity of $O(I_1 I_2 I_3 (\log(I_1 I_2 I_3) + \sum_{n=1}^{3} \min(I_n, I_n + 1)))$, where we define $I_1 = I_1$; calculating $\gamma$ has a complexity of $O(\sum_{n=1}^{N} r_n s_n)$. Then, the total complexity of the proposed algorithms can be obtained by counting the complexity of the above variables. For easily viewing, we list the total complexity of the proposed models as follows:

$$O(3I_n r_n^2 + 3I_n r_n s_n + 3r_n^2 s_n + I_1 I_2 I_3 (\log(I_1 I_2 I_3)$$

$$+ \sum_{n=1}^{3} \min(I_n, I_n + 1))) \quad (28)$$

VI. Numerical Experiments

Three types of public tensor data-sets, i.e., video, MRI and hyperspectral image, are selected for verification experiments to evaluate the performance of the proposed model. Five state-of-the-art techniques proposed between 2013-2020, i.e., five tensor completion models related to the proposed models: TMac(2013) [18], MF-TV method(2016) [46], TNN(2016) [28], PSTNN(2020) [29] and FTNN(2020) [30], were chosen for comparison.

Two types of standards for evaluation: qualitative visual evaluation of the restored data, five widely used quantitative picture quality indices (PQIs): PSNR [47], SSIM [48], FSIM [49], ERGAS [50] and SAM, were utilized to assess the quality of restored tensor. All experiments were performed on MATLAB 2018b, Intel core i7@2.2GHz and 64.0 GB RAM. For a tensor $Y \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ the sampling ratio (SR) can be defined as: $SR = \frac{\text{Number of observed entries}}{\text{Number of total entries}}$, where $\text{Number of observed entries}$ denotes the number of sampled entries, $\Omega$ represents index set. The sampled entries are chosen randomly from a tensor $Y$ by a uniform distribution.

A. MRI

In this subsection, to further verify the versatility of the proposed models for different datasets, we conduct experiments on cubical MRI data with size 150 \times 150 \times 181. SRs are set as follows: 5%, 10%, 20% and 30%. Here, we set the rank to $(T_1, T_2, T_3)$, where $T_1$, $T_2$, $T_3$ denote the number of the largest 0.5% singular values of model-1, model-2 and model-3, respectively.

For quantitative evaluation, the table II lists the PQI of all recovery results in detail, the best results are marked in bold. It can be seen from the table that the proposed NC-MCTF obtains the best PQI, and the second best is the proposed MCTF, both of which are superior to competing methods of the same type. Fig. 13 shows the detailed PSNR, SSIM and FSIM of all slices of the restored data. The same advantages of our model can also be seen from it.

Furthermore, the proposed model is evaluated in terms of visual evaluation. We choose the restoration result of 0.1 sampling rate as an example. Fig. 7 shows the original MRI data, sampled data and recovery results of different methods. Compared with competing methods, the image recovered by MCTF and NC-MCTF show richer details and clearer structures. In addition, one can also see from the figure that the non-convex metric, as shown in NC-MCTF, produces impressive improvement over MCTF.
to impose additional prior constraints on the solution space of relatively accurately, as at high sampling rates, it is necessary that the image is scarce. If it is desired to restore the data image information contained in the input observation, then it is seen from the figure that the proposed model has achieved a significant advantage in restored images, especially at low sampling rates. Because when the sampling rate is low, the restored result is to the original reference image, the closer the restored data under different sampling conditions. The quantitative comparison: Table III and table IV give the detailed PQI of all recovered data under three sampling rates. The best value is highlighted in bolded fonts.

**TABLE III**

| SAMPLING RATES | method | noisy MF-TV | TMac | FTNN | PSTNN | TNN | MCTF | NC-MCTF |
|----------------|--------|-------------|------|------|-------|-----|------|---------|
| PSNR           |        | 7.259       | 13.810 | 23.385 | 27.294 | 17.447 | 22.005 | 27.130  |
| SSIM           | 0.009  | 0.094       | 0.622 | 0.465 | 0.912 | 0.563 |       | 0.948   |
| FSIM           | 0.454  | 0.42        | 0.792 | 0.555 | 0.59  | 0.776 |       | 0.842   |
| ERGAS          | 1057.282 | 501.117    | 167.927 | 129.27 | 327.678 | 194.844 | 292.455 | 64.698   |

Further, the proposed model was comprehensively evaluated from both quantitative and qualitative perspectives. Here, we set the rank to \((T_1, T_2, T_3)\), where \(T_1, T_2, T_3\) denote the number of the largest 0.5\% singular values of model-1, model-2 and model-3, respectively.

Quantitative comparison: Table III and table IV give the detailed PQI of all recovered data under three sampling rates. The boldface indicates the best PQI for each sampling rate. It can be clearly seen from the table III and table IV that among all the test methods, the proposed NC-MCTF obtains the best results, and the evaluation index obtained by the proposed MCTF is also superior to competitive methods.

Based on above quantitative comparison, we further conduct a quantitative evaluation of the proposed model in terms of vision. Fig. 9 and Fig. 11 show the partial slice images of restored data under different sampling conditions. The closer the restored result is to the original reference image, the better the performance of the corresponding model. It can be seen from the figure that the proposed model has achieved a significant advantage in restored images, especially at low sampling rates. Because when the sampling rate is low, the original image information contained in the input observation image is scarce. If it is desired to restore the data image relatively accurately, as at high sampling rates, it is necessary to impose additional prior constraints on the solution space of the optimization model to increase the accuracy of the obtained solution, as the proposed model does.

**C. Hyperspectral Image**

In this subsection, we choose two HSI data to apply simulation experiments. The first dataset is five sequential images\(^3\) were acquired by the Sentinel-2 MSI on 05/09, 15/09, 20/09, 5/10 and 15/10, 2018, in Belgium, with 20 m spatial-resolution, 10 \times 10 km. For this dataset, the HSIs are corrupted by various types of missing areas with cloud shape (see the second row of Fig. 14).

\(^3\)http://trace.eas.asu.edu/yuv/

\(^4\)https://drive.google.com/file/d/1LlvUKtUWAkoF6R6ghREwwP2Wfja0U/view
Fig. 9. One slice of the recovered video for “suzie” by MF-TV, TMac, FTNN, PSTNN, TNN, our MCTF and NC-MCTF. The sampling rate is 5%.

Fig. 10. One slice of the recovered video for “suzie” by MF-TV, TMac, FTNN, PSTNN, TNN, our MCTF and NC-MCTF. The sampling rates are 10%, 20%, respectively.

Fig. 11. One slice of the recovered video for “hall” by MF-TV, TMac, FTNN, PSTNN, TNN, our MCTF and NC-MCTF. The sampling rates are 5% and 10%, respectively.

TABLE IV
THE AVERAGED PSNR, SSIM, FSIM, ERGA AND SAM OF THE RECOVERED RESULTS ON VIDEO “HALL” BY MF-TV, TMAC, PSTNN, TNN, FTNN, OUR MCTF AND NC-MCTF WITH DIFFERENT SAMPLING RATES. THE BEST VALUE IS HIGHLIGHTED IN BOLDER FONTS.

| Method | Noisy | MF-TV | TMac | FTNN | PSTNN | TNN | MCTF | NC-MCTF |
|--------|-------|-------|------|------|-------|-----|------|--------|
| PSNR   |       |       |      |      |       |     |      |        |
| SR = 0.05 |      |       |      |      |       |     |      |        |
| Noisy  | 4.82  | 13.53 | 22.10| 30.02| 16.075| 20.78| 26.215| 27.415 |
| SSIM   | 0.007 | 0.412 | 0.675| 0.792| 0.36   | 0.636| 0.856 | 0.982  |
| FSIM   | 0.307 | 0.612 | 0.708| 0.835| 0.672  | 0.792| 0.840 | 0.906  |
| ERGAS  | 1225.779 | 452.351| 168.866| 98.14 | 195.315| 105.199| 91.728 |
| SR = 0.1 |      |       |      |      |       |     |      |        |
| Noisy  | 5.055 | 24.855| 26.936| 32.790| 29.014| 24.33 | 28.273 | 31.481 |
| SSIM   | 0.013 | 0.829 | 0.854| 0.892| 0.905  | 0.933| 0.942 | 0.952  |
| FSIM   | 0.393 | 0.873 | 0.888| 0.934| 0.936  | 0.935| 0.936 | 0.935  |
| ERGAS  | 1193.075| 131.422| 97.185| 59.375| 50.735 | 42.259| 39.723| 37.080 |
| SR = 0.2 |      |       |      |      |       |     |      |        |
| Noisy  | 5.567 | 33.006| 27.646| 35.755| 33.626| 33.961| 33.052| 34.097 |
| SSIM   | 0.025 | 0.94  | 0.869| 0.902| 0.961  | 0.982| 0.956 | 0.962  |
| FSIM   | 0.403 | 0.954 | 0.987| 0.972| 0.974  | 0.985| 0.970 | 0.970  |
| ERGAS  | 1124.737| 50.971| 80.271| 44.144| 45.725 | 45.851| 48.414| 43.191 |

The second data-set is the airborne visible/infrared imaging spectrometer (AVIRIS) copper salt data with size 150 × 150 × 210. SRs are set as follows: 0.025, 0.05 and 0.1. Here, the missing values are also random sampled, and we set the

5http://aviris.jpl.nasa.gov/html/aviris.freedata.html
TABLE V
THE AVERAGED PSNR, SSIM, FSIM, ERGAS AND SAM OF THE RECOVERED RESULTS ON HYPERSPECTRAL IMAGE "CUPRITE" BY MF-TV, TMAC, FTNN, PSTNN, TNN, OUR MCTF AND NC-MCTF WITH DIFFERENT SAMPLING RATES. THE BEST VALUE IS HIGHLIGHTED IN BOLDER FONTS.

| SR       | Method | Noisy | MF-TV | TMac | PSTNN | TNN | MCTF | NC-MCTF |
|----------|--------|-------|-------|------|-------|-----|------|---------|
| 0.05     | PSNR   | 7.779 | 34.684| 28.945| 20.621| 26.579| 34.739|         |
|          | SSIM   | 0.01  | 0.915 | 0.805 | 0.907 | 0.912 |       |         |
|          | FSIM   | 0.471 | 0.920 | 0.978 | 0.951 | 0.943 |       |         |
|          | ERGAS  | 1043.633| 40.888| 35.627| 35.510| 35.015| 37.449|         |
| 0.1      | PSNR   | 10.047| 34.845| 29.015| 21.621| 27.579| 35.739|         |
|          | SSIM   | 0.013 | 0.920 | 0.805 | 0.907 | 0.912 |       |         |
|          | FSIM   | 0.471 | 0.920 | 0.978 | 0.951 | 0.943 |       |         |
|          | ERGAS  | 1030.139| 40.888| 35.627| 35.510| 35.015| 37.449|         |
| 0.2      | PSNR   | 12.047| 34.845| 29.015| 21.621| 27.579| 35.739|         |
|          | SSIM   | 0.013 | 0.920 | 0.805 | 0.907 | 0.912 |       |         |
|          | FSIM   | 0.471 | 0.920 | 0.978 | 0.951 | 0.943 |       |         |
|          | ERGAS  | 1020.75| 40.888| 35.627| 35.510| 35.015| 37.449|         |

VII. CONCLUSIONS
In this paper, based on basic tensor decomposition theory, we first defined a novel tensor decomposition to explore the low-rank structure of underlying tensors. Then, we studied the structural characteristics of the factors obtained by the decomposition, and proposed a novel tensor low-rankness measure. Furthermore, we performed nonconvex relaxation on the submeasure, and obtained a better low-rankness measure. Thanks to the “good” properties of TNN and Tucker decomposition, our model and theoretical guarantee are both natural extensions of TNN and Tucker. We have also developed an effective method to calculate the optimization problem corresponding to the proposed model. Numerical experiments verify our theory, and the results of hyperspectral image, MRI and video prove that our algorithms can recover a variety of low-rank tensors with significantly fewer samples than the compared methods.

APPENDIX A
DETAILS OF THE OPTIMIZATION
A. Update $X_n$ with fixing others
The $X_n$-subproblem in (14) can be written as follows:

$$X_n^{k+1} = \arg\min_{X_n} \sum_{n=1}^{3} \frac{\alpha_n}{2} \|Y - G_n \times_n X_n\|_F^2 + \tau_n \|X_n\|_{* \ or \ log} + \frac{\rho_n}{2} \|X_n - X_n^k\|_F^2.$$  
(29)

To efficiently solve it, we first introduce one auxiliary variable $Z_n$, then (29) can be rewritten as

$$\arg\min_{X_n, Z_n} \sum_{n=1}^{3} \frac{\alpha_n}{2} \|Y - G_n \times_n X_n\|_F^2 + \tau_n \|Z_n\|_{* \ or \ log} + \frac{\rho_n}{2} \|X_n - X_n^k\|_F^2, s.t., X_n = Z_n.$$  
(30)

Based on the augmented Lagrange multiplier (ALM) method, the above minimization problem (30) can be transformed into

$$\arg\min_{X_n, Z_n} \sum_{n=1}^{3} \frac{\alpha_n}{2} \|Y - G_n \times_n X_n\|_F^2 + \tau_n \|Z_n\|_{* \ or \ log} + \frac{\rho_n}{2} \|X_n - X_n^k\|_F^2 + \tau_n \|X_n - Z_n\|_{L_1}^2 + \frac{\rho_n}{2} \|Z_n - Z_n^k\|_{L_1}^2,$$  
(31)

where $\Gamma_n$ is a Lagrange multiplier.

With other variables fixed, the minimization subproblem for $Z_n$ can be deduced from (31) as follows:

$$Z_n^{k+1} = \arg\min_{Z_n} \tau_n \|Z_n\|_{* \ or \ log} + \frac{\rho_n}{2} \|X_n - Z_n + \Gamma_n/\rho_n\|_F^2.$$  
(32)
Fig. 14. Sentinel-2 MSI multi-temporal data sequence cloud removal experiments (20 m spatial-resolution; 10 × 10 km; five temporal images) over Mechelen in Belgium.

Fig. 15. One slice of the recovered HSI "Cuprite" by MF-TV, TMac, PSTNN, TNN, our MCTF and NC-MCTF. The sampling rate is 5%.

Fig. 16. Performance variation of the proposed method in terms of the NC-MCTF on different C and different sampling rate.

For MCTF based on ∥.∥∗, by using the SVT operator (V-A), it is easy to get

\[ Z^{k+1}_n = D_{\rho_n} (X^k_n + \Gamma^{k-1}_n / \rho_n), \ n = 1, 2, \cdots, N; \quad (33) \]

while for the \( Z_n \) related subproblem of nonconvex low-rankness measure (NC-MCTF):

\[ Z^{k+1}_n = \arg \min_{Z_n} \tau_n \|Z_n\|_{\log} + \frac{\rho_n}{2} \|X^k_n - Z_n + \Gamma_n / \rho_n\|_F^2, \]

it can be solved by WNMM operator, i.e.,

\[ Z^{k+1}_n = W_{\tau_n, \epsilon} (X^k_n + \Gamma^k_n / \rho_n), n = 1, 2, \cdots, N. \quad (35) \]

With other variables fixed, the minimization subproblem for \( X_n (n \neq 3) \) can be deduced from (31) as follows:

\[ X^{k+1}_n = \arg \min_{X_n} \alpha_n \|Y - G^k_n \times X_n\|_F^2 \]

\[ + \frac{\rho_n}{2} \left\| X_n - \frac{Z^{k+1}_n - \Gamma_n / \mu_n + X^k_n}{2} \right\|_F^2. \quad (36) \]

They are convex and have the following closed-form solutions

\[ X^{k+1}_n = (\alpha_n G^T_n G_n + 2 \rho_n I_n)^{-1} [\alpha_n G^T_n Y_n] \]

\[ + \mu_n (\frac{Z^{k+1}_n - \Gamma_n / \mu_n + X^k_n}{2}). \quad (37) \]
Based on the ALM method, the multipliers are updated by the following equations:

$$\Gamma_n^X = \Gamma_n^X + X_n - Z_n.$$  \hfill (38)

**B. Update \(G_n\) with fixing others**

The \(G_n\)-subproblem in (14) can be written as follows:

$$G^{k+1} = \arg\min_G \frac{\alpha_G}{2} \left\| Y - G_n \times_n X_n \right\|_2^2 + \lambda_n \left\| G_n \right\|_{\text{A}}^*, \quad \text{s.t.}, \ G_n = J_n.$$ \hfill (39)

By introducing an auxiliary variable, (39) can be rewritten as

$$\arg\min_{\hat{G}_n} \frac{\alpha_n}{2} \left\| Y - G_n \times_n X_n \right\|_2^2 + \lambda_n \left\| J_n \right\|_{\text{A}}^*, \quad \text{s.t.}, \ \hat{G}_n = G_n.$$ \hfill (40)

Then, (20) can also be reformulated as

$$\arg\min_{A_n, J_n} \frac{\alpha_n}{2} \left\| Y - G_n \times_n X_n \right\|_2^2 + \lambda_n \left\| J_n \right\|_{\text{A}}^* + \rho_n \left\| G_n - \hat{G}_n \right\|_F^2 + \Gamma_n^G/\rho_n.$$ \hfill (41)

Firstly, with other variables fixed, the minimization subproblem for \(J_n\) can be deduced from (41) as follows:

$$J_n^{k+1} = \arg\min_{J_n} \lambda_n \left\| J_n \right\|_{\text{A}}^* + \frac{\rho_n}{2} \left\| G_n^k - J_n + \Gamma_n^G/\rho_n \right\|_F^2.$$ \hfill (42)

The updating of \(J_n\) in (42) has a closed-form solution. For the sake of simplicity, we denote the iteration of \(J_n\) as

$$J_{n+1} = \arg\min_{J_n} \left\| J_n \right\|_{\text{A}}^* + \frac{\rho_n}{2} \left\| J_n - U_n \right\|_F^2,$$ \hfill (43)

where \(U_n = \hat{G}_n^k + \Gamma_n^G/\rho_n\).

Solving the optimization subproblem (43) is equivalent to solving the following tensor recovery problem in the frequency domain:

$$n \hat{J}_n^{k+1} = \arg\min_{n \hat{J}_n^{q}} \frac{1}{p} \sum_{q=1}^{p} \left\| n \hat{J}_n^{q} \right\|_2 + \frac{\rho_n}{2} \left\| n \hat{J}_n^{q} - U_n \right\|_F^2.$$ \hfill (44)

Finally, the Lagrangian multiplier can be updated by the following equations

$$\Gamma_n^G = \Gamma_n^G + G_n - J_n.$$ \hfill (53)

**C. Update \(Y\) with fixing others**

With other variables fixed, the minimization subproblem for \(Y_n\) in (14) can be written as

$$Y^{k+1} = \arg\min_{Y_n} \frac{\alpha_n}{2} \left\| Y - G_n \times_n X_n \right\|_2^2 + \frac{\rho_n}{2} \left\| Y - Y^k \right\|_F^2,$$ \hfill (54)

Then, the \((k+1)\)-th updating of \(J_n^{k+1}\) can be obtained via inverse Fourier transform

$$J_n^{k+1} = \text{ift} \left( n \hat{J}_n^{k+1}, [\cdot], n \right).$$ \hfill (57)

Similarly, the \(J_n\) related subproblem of nonconvex low-rankness measure

$$n \hat{J}_n^{k+1}(q) = \arg\min_{n \hat{J}_n^{q}} \left\| n \hat{J}_n^{q} \right\|_F + \frac{\rho_n}{2} \left\| n \hat{J}_n^{q} - n \hat{F}_n^{q} \right\|_F^2,$$ \hfill (48)

can be solved by the WNNM operator (10), i.e.,

$$n \hat{J}_n^{k+1}(q) = W_{\frac{1}{\rho_n}, \epsilon} \left( n \hat{U}_n^{q} \right), \quad q = 1, 2, \cdots, p.$$ \hfill (49)

Then, the \((k+1)\)-th updating of \(J_n^{k+1}\) can be obtained via inverse Fourier transform

$$J_n^{k+1} = \text{ift} \left( n \hat{J}_n^{k+1}, [\cdot], n \right).$$ \hfill (50)

Secondly, with other variables fixed, the minimization subproblem for \(G_n\) can be deduced from (41) as follows:

$$G_n^{k+1} = \arg\min_{G_n} \frac{\alpha_n}{2} \left\| Y - G_n \times_n X_n \right\|_2^2 + \frac{\rho_n}{2} \left\| G_n - \hat{G}_n \right\|_F^2.$$ \hfill (51)

It is also convex and has the following closed-form solution

$$G_n^{k+1} = \text{fold} \left( \left( Y_n^k (X_n^{k+1})^T + 2\rho_n (J_n^{k+1} - \hat{G}_n^k/\rho_n + G_n^{k}) \right) \left( X_n^{k+1} (X_n^{k+1})^T + 2\rho_n I_n \right)^{-1} \right), \quad n = 1, 2, \cdots, N.$$ \hfill (52)

By using the SVT operator (V-A), it is easy to obtain the solutions of the minimization subproblems (45)

$$n \hat{J}_n^{k+1}(q) = D_{\frac{1}{\rho_n}} \left( n \hat{U}_n^{q} \right), \quad q = 1, 2, \cdots, p.$$ \hfill (46)
REFERENCES

[1] A. Shashua, “On photometric issues in 3d visual recognition from a single 2d image,” International Journal of Computer Vision, vol. 21, no. 1-2, pp. 99–122, 1997.

[2] S. Raychaudhuri, J. M. Stuart, and R. B. Altman, “Principal components analysis to summarize microarray experiments: application to sporulation time series,” in Biocomputing 2000. World Scientific, 1999, pp. 455–466.

[3] F. Heide, M. Steinberger, Y.-T. Tsai, M. Rouf, D. Pajak, D. Reddy, O. Gallo, J. Liu, W. Heidrich, K. Egiazarian et al., “Flexisp: A flexible camera image processing framework,” ACM Transactions on Graphics (TOG), vol. 33, no. 6, pp. 1–13, 2014.

[4] A. M. Teodoro, J. M. Bioucas-Dias, and M. A. Figueiredo, “Image restoration and reconstruction using variable splitting and class-adapted image priors,” in 2016 IEEE International Conference on Image Processing (ICIP). IEEE, 2016, pp. 3518–3522.

[5] 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 Transactions on Knowledge and Data Engineering, vol. 19, no. 3, pp. 355–369, 2007.

[6] Q. Xie, Q. Zhao, D. Meng, and Z. Xu, “Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 8, pp. 1888–1902, 2017.

[7] H.-J. Zeng, X.-Z. Xie, K. Wen-Feng, S. Cui, and J.-F. Ning, “Hyperspectral image denoising via combined non-local self-similarity and local low-rank regularization,” IEEE Access, vol. 8, pp. 50 190–50 208, 2020.

[8] Y. Wu, H. Tan, Y. Li, F. Li, and H. He, “Robust tensor decomposition based on cauchy distribution and its applications,” Neurocomputing, vol. 223, pp. 107–117, 2017.

[9] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. A. Phan, “Tensor decompositions for signal processing applications: From two-way to multiway component analysis,” IEEE signal processing magazine, vol. 32, no. 2, pp. 145–163, 2015.

[10] L. R. Tucker, “Some mathematical notes on three-mode factor analysis,” Psychometrika, vol. 31, no. 3, pp. 279–311, 1966.

[11] R. A. Harshman and M. E. Lundy, “Parafac: Parallel factor analysis,” Computational Statistics & Data Analysis, vol. 18, no. 1, pp. 39–72, 1994.

[12] M. E. Kilmer, K. Braman, N. Hao, and R. C. Hoover, “Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging,” SIAM Journal on Matrix Analysis and Applications, vol. 34, no. 1, pp. 148–172, 2013.

[13] J. Xue, Y. Zhao, W. Liao, J. C.-W. Chan, and S. G. Kong, “Enhanced sparsity prior model for low-rank tensor completion,” IEEE Transactions on Neural Networks and Learning Systems, vol. 31, no. 11, pp. 4567–4581, 2020.

[14] J. Håstad, “Tensor rank is np-complete,” Journal of algorithms (Print), vol. 11, no. 4, pp. 644–654, 1990.

[15] 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.

[16] J. Liu, P. Musialski, P. Wonka, and J. Ye, “Tensor completion for estimating missing values in visual data,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 1, pp. 208–220, 2012.

[17] B. Romera-Paredes and M. Pontil, “A new convex relaxation for tensor completion,” in Advances in Neural Information Processing Systems, 2013, pp. 2967–2975.

[18] Y. Xu, R. Hao, W. Yin, and Z. Su, “Parallel matrix factorization for low-rank tensor completion,” arXiv preprint arXiv:1312.1254, 2013.

[19] H. Zeng, X. Xie, and J. Ning, “Tensor completion using enhanced multiple modes low-rank prior and total variation,” arXiv preprint arXiv:2004.08747, 2020.

[20] T. G. Kolda and B. W. Bader, “Tensor decompositions and applications,” SIAM review, vol. 51, no. 3, pp. 455–500, 2009.

[21] M. E. Kilmer, K. Braman, N. Hao, and R. C. Hoover, “Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging,” SIAM Journal on Matrix Analysis and Applications, vol. 34, no. 1, pp. 148–172, 2013.

[22] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, “Tensor robust principal component analysis with a new tensor nuclear norm,” IEEE transactions on pattern analysis and machine intelligence, vol. 42, no. 4, pp. 925–938, 2019.

[23] P. Zhou, C. Lu, Z. Lin, and C. Zhang, “Tensor factorization for low-rank tensor completion,” IEEE Transactions on Image Processing, vol. 27, no. 3, pp. 1152–1163, 2017.

[24] A. M. Buchanan and A. W. Fitzgibbon, “Damped newton algorithms for matrix factorization with missing data,” in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), vol. 2. IEEE, 2005, pp. 316–322.

[25] E. J. Candès, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis?” Journal of the ACM (JACM), vol. 58, no. 3, pp. 1–37, 2011.

[26] 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 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2010, pp. 771–778.

[27] Q. Ke and T. Kanade, “Robust l/sub 1/norm factorization in the presence of outliers and missing data by alternative convex programming,” in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), vol. 1. IEEE, 2005, pp. 739–746.

[28] Z. Zhang and S. Aeron, “Exact tensor completion using t-svd,” IEEE Transactions on Signal Processing, vol. 65, no. 6, pp. 1511–1526, 2016.

[29] T.-X. Jiang, T.-Z. Huang, X.-L. Zhao, and L.-J. Deng, biocomputing 2000. World Scientific, 1999, pp. 455–466.
“Multi-dimensional imaging data recovery via minimizing the partial sum of tubal nuclear norm,” *Journal of Computational and Applied Mathematics*, vol. 372, p. 112680, 2020.

[30] T.-X. Jiang, M. K. Ng, X.-L. Zhao, and T.-Z. Huang, “Framelet representation of tensor nuclear norm for third-order tensor completion,” *IEEE Transactions on Image Processing*, vol. 29, pp. 7233–7244, 2020.

[31] H. Zeng, X. Xie, H. Cui, Y. Zhao, and J. Ning, “Hyperspectral image restoration via cnn denoiser prior regularized low-rank tensor recovery,” *Computer Vision and Image Understanding*, vol. 197-198, p. 103004, 2020.

[32] Y. Wang, J. Peng, Q. Zhao, Y. Leung, X.-L. Zhao, and D. Meng, “Hyperspectral image restoration via total variation regularized low-rank tensor decomposition,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 4, pp. 1227–1243, 2017.

[33] J. Xue, Y. Zhao, W. Liao, and J. C.-W. Chan, “Nonlocal low-rank regularized tensor decomposition for hyperspectral image denoising,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 5174–5189, 2019.

[34] Y. Xie, D. Tao, W. Zhang, Y. Liu, L. Zhang, and Y. Qu, “On unifying multi-view self-representations for clustering by tensor multi-rank minimization,” *International Journal of Computer Vision*, vol. 126, no. 11, pp. 1157–1179, 2018.

[35] X.-Y. Liu, S. Aeron, V. Aggarwal, X. Wang, and M.-Y. Wu, “Adaptive sampling of rf fingerprints for fine-grained indoor localization,” *IEEE Transactions on Mobile Computing*, vol. 15, no. 10, pp. 2411–2423, 2015.

[36] M. Mørup, L. K. Hansen, and S. M. Arnfred, “Algorithms for sparse nonnegative tucker decompositions,” *Neural Computation*, vol. 20, no. 8, pp. 2112–2131, 2008.

[37] 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 5249–5257.

[38] C. Lu, J. Feng, Z. Lin, and S. Yan, “Exact low tubal rank tensor recovery from gaussian measurements,” *arXiv preprint arXiv:1806.02511*, 2018.

[39] M. Yin, J. Gao, S. Xie, and Y. Guo, “Multiview subspace clustering via tensorial t-product representation,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 3, pp. 851–864, 2018.

[40] W. Hu, D. Tao, W. Zhang, Y. Xie, and Y. Yang, “The twist tensor nuclear norm for video completion,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 12, pp. 2961–2973, 2016.

[41] L. De Lathauwer and B. De Moor, “From matrix to tensor: Multilinear algebra and signal processing,” in *Institute of mathematics and its applications conference series*, vol. 67. Citeseer, 1998, pp. 1–16.

[42] S. Gu, L. Zhang, W. Zuo, and X. Feng, “Weighted nuclear norm minimization with application to image denoising,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2862–2869.

[43] M. Fazel, H. Hindi, and S. P. Boyd, “Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices,” in *Proceedings of the 2003 American Control Conference*, 2003., vol. 3. IEEE, 2003, pp. 2156–2162.

[44] K. Mohan and M. Fazel, “Iterative reweighted algorithms for matrix rank minimization,” *Journal of Machine Learning Research*, vol. 13, no. Nov, pp. 3441–3473, 2012.

[45] M. Razaviyayn, M. Hong, and Z.-Q. Luo, “A unified convergence analysis of block successive minimization methods for nonsmooth optimization,” *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 1126–1153, 2013.

[46] T.-Y. Ji, T.-Z. Huang, X.-L. Zhao, T.-H. Ma, and G. Liu, “Tensor completion using total variation and low-rank matrix factorization,” *Information Sciences*, vol. 326, pp. 243–257, 2016.

[47] Q. Huynh-Thu and M. Ghanbari, “Scope of validity of psnr in image/video quality assessment,” *Electronics Letters*, vol. 44, no. 13, pp. 800–801, 2008.

[48] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli et al., “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[49] L. Zhang, L. Zhang, X. Mou, and D. Zhang, “FSIM: A feature similarity index for image quality assessment,” *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.

[50] L. Wald, *Data fusion: definitions and architectures: fusion of images of different spatial resolutions*. Presses des MINES, 2002.