Journal of Machine Learning Research 23 (2022) 1-40 Submitted 4/22; Revised 6/22; Published 7/22

Smooth Robust Tensor Completion for Background/Foreground Separation with Missing Pixels: Novel Algorithm with Convergence Guarantee

Bo Shen BOSHEN@VT.EDU
Weijun Xie WXIE@VT.EDU
Zhenyu (James) Kong[*] ZKONG@VT.EDU
Department of Industrial and Systems Engineering, Virginia Tech
Blacksburg, Virginia 24061, USA
[*]Corresponding Author

Editor: Julien Mairal

Abstract

Robust PCA (RPCA) and its tensor extension, namely, Robust Tensor PCA (RTPCA), provide an effective framework for background/foreground separation by decomposing the data into low-rank and sparse components, which contain the background and the foreground (moving objects), respectively. However, in real-world applications, the presence of missing pixels is a very common and challenging issue due to errors in the acquisition process or manufacturer defects. RPCA and RTPCA are not able to recover the background and foreground simultaneously with missing pixels. This study aims to address the problem of background/foreground separation with missing pixels by combining video recovery and background/foreground separation into a single framework. To achieve this goal, a smooth robust tensor completion (SRTC) model is proposed to recover the data and decompose it into the static background and smooth foreground, respectively. An efficient algorithm based on tensor proximal alternating minimization (tenPAM) is implemented to solve the proposed model with a global convergence guarantee under very mild conditions. Extensive experiments on actual data demonstrate that the proposed method significantly outperforms the state-of-the-art approaches for background/foreground separation with missing pixels.

Keywords: Robust Tensor Completion (RTC), Spatio-temporal Continuity, Low-rank Property, Tensor Proximal Alternating Minimization (tenPAM), Global Convergence

Nomenclature

H, W, T	The height, width, and number of image frames
(r_1, r_2, r_3)	The multi-linear rank in Tucker Decomposition
λ	The balance coefficient in the proposed objective function
Ω	The index set of the observed elements
$\boldsymbol{\mathcal{X}}$	The order-3 tensor in $\mathbb{R}^{H \times W \times T}$ represented by $\{\mathbf{X}_1, \cdots, \mathbf{X}_T\}$
$\mathbf{X}_t$	t-th image frame in $\mathbb{R}^{H \times W}$
$\boldsymbol{\mathcal{L}}$	The low-rank tensor (static video background)
$\boldsymbol{\mathcal{S}}$	The smooth tensor (smooth moving objects)
$\boldsymbol{\mathcal{X}} \times_n \mathbf{V}$	The mode-n multiplication of a tensor $\boldsymbol{\mathcal{X}}$ with a matrix $\mathbf{V}$

$\boldsymbol{\mathcal{C}}$	The core tensor in Tucker decomposition
$\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3$	The factor matrices in Tucker decomposition
$\mathbf{U}$	The set of factor matrices in Tucker decomposition, namely, $\{\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3\}$
$\mathbf{U}\mathbf{U}^\top$	The set of projection matrices in Tucker decomposition, namely, $\{\mathbf{U}_1\mathbf{U}_1^\top, \mathbf{U}_2\mathbf{U}_2^\top, \mathbf{U}_3\mathbf{U}_3^\top\}$
$\boldsymbol{f}$	The auxiliary variable
$\mathbf{D}_h, \mathbf{D}_v, \mathbf{D}_t$	Three vectorizations of the difference operation along with the horizontal, vertical, and temporal directions
$\mathbf{D}$	The concatenated difference operation, namely, $[\mathbf{D}_h^\top, \mathbf{D}_v^\top, \mathbf{D}_t^\top]^\top$
$\|\cdot\|_F$	The Frobenius norm
$\|\cdot\|_1$	The ℓ_1 norm
$\|\cdot\|_2$	The ℓ_2 norm
$\|\cdot\|$	The 2-operator norm
$\|\cdot\|_{TV1}$	The anisotropic total variation norm
ρ	The positive coefficient for the proximal term
$\texttt{vec}(\cdot)$	The vectorization operator
$\texttt{ten}(\cdot)$	The tensorization operator
$\boldsymbol{\lambda}^{\boldsymbol{f}}$	The Lagrange multiplier tensor
$\beta^{\boldsymbol{f}}$	The positive penalty scalar
c_1, c_2	Coefficients in the adaptive updating scheme for $\beta^{\boldsymbol{f}}$
$\texttt{fftn}(\cdot)$	The fast 3D Fourier transform
$\texttt{ifftn}(\cdot)$	The inverse fast 3D Fourier transform
$\mathrm{soft}(\cdot,\cdot)$	The soft-thresholding operator
γ	The parameter associated with the convergence rate in ADMM
$\mathrm{Err}(\cdot)$	The error of the auxiliary variable

1. Introduction

Background/foreground separation is a fundamental step for moving object detection in many video data applications (Zhou et al., 2012). It is usually performed by separating the moving objects called "foreground" from the static objects called "background" (Bouwmans et al., 2017b). In many real-world applications, the presence of missing pixels is a very common but challenging issue (Firtha et al., 2008; Liu et al., 2012; Ren et al., 2021) due to errors in the acquisition process or manufacturer defects. In this paper, we are interested in background/foreground separation with missing pixels, as shown in Figure 1, which aims to recover the original video with high fidelity and meanwhile accurately separate the moving objects from the video background based on partially observed pixels. The video imaging system first captures missing measurements from the scenes and then transmits these measurements to the processing center for recovery and background/foreground separation.

In the current literature, research on background/foreground separation is based on the decomposition of video data into low-rank and sparse components. It is an effective framework to separate the foreground from the background, which are modeled by the sparse and low-rank components, respectively. Among them, the most representative problem formulation is the robust principal component analysis (RPCA) (Candès et al., 2011), which is a modification of the widely used statistical procedure named principal component analysis

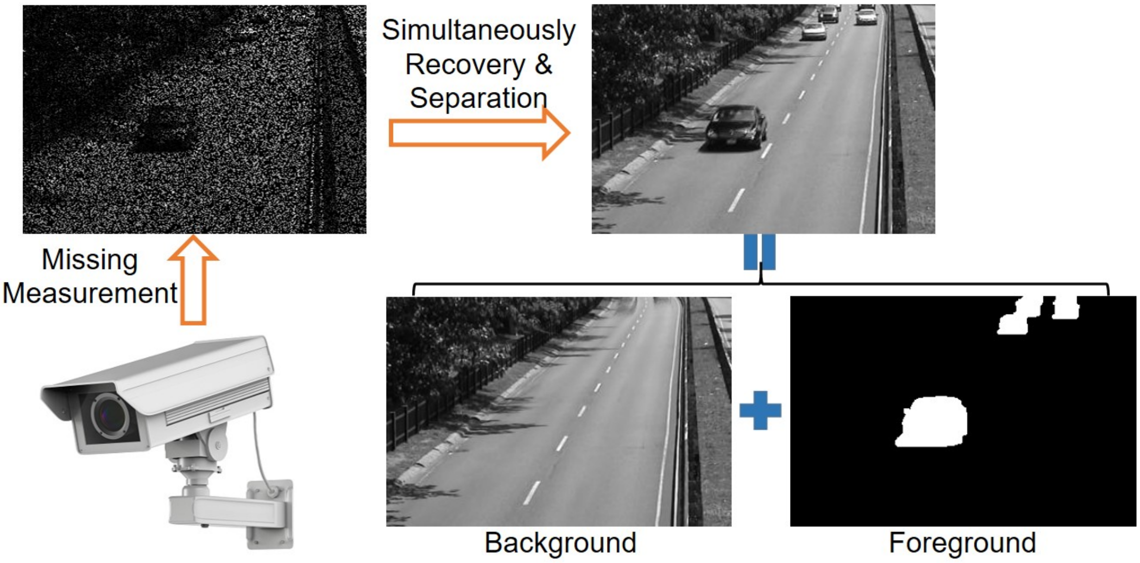


Figure 1: The framework of the imaging system with missing pixels.

(PCA). RPCA decomposes the video data $\mathbf{X}$ into the sum of a low-rank component $\mathbf{L}$ and a sparse component $\mathbf{E}$, where the low-rank property and sparsity are measured by the nuclear norm $\|\cdot\|_*$ and ℓ_1 norm $\|\cdot\|_1$, respectively. One major disadvantage of RPCA is that it can only process 2-D matrix data since the nuclear norm $\|\cdot\|_*$ is designed for matrices. However, real-world data is usually multi-dimensional in nature, where rich information is stored in multi-way arrays known as tensors (Kolda and Bader, 2009). For example, a grayscale video is 3-D data, which stacks multiple images along with the time domain; a color image is also 3-D data with three channels: red, green, and blue, where each channel is a 2-D image. To apply RPCA to these data sets, the multi-way tensor data has to be reconstructed into a matrix. Such a preprocessing usually leads to information loss and performance degradation since the structure information in the data is deteriorated. To address this issue, it is necessary to consider extending RPCA to manipulate the tensor data directly by taking advantage of its multi-dimensional structure.

Motivated by the tensor multiplication scheme on t-SVD (Kilmer and Martin, 2011), Zhang et al. (2014) proposed the tensor tubal rank as well as the tensor nuclear norm for image denoising. Based on the tensor nuclear norm, Lu et al. (2016) developed robust tensor PCA (RTPCA) by extending RPCA from a 2-D matrix to 3-D tensor data, aiming to exactly recover a low-rank tensor contaminated by sparse errors. More specifically, it tries to recover the low-rank tensor $\boldsymbol{\mathcal{L}}$ and sparse tensor $\boldsymbol{\mathcal{E}}$ from the data tensor $\boldsymbol{\mathcal{X}}$, which can be represented as $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{L}} + \boldsymbol{\mathcal{E}}$.

Recent research on the missing value estimation problem in video data is focused on matrix completion (MC) (Candès and Recht, 2009). MC is able to recover the original signal $\boldsymbol{\mathcal{X}}$ from a partially observed signal $\boldsymbol{\mathcal{X}}_\Omega$ (or called the undersampled/incomplete signal), where Ω is a subset containing 2D coordinates of sampled entries. In the pioneering work of (Candès and Recht, 2009), the signal $\boldsymbol{\mathcal{X}}$ is recovered by solving a convex relaxation of the rank minimization problem based on the nuclear norm $\|\cdot\|_*$. Furthermore, Zhou et al. (2017) extended matrix completion to tensor completion (TC) based on the tensor nuclear norm. However, none of RPCA/RTPCA and MC/TC is able to address the background/foreground separation with missing pixels because RPCA/RTPCA cannot recover the video data and MC/TC cannot separate the background and foreground. There is some work on robust matrix completion/tensor completion (Chen et al., 2015; He et al., 2019; He and Atia, 2020; Li and So, 2021; Shang et al., 2017; Huang et al., 2021), which combines RPCA/RTPCA and MC/TC. Specifically, they aim to reconstruct a signal from the noisy observations of a small, random subset of its entries. The problem with these methods is that their learned sparse component cannot represent the foreground since the sparse component tends to set the foreground to zero for the positions of those missing pixels. As a result, the obtained foreground is incomplete due to the sparsity modeling.

This study aims to address the problem of background/foreground separation with missing pixels by combining video recovery and background/foreground separation into one single framework. Compared to the conventional method, this new method need not fully sense all the video pixels and thus heavily reduces the computational and storage costs and even the energy consumption of imaging sensors. To achieve this objective, a smooth robust tensor completion (SRTC) is proposed to recover the data tensor $\boldsymbol{\mathcal{X}}$ and decompose it into a low-rank tensor (background) $\boldsymbol{\mathcal{L}}$ and a smooth tensor (foreground) $\boldsymbol{\mathcal{S}}$, namely, $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{L}} + \boldsymbol{\mathcal{S}}$. In the SRTC, the background is modeled by the low-rank Tucker decomposition (Kolda and Bader, 2009). The spatio-temporal continuity is applied to formulate the moving objects (foreground) (Cao et al., 2015, 2016; Shen et al., 2021). That is, the moving objects in the video foreground are spatially continuous in both their support regions and their intensity values in these regions. Moreover, the moving objects are also temporally continuous among succeeding frames. To summarize, the contributions of this paper are as follows:

- A smooth robust tensor completion model is proposed for background/foreground separation with missing pixels by simultaneously recovering the tensor data and decomposing it into a low-rank tensor and a smooth tensor, respectively;
- An efficient tensor proximal alternating minimization (namely, tenPAM) algorithm is implemented to solve the proposed model;
- Convergence of the iterative sequence generated by the tenPAM algorithm is analyzed, which is proved to be globally convergent to the stationary points under very mild conditions.

The remainder of this paper is organized as follows. A brief review of notation and related research work is provided in Section 2. The proposed model and algorithm to solve this model are introduced in Section 3, followed by the convergence analysis in Section 4. Numerical studies in Section 5 are provided for testing and validation of the proposed method. Finally, the conclusions are discussed in Section 6.

Remark 1 (Background Initialization) *Background/foreground separation aims to detect moving foreground objects from a video stream (Garcia-Garcia et al., 2020). It includes the following steps:*

(1) the background initialization module provides the clean background image from all training frames,

(2) Foreground detection classifying pixels as foreground or background is achieved by comparing the background and current images, and

(3) If there is a future unseen image, the background maintenance module updates the background image using the previous background, the current image, and the foreground detection mask.

Background initialization is a fundamental step in obtaining a clean background without foreground objects (Bouwmans et al., 2017a).
Background initialization can be formulated as a tensor completion task (Sobral and Zahzah, 2017; Javed et al., 2017; Kajo et al., 2018, 2019). The missing pixels are induced from the moving regions through a simple joint motion-detection and frame-selection operation. Next, tensor completion is applied to recover the background as the low-rank component from the partially observed data. However, these approaches still have the full observations of all pixels, where missing pixels are generated by some algorithms for background initialization. In our paper, different from these approaches (Sobral and Zahzah, 2017; Javed et al., 2017; Kajo et al., 2018, 2019), missing pixels are generated in the acquisition process or manufacturer defects, where we want to recover the information of these missing pixels using our proposed method.

2. Notation and Research Background

In Section 2.1, the notation and basics of multi-linear algebra used in this paper are reviewed. Then, the robust tensor PCA, tensor completion, and robust matrix/tensor completion are reviewed briefly in Section 2.2. Afterward, the research gaps in the existing work are identified in Section 2.3.

2.1 Notation and Tensor Basis

Throughout this paper, scalars are denoted by lowercase letters, e.g., x; vectors are denoted by lowercase boldface letters, e.g., $\boldsymbol{x}$; matrices are denoted by uppercase boldface, e.g., $\mathbf{X}$; and tensors are denoted by calligraphic letters, e.g., $\boldsymbol{\mathcal{X}}$. The order of a tensor is the number of its modes or dimensions. A real-valued tensor of order-N is denoted by $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and its entries by $\boldsymbol{\mathcal{X}}(i_1, i_2, \cdots, i_N)$. The multi-linear Tucker rank of an order-N tensor is the tuple of the ranks of the mode-n unfolding $\mathbf{X}_{(n)} \in \mathbb{R}^{I_n \times (I_1 \times \cdots \times I_{n-1} \times I_{n+1} \times \cdots \times I_N)}$. The inner product of two same-sized tensors $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ is the sum of the products of their entries, namely, $\langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}} \rangle = \sum_{i_1} \cdots \sum_{i_N} \boldsymbol{\mathcal{X}}(i_1, \ldots, i_N) \cdot \boldsymbol{\mathcal{Y}}(i_1, \ldots, i_N)$. Following the definition of inner product, the Frobenius norm of a tensor $\boldsymbol{\mathcal{X}}$ is defined as $\|\boldsymbol{\mathcal{X}}\|_F = \sqrt{\langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{X}} \rangle}$. The mode-$n$ multiplication of a tensor $\boldsymbol{\mathcal{X}}$ with a matrix $\mathbf{U}$ amounts to the multiplication of all mode-n vector fibers with $\mathbf{U}$, namely, $(\boldsymbol{\mathcal{X}} \times_n \mathbf{U})(i_1, \cdots, i_{n-1}, j_n, i_{n+1}, \cdots, i_N) = \sum_{i_n} \boldsymbol{\mathcal{X}}(i_1, \cdots, i_N) \cdot \mathbf{U}(j_n, i_n)$. Unfolding $\boldsymbol{\mathcal{X}}$ along the n-mode is denoted as $\mathbf{X}_{(n)} \in \mathbb{R}^{I_n \times (I_1 \times \cdots \times I_{n-1} \times I_{n+1} \times \cdots \times I_N)}$, in particular, if $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{C}} \times_1 \mathbf{U}^{(1)} \times_2 \cdots \times_N \mathbf{U}^{(N)}$ with $\boldsymbol{\mathcal{C}} \in \mathbb{R}^{P_1 \times \cdots \times P_N}$ and $\mathbf{U}$, then $\mathbf{X}_{(n)} = \mathbf{U}^{(n)} \mathbf{C}_{(n)} (\mathbf{U}^{(N)} \otimes \cdots \otimes \mathbf{U}^{(n+1)} \otimes \mathbf{U}^{(n-1)} \otimes \cdots \otimes \mathbf{U}^{(1)})^\top$, where $\otimes$ is the Kronecker product.

2.2 Related Work

In this subsection, three directions of related work to motivate the research in this paper are introduced here.

2.2.1 Robust Tensor PCA

As the tensor extension of the popular robust PCA (Candès et al., 2011), the recent proposed RTPCA (Lu et al., 2016) aims to recover the low-rank tensor $\boldsymbol{\mathcal{L}}_0 \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ and sparse tensor $\boldsymbol{\mathcal{E}}_0 \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ from their sum. RTPCA solves the following convex optimization problem

$$\min_{\boldsymbol{\mathcal{L}},\boldsymbol{\mathcal{S}}} \|\boldsymbol{\mathcal{L}}\|_{TNN} + \lambda \|\boldsymbol{\mathcal{E}}\|_1, \text{ s.t. } \boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{L}} + \boldsymbol{\mathcal{E}},$$

where $\|\cdot\|_{TNN}$ is their proposed tensor nuclear norm, which is a convex relaxation of the tensor tubal rank. The tensor nuclear norm and tensor tubal rank are defined based on the t-SVD proposed in (Zhang et al., 2014). Following this direction, to further exploit the low-rank structures in tensor data, Liu et al. (2018) extracted a low-rank component for the core matrix whose entries are from the diagonal elements of the core tensor. Based on this idea, they defined a new tensor nuclear norm and proposed a creative algorithm to deal with RTPCA problems. Other than the work based on the tensor tubal rank, Yang et al. (2020) considered a new model for RTPCA based on tensor train rank. These methods are applied to background/foreground separation, image/video denoising, etc.

2.2.2 Tensor Completion

Motivated by tensor nuclear norm, Zhou et al. (2017) proposed a novel low-rank tensor factorization method for efficiently solving the order-3 tensor completion problem. It aims at exactly recovering a low-rank tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ from an incomplete observation $\boldsymbol{\mathcal{F}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$. Accordingly, its mathematical model is written as

$$\min_{\boldsymbol{\mathcal{X}}} \|\boldsymbol{\mathcal{X}}\|_{TNN}, \text{ s.t. } \boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{X}}) = \boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{F}}),$$

where Ω is the index set of the observed elements, $\boldsymbol{\mathcal{P}}$ is a linear operator that extracts entries in Ω and fills the entries not in Ω with zeros. In the optimization process, their method only needs to update two smaller tensors, which can be more efficiently conducted than computing t-SVD. Furthermore, they prove that the proposed alternating minimization algorithm can converge to a Karush–Kuhn–Tucker point.

2.2.3 Robust Matrix/Tensor Completion

Robust matrix completion aims to recover a low-rank matrix $\mathbf{L} \in \mathbb{R}^{I_1 \times I_2}$ from a subset of noisy entries perturbed by complex noises. Chen et al. (2015) provided a robust matrix completion model as follows

$$\min_{\mathbf{L},\mathbf{E}} \|\mathbf{L}\|_* + \lambda \|\mathbf{E}\|_{2,1}, \text{ s.t. } \boldsymbol{\mathcal{P}}_\Omega(\mathbf{X}) = \boldsymbol{\mathcal{P}}_\Omega(\mathbf{L} + \mathbf{E}),$$

where $\|\cdot\|_*$ is the matrix nuclear norm, and $\|\cdot\|_{2,1}$ is the sum of the column $\ell 2$ norms of a matrix and a convex surrogate of its column sparsity. Fan et al. (2017) proposed novel bilinear factor matrix norm minimization models by defining the double nuclear norm and Frobenius/nuclear hybrid norm. He et al. (2019) proposed a robust and fast matrix completion method based on the maximum correntropy criterion, which is extended to the tensor version in the work of (He and Atia, 2020). Li et al. (2021) considered column outliers and sparse noise. The $\ell_{2,1}$ norm based objective function makes the recovered matrix keeps a low-rank structure and lets the algorithm robust to column outliers, while the regularization term based on ℓ_1 norm can alleviate the influence of sparse noise. Huang et al. (2021) proposed robust tensor ring completion, where the low-rank tensor component is constrained by the weighted sum of nuclear norms of its balanced unfolding, while the sparse component is regularized by its ℓ_1 norm.

2.3 Research Gap Identification

In real-world applications, video data often contains missing pixels due to errors in the acquisition process or manufacturer defects. If the RPCA/RTPC (Candès et al., 2011; Lu et al., 2016) in Section 2.2.1 is applied to the video with missing pixels, the background and foreground are incomplete since RPCA/RTPCA cannot recover missing pixels. If matrix/tensor completion in Section 2.2.2 is applied, the background and foreground cannot be separated because they cannot decompose the video data. If robust matrix/tensor completion in Section 2.2.3 is applied, the foreground recovery is not guaranteed. This is due to the fact that the foreground is represented by the sparse component. Specifically, suppose a part of the foreground is missing in the data. In that case, there is no way to recover the missing part since the optimization problem will set this part to zero because it is modeled by the sparse component. Therefore, this work seeks to address these research gaps by devising a new smooth robust tensor completion (SRTC) model. The proposed model can be considered as separating background/foreground together with video recovery by providing a new decomposition methodology with missing pixels.

3. Proposed Method

In Section 3.1, the proposed smooth RTC for the background/foreground separation with missing pixels is presented. Specifically, the low-rank property and spatio-temporal continuity are formulated by Tucker decomposition and total variation (TV) regularization, respectively. In Section 3.2, an efficient algorithm based on proximal alternating minimization (PAM) (Attouch et al., 2010) is designed to solve the proposed model.

3.1 Proposed Model

This work is focused on videos that can be represented as an order-3 tensor $\boldsymbol{\mathcal{X}} := \{\mathbf{X}_1, \cdots, \mathbf{X}_T\} \in \mathbb{R}^{H\times W\times T}$, where each matrix $\mathbf{X}_t \in \mathbb{R}^{H\times W}$ represents t-th image frame $t = 1, \ldots, T$. H, W, and T denote the height, width of an image frame, and the number of image frames, respectively. The three modes of tensor $\boldsymbol{\mathcal{X}}$ are the height, width, and time of the video, respectively.

As discussed in Section 2.3, for a video with missing pixels, it is necessary to recover the video data $\boldsymbol{\mathcal{X}}$ and decompose it into the low-rank tensor $\boldsymbol{\mathcal{L}}$ (the static video background), the smooth tensor $\boldsymbol{\mathcal{S}}$ (the smooth moving objects in the foreground), respectively. In the static background, the image frames are the same along with the time domain. This can be achieved by restricting $\boldsymbol{\mathcal{L}}$ to be a low-rank tensor in the time domain. The moving objects in the video foreground are spatially and temporally continuous, so they can be represented as a smooth tensor $\boldsymbol{\mathcal{S}}$. An illustration of the video decomposition strategy for our proposed

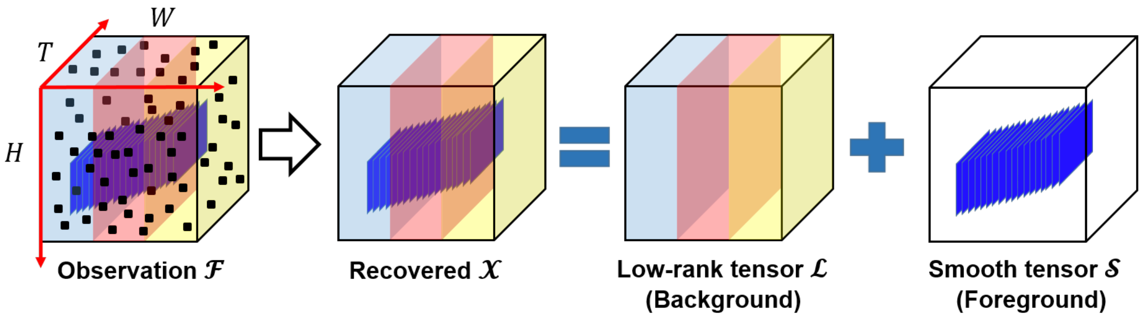

W
T
H
Observation $\mathcal{F}$
Recovered $\mathcal{X}$
Low-rank tensor $\mathcal{L}$
(Background)
Smooth tensor $\mathcal{S}$
(Foreground)

Figure 2: Illustration of the decomposition strategy of a video in the proposed method.

method is provided in Figure 2. Specifically, it has the following form $\mathcal{X} = \mathcal{L} + \mathcal{S}$, as mentioned in Section 1.

To model the low-rank property, the static background $\mathcal{L}$ is approximated by the well-known Tucker decomposition (Kolda and Bader, 2009) with rank-(r_1, r_2, r_3). Specifically, Tucker decomposition has the following form

$$\mathcal{L} = \mathcal{C} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \mathbf{U}^{(3)}, \tag{1}$$

where $\mathbf{U}^{(1)} \in \mathbb{R}^{H \times r_1}, r_1 < H$ and $\mathbf{U}^{(2)} \in \mathbb{R}^{W \times r_2}, r_2 < W$ are orthogonal factor matrices for two spatial domains, $\mathbf{U}^{(3)} \in \mathbb{R}^{T \times r_3}, r_3 < T$ is the orthogonal factor matrix for the temporal domain, the core tensor $\mathcal{C} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ interacts these factors. By formulating the low-rank tensor $\mathcal{L}$ using Tucker decomposition, it can reconstruct a more accurate video background than the low-rank model based on matrices. Because Tucker decomposition considers not only the spatial but also the temporal correlations in the video background.

The smooth tensor $\mathcal{S}$ (moving objects) is assumed to have the spatio-temporal continuity property such that the foreground moves smoothly and coherently in the spatial and temporal directions. In the literature, imposing the spatio-temporal continuity constraints on moving objects in the foreground is well studied and proven to be effective (Cao et al., 2015, 2016). To measure the sensitivity to change of a quantity function, the derivative is often applied in mathematics. For discrete functions, difference operators are the approximation to derivative. Given an order-3 tensor $\mathcal{S} \in \mathbb{R}^{H \times W \times T}$, $\mathcal{S}(x, y, t)$ indicates the intensity of position (x, y) at time t, and

$$\begin{aligned}
\mathcal{S}_h(x, y, t) &= \mathcal{S}_h(x+1, y, t) - \mathcal{S}_h(x, y, t),\\
\mathcal{S}_v(x, y, t) &= \mathcal{S}_v(x, y+1, t) - \mathcal{S}_v(x, y, t),\\
\mathcal{S}_t(x, y, t) &= \mathcal{S}_t(x, y, t+1) - \mathcal{S}_t(x, y, t)
\end{aligned}$$

denote three difference operation results of position (x, y) at time t with periodic boundary conditions along with the horizontal, vertical, and temporal directions, respectively. For simplicity of computation, all the entries of $\mathcal{S}$ can be stacked into a column vector $\boldsymbol{s} = \mathtt{vec}(\mathcal{S})$, in which $\mathtt{vec}(\cdot)$ represents the vectorization operator. $\mathbf{D}_h\boldsymbol{s} = \mathtt{vec}(\mathcal{S}_h)$, $\mathbf{D}_v\boldsymbol{s} = \mathtt{vec}(\mathcal{S}_v)$, and $\mathbf{D}_t\boldsymbol{s} = \mathtt{vec}(\mathcal{S}_t)$ are used to denote the vectorizations of the three difference operation results, respectively, in which $\mathbf{D}_h$, $\mathbf{D}_v$, and $\mathbf{D}_t \in \mathbb{R}^{HWT \times HWT}$. Furthermore, $\mathbf{D}\boldsymbol{s} = [\mathbf{D}_h\boldsymbol{s}^\top, \mathbf{D}_v\boldsymbol{s}^\top, \mathbf{D}_t\boldsymbol{s}^\top]^\top$ is used to represent the concatenated difference operation, in which $\mathbf{D} = [\mathbf{D}_h^\top, \mathbf{D}_v^\top, \mathbf{D}_t^\top]^\top \in \mathbb{R}^{3HWT \times HWT}$. Note that the i-th element in $\mathbf{D}_h\boldsymbol{s}$, $\mathbf{D}_v\boldsymbol{s}$, and

$\mathbf{D}_t\boldsymbol{s}$ (namely, $[\mathbf{D}_h\boldsymbol{s}]_i$, $[\mathbf{D}_v\boldsymbol{s}]_i$, and $[\mathbf{D}_t\boldsymbol{s}]_i$) describes the intensity changes in i-th point in $\boldsymbol{s}$ along with the horizontal, vertical, and temporal directions, respectively. To quantify the changes of intensity, any vector norm of $\big[[\mathbf{D}_h\boldsymbol{s}]_i, [\mathbf{D}_v\boldsymbol{s}]_i, [\mathbf{D}_t\boldsymbol{s}]_i\big]$ can be applied. The commonly used vector norm is the ℓ_1 norm. Specifically, the anisotropic total variation norm is defined as

$$\|\boldsymbol{\mathcal{S}}\|_{TV1} = \sum_i (|[\mathbf{D}_h\boldsymbol{s}]_i| + |[\mathbf{D}_v\boldsymbol{s}]_i| + |[\mathbf{D}_t\boldsymbol{s}]_i|), \tag{2}$$

which is the ℓ_1 norm of $[\mathbf{D}_h\boldsymbol{s}, \mathbf{D}_v\boldsymbol{s}, \mathbf{D}_t\boldsymbol{s}]^\top$. The total variation regularization has been widely used in image and video denoising and restoration (Wang et al., 2017; Zhang et al., 2019) due to its superiority in detecting discontinuous changes in image processing.

By combining the advantages of Tucker decomposition for the low-rank tensor and total variation regularization for the smooth tensor, the proposed problem has the following formulation

$$\begin{aligned} \min_{\boldsymbol{\mathcal{C}},\mathbf{U},\boldsymbol{\mathcal{S}},\boldsymbol{\mathcal{X}}} \quad & \|\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{C}} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \mathbf{U}^{(3)}\|_F^2 + \lambda\|\boldsymbol{\mathcal{S}}\|_{TV1} \\ \text{s.t.} \quad & \boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{X}}) = \boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{F}}), \\ & \mathbf{U}^{(n)\top}\mathbf{U}^{(n)} = \mathbf{I}, n = 1, 2, 3, \end{aligned} \tag{3}$$

where $\boldsymbol{\mathcal{F}}$ is the partially observed tensor, Ω is the index set of the observed elements, the first term is the fitting error, and the second term is the regularization term to measure the spatio-temporal continuity of $\boldsymbol{\mathcal{S}}$. For notational convenience, let the core tensor be $\boldsymbol{\mathcal{C}} \in \mathbb{R}^{r_1\times r_2\times r_3}$. The set of multilinear subspaces, namely, $\mathbf{U}$, is defined as $\{(\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \mathbf{U}^{(3)}) : \mathbf{U}^{(n)\top}\mathbf{U}^{(n)} = \mathbf{I}, n = 1, 2, 3\}$. $\lambda > 0$ is the coefficient for the regularization term in (3). The optimization problem in (3) is based on the decision variables $\{\boldsymbol{\mathcal{C}}, \mathbf{U}, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{X}}\}$. The following proposition states that $\boldsymbol{\mathcal{C}}$ can be projected out from the original formulation.

Proposition 2 *Suppose* $(\boldsymbol{\mathcal{C}}^*, \mathbf{U}^*, \mathbf{S}^*, \boldsymbol{\mathcal{X}}^*)$ *is an optimal solution of the proposed formulation* (3)*, then*

$$\boldsymbol{\mathcal{C}}^* = (\boldsymbol{\mathcal{X}}^* - \boldsymbol{\mathcal{S}}^*) \times_1 \mathbf{U}^{(1)*\top} \times_2 \mathbf{U}^{(2)*\top} \times_3 \mathbf{U}^{(3)*\top}.$$

Proof Suppose the optimal $\mathbf{U}^*, \mathbf{S}^*, \boldsymbol{\mathcal{X}}^*$ are given. Then the first-order optimality condition with respect to $\boldsymbol{\mathcal{C}}$ is

$$2\big[-(\boldsymbol{\mathcal{X}}^* - \boldsymbol{\mathcal{S}}^*) \times_1 \mathbf{U}^{(1)*\top} \times_2 \mathbf{U}^{(2)*\top} \times_3 \mathbf{U}^{(3)*\top} + \boldsymbol{\mathcal{C}}\big] = \mathbf{0}.$$

We must have

$$\boldsymbol{\mathcal{C}}^* = (\boldsymbol{\mathcal{X}}^* - \boldsymbol{\mathcal{S}}^*) \times_1 \mathbf{U}^{(1)*\top} \times_2 \mathbf{U}^{(2)*\top} \times_3 \mathbf{U}^{(3)*\top}.$$

■

Proposition 2 implies that there are only three types of decision variables $\mathbf{U}, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{X}}$ in the proposed formulation (3), namely, by projecting out variables $\boldsymbol{\mathcal{C}}$, which can be simplified as

$$\begin{aligned} \min_{\mathbf{U},\boldsymbol{\mathcal{S}},\boldsymbol{\mathcal{X}}} \quad & \hat{F}(\mathbf{U}\mathbf{U}^\top, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{X}}) \\ \text{s.t.} \quad & \boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{X}}) = \boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{F}}), \\ & \mathbf{U}^{(n)\top}\mathbf{U}^{(n)} = \mathbf{I}, n = 1, 2, 3, \end{aligned} \tag{4}$$

where $\hat{F}(\mathbf{U}\mathbf{U}^\top, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{X}}) := \|\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{S}} - (\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{S}}) \times_1 \mathbf{U}^{(1)}\mathbf{U}^{(1)\top} \times_2 \mathbf{U}^{(2)}\mathbf{U}^{(2)\top} \times_3 \mathbf{U}^{(3)}\mathbf{U}^{(3)\top}\|_F^2 + \lambda\|\boldsymbol{\mathcal{S}}\|_{TV1}$ and $\mathbf{U}\mathbf{U}^\top$ is defined as $\{(\mathbf{U}^{(1)}\mathbf{U}^{(1)\top}, \mathbf{U}^{(2)}\mathbf{U}^{(2)\top}, \mathbf{U}^{(3)}\mathbf{U}^{(3)\top}) : \mathbf{U}^{(n)\top}\mathbf{U}^{(n)} = \mathbf{I}, n = 1, 2, 3\}$.

The following proposition shows the representation of $\mathbf{U}\mathbf{U}^\top$ is unique, but not for $\mathbf{U}$.

Proposition 3 *Given $\mathbf{W}^\top\mathbf{W} = \mathbf{V}^\top\mathbf{V} = \mathbf{I}, \mathbf{W}, \mathbf{V} \in \mathbb{R}^{I\times r}, r < I$, $\|\mathbf{W}\mathbf{W}^\top - \mathbf{V}\mathbf{V}^\top\|_F^2 = 0$ if and only if there exists an orthogonal matrix $\mathbf{R} \in \mathbb{R}^{r\times r}$ such that $\mathbf{W} = \mathbf{V}\mathbf{R}$.*

Proof if: This follows from the straightforward calculation.

only if: $\|\mathbf{W}\mathbf{W}^\top - \mathbf{V}\mathbf{V}^\top\|_F^2 = 0$ implies that $\mathbf{W}\mathbf{W}^\top = \mathbf{V}\mathbf{V}^\top$. Since $\mathbf{W}^\top\mathbf{W} = \mathbf{V}^\top\mathbf{V} = \mathbf{I}$, we further have

$$\mathbf{W}\mathbf{W}^\top\mathbf{V} = \mathbf{V}\mathbf{V}^\top\mathbf{V} = \mathbf{V},$$

namely, columns of $\mathbf{V}$ are distinct eigenvectors of $\mathbf{W}\mathbf{W}^\top$, where their corresponding eigenvalues are equal to 1. Therefore, $\mathbf{W}$ and $\mathbf{V}$ have the same column spaces (Strang, 2016). ■

3.2 Tensor Proximal Alternating Minimization Algorithm

Note that (4) is a multivariate optimization problem. Alternating minimization algorithm (Attouch et al., 2013) is commonly used to solve multivariate optimization problems due to its simplicity and efficiency. To enhance the theoretical convergence and numerical stability of the alternating minimization algorithm, proximal terms are suggested to add in sub-problems arising from the alternating minimization algorithm, which is called proximal alternating minimization (PAM) algorithm. In this section, a tensor PAM (tenPAM) algorithm is developed for solving (4).

Given the solution from k-th iterations $(\mathbf{U}_k\mathbf{U}_k^\top, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k)$ for the problem (4), then the tenPAM iterates the following three parts: (1) fix $(\boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k)$, we solve the optimization problem over $\mathbf{U}$ in Section 3.2.1 to obtain $\mathbf{U}_{k+1}$; (2) fix $(\mathbf{U}_{k+1}, \boldsymbol{\mathcal{X}}^k)$, we solve the optimization problem over $\boldsymbol{\mathcal{S}}$ in Section 3.2.2 to obtain $\boldsymbol{\mathcal{S}}^{k+1}$; (3) fix $(\mathbf{U}_{k+1}, \boldsymbol{\mathcal{S}}^{k+1})$, we solve the optimization problem over $\boldsymbol{\mathcal{X}}$ in Section 3.2.3 to obtain $\boldsymbol{\mathcal{X}}^{k+1}$.

3.2.1 Optimization over U

At iteration $k+1$, assuming that $\boldsymbol{\mathcal{S}}^k$ and $\boldsymbol{\mathcal{X}}^k$ are fixed, we solve the below problem to obtain $\mathbf{U}_{k+1}^{(n)}$ sequentially $(n = 1, 2, 3)$:

$$\begin{aligned} \min_{\mathbf{U}^{(n)}} \quad & \hat{F}(\{\mathbf{U}_{k+1}^{(i)}\mathbf{U}_{k+1}^{(i)\top}\}_{i<n}, \mathbf{U}^{(n)}\mathbf{U}^{(n)\top}, \{\mathbf{U}_k^{(i)}\mathbf{U}_k^{(i)\top}\}_{i>n}, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k) \\ & + \rho\|\mathbf{U}^{(n)}\mathbf{U}^{(n)\top} - \mathbf{U}_k^{(n)}\mathbf{U}_k^{(n)\top}\|_F^2 \\ \text{s.t.} \quad & \mathbf{U}^{(n)\top}\mathbf{U}^{(n)} = \mathbf{I}, \end{aligned} \tag{5}$$

where $\rho\|\mathbf{U}^{(n)}\mathbf{U}^{(n)\top} - \mathbf{U}_k^{(n)}\mathbf{U}_k^{(n)\top}\|_F^2$ is the proximal term, $\rho > 0$ is the positive coefficient. Problem (5) can be equivalently formulated as a standard trace optimization problem as

$$\mathbf{U}_{k+1}^{(n)} \in \arg\max_{\mathbf{U}^{(n)}} \left\{ \mathrm{Tr}(\mathbf{U}^{(n)\top}\boldsymbol{\Psi}_k^{(n)}\mathbf{U}^{(n)}) - \rho\|\mathbf{U}^{(n)}\mathbf{U}^{(n)\top} - \mathbf{U}_k^{(n)}\mathbf{U}_k^{(n)\top}\|_F^2 : \mathbf{U}^{(n)\top}\mathbf{U}^{(n)} = \mathbf{I} \right\}, \tag{6}$$

where $\mathbf{\Psi}_k^{(n)} = (\mathbf{X}_{(n)}^k - \mathbf{S}_{(n)}^k) \cdot \mathbf{U}_{\mathbf{\Psi}_k^{(n)}} \cdot \mathbf{U}_{\mathbf{\Psi}_k^{(n)}}{}^\top \cdot (\mathbf{X}_{(n)}^k - \mathbf{S}_{(n)}^k)^\top$ and $\mathbf{U}_{\mathbf{\Psi}_k^{(n)}} = \mathbf{U}_k^{(3)} \otimes \cdots \otimes \mathbf{U}_k^{(n+1)} \otimes \mathbf{U}_{k+1}^{(n-1)} \otimes \cdots \otimes \mathbf{U}_{k+1}^{(1)}$. The following lemma shows that we can absorb the penalty term so that problem (6) has a closed-form optimal solution by redefining matrices $\{\mathbf{\Psi}_k^{(n)}\}_{n\in[3]}$ for all k.

Lemma 4 *Problem* (6) *is equivalent to*

$$\mathbf{U}_{k+1}^{(n)} \in \arg\max_{\mathbf{U}^{(n)}} \left\{ \mathrm{Tr}(\mathbf{U}^{(n)\top} \mathbf{\Phi}_k^{(n)} \mathbf{U}^{(n)}) : \mathbf{U}^{(n)\top}\mathbf{U}^{(n)} = \mathbf{I} \right\}, \tag{7}$$

where $\mathbf{\Phi}_k^{(n)} = \mathbf{\Psi}_k^{(n)} - 2\rho(\mathbf{I} - \mathbf{U}_k^{(n)}\mathbf{U}_k^{(n)\top})$. *In addition, problem* (7) *is a standard eigen-decomposition problem, which has a closed-form solution.*

Proof On the one hand, we have

$$\begin{aligned} &\|\mathbf{U}^{(n)}\mathbf{U}^{(n)\top} - \mathbf{U}_k^{(n)}\mathbf{U}_k^{(n)\top}\|_F^2 \\ =&\mathrm{Tr}\Big((\mathbf{U}^{(n)}\mathbf{U}^{(n)\top} - \mathbf{U}_k^{(n)}\mathbf{U}_k^{(n)\top})(\mathbf{U}^{(n)}\mathbf{U}^{(n)\top} - \mathbf{U}_k^{(n)}\mathbf{U}_k^{(n)\top})^\top\Big) \\ =&2\Big(r_n - \mathrm{Tr}(\mathbf{U}_k^{(n)}\mathbf{U}_k^{(n)\top}\mathbf{U}^{(n)}\mathbf{U}^{(n)\top})\Big). \end{aligned} \tag{8}$$

On the other hand, we have

$$\begin{aligned} &\mathrm{Tr}\Big(\mathbf{U}^{(n)\top}(\mathbf{I} - \mathbf{U}_k^{(n)}\mathbf{U}_k^{(n)\top})\mathbf{U}^{(n)}\Big) \\ =&\mathrm{Tr}\Big(\mathbf{U}^{(n)}\mathbf{U}^{(n)\top} - \mathbf{U}_k^{(n)}\mathbf{U}_k^{(n)\top}\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}\Big) \\ =&r_n - \mathrm{Tr}(\mathbf{U}_k^{(n)}\mathbf{U}_k^{(n)\top}\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}). \end{aligned} \tag{9}$$

According to the above two equations (8) and (9), the following can be derived

$$\|\mathbf{U}^{(n)}\mathbf{U}^{(n)\top} - \mathbf{U}_k^{(n)}\mathbf{U}_k^{(n)\top}\|_F^2 = 2\mathrm{Tr}\Big(\mathbf{U}^{(n)\top}(\mathbf{I} - \mathbf{U}_k^{(n)}\mathbf{U}_k^{(n)\top})\mathbf{U}^{(n)}\Big). \tag{10}$$

Thus,

$$\begin{aligned} &\mathrm{Tr}(\mathbf{U}^{(n)\top}\mathbf{\Psi}_k^{(n)}\mathbf{U}^{(n)}) - \rho\|\mathbf{U}^{(n)}\mathbf{U}^{(n)\top} - \mathbf{U}_k^{(n)}\mathbf{U}_k^{(n)\top}\|_F^2 \\ =&\mathrm{Tr}(\mathbf{U}^{(n)\top}\mathbf{\Psi}_k^{(n)}\mathbf{U}^{(n)}) - 2\rho\mathrm{Tr}\Big(\mathbf{U}^{(n)\top}(\mathbf{I} - \mathbf{U}_k^{(n)}\mathbf{U}_k^{(n)\top})\mathbf{U}^{(n)}\Big) \\ =&\mathrm{Tr}(\mathbf{U}^{(n)\top}\mathbf{\Phi}_k^{(n)}\mathbf{U}^{(n)}), \end{aligned}$$

where the first equality is due to (10) and $\mathbf{\Phi}_k^{(n)} = \mathbf{\Psi}_k^{(n)} - 2\rho(\mathbf{I} - \mathbf{U}_k^{(n)}\mathbf{U}_k^{(n)\top})$. ■

Remark 5 *There are two reasons to use* $\rho\|\mathbf{U}^{(n)}\mathbf{U}^{(n)\top} - \mathbf{U}_k^{(n)}\mathbf{U}_k^{(n)\top}\|_F^2$ *as the proximal term instead of* $\rho\|\mathbf{U}^{(n)} - \mathbf{U}_k^{(n)}\|_F^2$.

1. *Tucker decomposition of a tensor is not unique due to the possible orthogonal transformations of basis matrices. This is known as the rotation indeterminacy of tensors. For example, the solutions $\mathbf{U}^{(n)}$ and $\mathbf{U}^{(n)}\mathbf{A}$ can achieve the same performance, where $\mathbf{A}$ is an orthogonal matrix. But the $\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}$ is unique for Tucker decomposition, as shown in Proposition 3.*

2. *If one uses $\rho\|\mathbf{U}^{(n)}-\mathbf{U}_k^{(n)}\|_F^2$ in (5), then it becomes a very difficult optimization problem to solve. All current algorithms (Chen et al., 2020; Wang et al., 2020) can guarantee the convergence to a critical point. However, we need to solve the optimization problem (5) to optimality so that all further theoretical results will hold.*

3.2.2 Optimization over $\boldsymbol{\mathcal{S}}$

In this case, we fix all $(\mathbf{U}_{k+1}, \boldsymbol{\mathcal{X}}^k)$. Here, $\boldsymbol{\mathcal{S}}^{k+1}$ is updated through

$$\begin{aligned}\boldsymbol{\mathcal{S}}^{k+1} \in \arg\min_{\boldsymbol{\mathcal{S}}}\Big\{&\|\boldsymbol{\mathcal{X}}^k - \boldsymbol{\mathcal{S}} - (\boldsymbol{\mathcal{X}}^k - \boldsymbol{\mathcal{S}}^k) \times_1 \mathbf{U}_{k+1}^{(1)}\mathbf{U}_{k+1}^{(1)\top} \times_2 \mathbf{U}_{k+1}^{(2)}\mathbf{U}_{k+1}^{(2)\top} \times_3 \mathbf{U}_{k+1}^{(3)}\mathbf{U}_{k+1}^{(3)\top}\|_F^2 \\ &+ \lambda\|\boldsymbol{\mathcal{S}}\|_{TV1} + \rho\|\boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{S}}^k\|_F^2\Big\},\end{aligned} \tag{11}$$

where $\rho\|\boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{S}}^k\|_F^2$ is the proximal term. This problem (11) is a strictly convex optimization problem with the strongly convex objective function, which possesses a global and unique minimizer. There are many efficient solvers for finding the global minimizer of (11); see, e.g., the Bregman methods (Goldstein and Osher, 2009), proximal splitting methods (Combettes and Pesquet, 2011), and alternating direction of multiplier methods (ADMM) (Hong and Luo, 2017). We choose ADMM as a solver for the minimization problems (11), since the convergence of ADMM for (11) is theoretically guaranteed (Hong and Luo, 2017).

The optimization problem (11) is equivalent to the following equality constrained one

$$\begin{aligned}\min_{\boldsymbol{\mathcal{S}},\boldsymbol{f}} \quad & \|\boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{E}}_{k+1}\|_F^2 + \frac{\lambda}{1+\rho}\|\boldsymbol{f}\|_1 \\ \text{s.t.} \quad & \boldsymbol{f} = \mathbf{D}\mathtt{vec}(\boldsymbol{\mathcal{S}}),\end{aligned} \tag{12}$$

where $\boldsymbol{\mathcal{E}}_{k+1} = \frac{\boldsymbol{\mathcal{X}}^k - (\boldsymbol{\mathcal{X}}^k - \boldsymbol{\mathcal{S}}^k)\times_1\mathbf{U}_{k+1}^{(1)}\mathbf{U}_{k+1}^{(1)\top}\times_2\mathbf{U}_{k+1}^{(2)}\mathbf{U}_{k+1}^{(2)\top}\times_3\mathbf{U}_{k+1}^{(3)}\mathbf{U}_{k+1}^{(3)\top}+\rho\boldsymbol{\mathcal{S}}^k}{1+\rho}$. Note that the objective function in (12) is the summation of two single variable functions with $\boldsymbol{\mathcal{S}}$ and $\boldsymbol{f}$ as their individual variables. Thus, ADMM is applicable. The augmented Lagrangian function of problem (12) can be written as

$$L_A(\boldsymbol{\mathcal{S}}, \boldsymbol{f}) = \|\boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{E}}_{k+1}\|_F^2 + \frac{\lambda}{1+\rho}\|\boldsymbol{f}\|_1 - \langle \boldsymbol{\lambda}^{\boldsymbol{f}}, \boldsymbol{f} - \mathbf{D}\mathtt{vec}(\boldsymbol{\mathcal{S}})\rangle + \frac{\beta^{\boldsymbol{f}}}{2}\|\boldsymbol{f} - \mathbf{D}\mathtt{vec}(\boldsymbol{\mathcal{S}})\|_2^2, \tag{13}$$

where $\boldsymbol{\lambda}^{\boldsymbol{f}}$ is the Lagrange multiplier vector, $\beta^{\boldsymbol{f}}$ is a positive scalar. The optimization problem of L_A in (13) with respect to each variable can be solved by the following sub-problems:

1) $\boldsymbol{f}$ sub-problem: the sub-problem of L_A with respect to $\boldsymbol{f}$ can be rewritten as

$$\min_{\boldsymbol{f}} \frac{\lambda}{1+\rho}\|\boldsymbol{f}\|_1 + \frac{\beta^{\boldsymbol{f}}}{2}\|\boldsymbol{f} - (\mathbf{D}\mathtt{vec}(\boldsymbol{\mathcal{S}}) + \frac{\boldsymbol{\lambda}^{\boldsymbol{f}}}{\beta^{\boldsymbol{f}}})\|_2^2,$$

where the well-known soft-thresholding operator (Donoho, 1995) can be applied to solve this sub-problem as follows

$$\boldsymbol{f} = \text{soft}(\mathbf{D}\texttt{vec}(\boldsymbol{\mathcal{S}}) + \frac{\boldsymbol{\lambda}^{\boldsymbol{f}}}{\beta^{\boldsymbol{f}}}, \frac{\lambda}{(1+\rho)\beta^{\boldsymbol{f}}}), \tag{14}$$

where the soft-thresholding operator $\text{soft}(\boldsymbol{\mathcal{A}}, \tau) = \text{sign}(\boldsymbol{\mathcal{A}}) \cdot \max(|\boldsymbol{\mathcal{A}}| - \tau, 0)$ is performed element-wisely.

2) $\boldsymbol{\mathcal{S}}$ sub-problem: the sub-problem of L_A with respect to $\boldsymbol{\mathcal{S}}$ can be solved by the following linear system

$$(2\mathbf{I} + \beta^{\boldsymbol{f}}\mathbf{D}^*\mathbf{D})\texttt{vec}(\boldsymbol{\mathcal{S}}) = 2\texttt{vec}(\boldsymbol{\mathcal{E}}_{k+1}) + \mathbf{D}^*(\beta^{\boldsymbol{f}}\boldsymbol{f} - \boldsymbol{\lambda}^{\boldsymbol{f}}),$$

where $\mathbf{D}^*$ indicates the adjoint of $\mathbf{D}$. Let $\boldsymbol{\mathcal{C}} = \texttt{ten}\big(2\texttt{vec}(\boldsymbol{\mathcal{E}}_{k+1}) + \mathbf{D}^*(\beta^{\boldsymbol{f}}\boldsymbol{f} - \boldsymbol{\lambda}^{\boldsymbol{f}})\big)$. Thanks to the block-circulant structure of the matrix corresponding to the operator $\mathbf{D}^*\mathbf{D}$, it can be diagonalized by the 3D FFT matrix. Therefore, $\boldsymbol{\mathcal{S}}$ can be computed efficiently by

$$\texttt{ifftn}\Big(\frac{\texttt{fftn}(\boldsymbol{\mathcal{C}})}{2\cdot\mathbf{1} + \beta^{\boldsymbol{f}}(|\texttt{fftn}(\mathbf{D}_h)|^2 + |\texttt{fftn}(\mathbf{D}_v)|^2 + |\texttt{fftn}(\mathbf{D}_t)|^2)}\Big), \tag{15}$$

where $\texttt{fftn}(\cdot)$ and $\texttt{ifftn}(\cdot)$ indicate the fast 3D Fourier transform and its inverse transform, respectively. Note that the denominator in the equation can be pre-calculated outside the main loop, avoiding the extra computational cost.

3) Updating Multipliers: According to the ADMM, the multipliers associated with L_A are updated by the following formulas

$$\boldsymbol{\lambda}^{\boldsymbol{f}} \leftarrow \boldsymbol{\lambda}^{\boldsymbol{f}} - \gamma\beta^{\boldsymbol{f}}\big(\boldsymbol{f} - \mathbf{D}\texttt{vec}(\boldsymbol{\mathcal{S}})\big) \tag{16}$$

where $\gamma > 0$ is a parameter associated with convergence rate, and the penalty parameter $\beta^{\boldsymbol{f}}$ follow an adaptive updating scheme as suggested in (Chan et al., 2011). Take $\beta^{\boldsymbol{f}}$ as an example,

$$\beta^{\boldsymbol{f}} = \begin{cases} c_1\beta^{\boldsymbol{f}}, \text{ if } \text{Err}(\boldsymbol{f}^{\text{iter}+1}) \geq c_2\text{Err}(\boldsymbol{f}^{\text{iter}}) \\ \quad \beta^{\boldsymbol{f}}, \text{ otherwise.} \end{cases} \tag{17}$$

where $\text{Err}(\boldsymbol{f}^{\text{iter}}) = \|\boldsymbol{f}^{\text{iter}} - \mathbf{D}\texttt{vec}(\boldsymbol{\mathcal{S}})\|_2$. As suggested in (Cao et al., 2016), $\gamma = 1.1$, and c_1, c_2 can be taken as 1.15 and 0.95, respectively.

Repeating the iteration process (14), (15), and (17) sufficiently many times, the so-obtained $\boldsymbol{\mathcal{S}}$ is taken as an iterative solution to $\boldsymbol{\mathcal{S}}^{k+1}$. Note that among these many rounds of iteratively repeating (14), (15), and (17), the first round requires initial guesses of $\boldsymbol{\mathcal{S}}$ and $\boldsymbol{\lambda}^{\boldsymbol{f}}$, for which we set $\boldsymbol{\mathcal{S}}^k$ as the initial guess of $\boldsymbol{\mathcal{S}}$ and $\mathbf{0}$ (zero vector) as the initial guess of $\boldsymbol{\lambda}^{\boldsymbol{f}}$.

3.2.3 Optimization over $\boldsymbol{\mathcal{X}}$

In this case, we fix all $(\mathbf{U}_{k+1}, \boldsymbol{\mathcal{S}}^{k+1})$. Here, $\boldsymbol{\mathcal{X}}^{k+1}$ is updated through

$$\begin{aligned}\boldsymbol{\mathcal{X}}^{k+1} \in \arg\min_{\boldsymbol{\mathcal{X}}}\Big\{&\|\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{S}}^{k+1} - (\boldsymbol{\mathcal{X}}^k - \boldsymbol{\mathcal{S}}^{k+1}) \times_1 \mathbf{U}_{k+1}^{(1)}\mathbf{U}_{k+1}^{(1)\top} \times_2 \mathbf{U}_{k+1}^{(2)}\mathbf{U}_{k+1}^{(2)\top} \times_3 \mathbf{U}_{k+1}^{(3)}\mathbf{U}_{k+1}^{(3)\top}\|_F^2 \\ &+ \rho\|\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{X}}^k\|_F^2 : \boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{X}}) = \boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{F}})\Big\},\end{aligned} \tag{18}$$

Algorithm 1 ADMM algorithm for solving (12)

```
Input: $(\mathbf{U}_{k+1}, \boldsymbol{\mathcal{X}}^k)$, $\rho = 0.001$; The algorithm parameter: $\lambda$.
Initialization: $\boldsymbol{\mathcal{S}} = \boldsymbol{\mathcal{S}}^k$; $\beta^{\boldsymbol{f}}$ is initialized by $\frac{1e^{+1}}{\text{mean}(\boldsymbol{\mathcal{X}})}$; other variables are initialized by $\mathbf{0}$.
while $\frac{\|\boldsymbol{\mathcal{S}}_{\text{iter}} - \boldsymbol{\mathcal{S}}_{\text{iter}-1}\|_F}{\max\{1, \|\boldsymbol{\mathcal{S}}_{\text{iter}-1}\|_F\}} > 10^{-6}$ & iter $\leq 100$ do
    Updating $\boldsymbol{f}$ via (14);
    Updating $\boldsymbol{\mathcal{S}}$ via (15);
    Updating multipliers and the related parameters via (16) and (17);
    iter = iter + 1;
end while
Output: $\boldsymbol{\mathcal{S}}^{k+1} = \boldsymbol{\mathcal{S}}$
```

Now let $\boldsymbol{\mathcal{F}}_{k+1} \coloneqq \boldsymbol{\mathcal{S}}^{k+1} + (\boldsymbol{\mathcal{X}}^k - \boldsymbol{\mathcal{S}}^{k+1}) \times_1 \mathbf{U}_{k+1}^{(1)} \mathbf{U}_{k+1}^{(1)\top} \times_2 \mathbf{U}_{k+1}^{(2)} \mathbf{U}_{k+1}^{(2)\top} \times_3 \mathbf{U}_{k+1}^{(3)} \mathbf{U}_{k+1}^{(3)\top}$. This problem has a closed-form solution as follows

$$\boldsymbol{\mathcal{X}}^{k+1} = \boldsymbol{\mathcal{P}}_{\bar{\Omega}}\Big(\frac{\boldsymbol{\mathcal{F}}_{k+1} + \rho \boldsymbol{\mathcal{X}}^k}{1+\rho}\Big) + \boldsymbol{\mathcal{P}}_{\Omega}(\boldsymbol{\mathcal{F}}), \tag{19}$$

where $\bar{\Omega}$ is the complement set of Ω. The detailed implementation of tenPAM algorithm is summarized in Algorithm 2.

Algorithm 2 tenPAM Algorithm

```
Input: Partial observed tensor $\boldsymbol{\mathcal{F}}_{\Omega}$, $\lambda > 0$
Initialization: $r_1 = \text{ceil}(0.8 \times H)$, $r_2 = \text{ceil}(0.8 \times W)$, $r_3 = 1$; $\gamma = 1.1$; $c_1 = 1.15, c_2 = 0.95$; $\boldsymbol{\mathcal{X}}^0 = \boldsymbol{\mathcal{P}}(\boldsymbol{\mathcal{F}})$; $\{\mathbf{U}_0^{(n)}\}_{n \in [3]}$ is initialized by $(r_1, r_2, r_3)$-Tucker decomposition of $\boldsymbol{\mathcal{X}}^0$; $\boldsymbol{\mathcal{S}} = \boldsymbol{\mathcal{X}}^0 - \boldsymbol{\mathcal{L}}^0$; $\rho = 0.001$; Other variables are initialized as $\mathbf{0}$.
while $\frac{\|\boldsymbol{\mathcal{P}}_{\Omega}(\boldsymbol{\mathcal{X}}^k - \boldsymbol{\mathcal{F}})\|_F}{\max\{1, \|\boldsymbol{\mathcal{F}}\|_F\}} > 10^{-6}$ & $k \leq 50$ do
    for $n = 1$ to 3 do
        Get $\mathbf{U}_{k+1}^{(n)}$ by solving problem (7)
    end for
    Get $\boldsymbol{\mathcal{S}}^{k+1}$ by solving problem (11) using Algorithm 1
    Get $\boldsymbol{\mathcal{X}}^{k+1}$ by solving problem (18)
    $k = k + 1$
end while
Output: $\{\mathbf{U}_k \mathbf{U}_k^{\top}, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k\}_{k \geq 0}$
```

3.3 Implementation Details

In the SRTC model (3), there exist four parameters, namely, r_1, r_2, r_3, and λ, respectively, where r_1 and r_2 control the complexity of spatial redundancy, r_3 controls the complexity of temporal redundancy, and λ handles a trade-off between noise and foreground modeling. In all experiments, r_1 and r_2 are set to values of $\text{ceil}(0.80 \times H)$ and $\text{ceil}(0.80 \times W)$, respectively, where $\text{ceil}(\cdot)$ is the operator to round the element to the nearest integer greater than or equal to that element. By doing so, the accumulation energy ratio of top normalized singular values (AccEgyR) attains a ratio over 0.9 for various natural images, as reported in (Cao et al., 2016). For r_3, it takes the value 1 for all experiments so that each image frame in $\boldsymbol{\mathcal{L}}$ is the same (Sobral et al., 2016). In terms of λ, it needs to be carefully tuned based on the data. Specifically, λ is taken in the range $[0.2, 1]$.

4. Convergence Analysis of tenPAM Algorithm

In this section, we will prove the global convergence of tenPAM algorithm 2. For notational convenience in our analysis, set $I_1 = H$, $I_2 = W$, $I_3 = T$, and $N = 3$. We first note that the proposed formulation (4) can be reformulated as an equivalent unconstrained optimization problem as follows

$$\min_{\mathbf{U}\mathbf{U}^\top, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{X}}} G(\mathbf{U}\mathbf{U}^\top, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{X}}) = \hat{F}(\mathbf{U}\mathbf{U}^\top, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{X}}) + \delta_S(\boldsymbol{\mathcal{X}}) + \sum_{n\in[3]} \delta_{S_n}(\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}), \tag{20}$$

where the set $S_n := \{\mathbf{X} \in \mathbb{R}^{I_n \times I_n} : \operatorname{rank}(\mathbf{X}) = r_n, \mathbf{X} = \mathbf{X}^\top, \text{eigenvalue of } \mathbf{X} \text{ is either 1 or 0}\}$, and $S := \{\boldsymbol{\mathcal{X}} \in \mathbb{R}^{H\times W\times T} : \boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{X}}) = \boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{F}})\}$. For a given set A, its characteristic function is defined as

$$\delta_A(\boldsymbol{x}) = \begin{cases} 0 & \boldsymbol{x} \in A \\ +\infty & \text{otherwise} \end{cases},$$

which is a proper and lower semicontinuous (PLSC) function. Therefore, $G(\cdot)$ is a PLSC function. $\partial G(\mathbf{U}\mathbf{U}^\top, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{X}})$ is called the Subdifferential of G at $\{\mathbf{U}\mathbf{U}^\top, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{X}}\}$, which has the following definition.

Definition 6 (*Subdifferentials* (Attouch and Bolte, 2009; Attouch et al., 2010)) *Assume that $f : \mathbb{R}^d \to (-\infty, +\infty)$ is a proper and lower semicontinuous function.*

1. *The domain of f is defined and denoted by* $\operatorname{dom} f := \{\boldsymbol{x} \in \mathbb{R}^n : f(\boldsymbol{x}) < +\infty\}$

2. *For a given $\boldsymbol{x} \in \operatorname{dom} f$, the Fréchet subdifferential of f at $\boldsymbol{x}$, written $\hat{\partial} f(\boldsymbol{x})$, is the set of all vectors $\boldsymbol{u} \in \mathbb{R}^d$ that satisfy*

$$\lim_{\boldsymbol{y}\neq\boldsymbol{x}} \inf_{\boldsymbol{y}\to\boldsymbol{x}} \frac{f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle \boldsymbol{u}, \boldsymbol{y} - \boldsymbol{x}\rangle}{\|\boldsymbol{y} - \boldsymbol{x}\|} \geq 0.$$

3. *The limiting-subdifferential, or simply the subdifferential, of f at $\boldsymbol{x}$, written $\partial f(\boldsymbol{x})$ is defined through the following closure process*

$$\partial f(\boldsymbol{x}) := \{\boldsymbol{u} \in \mathbb{R}^d : \exists \boldsymbol{x}^k \to \boldsymbol{x}, f(\boldsymbol{x}^k) \to f(\boldsymbol{x}) \text{ and } \boldsymbol{u}^k \in \hat{\partial} f(\boldsymbol{x}^k) \to \boldsymbol{u} \text{ as } k \to \infty\}.$$

Before introducing our key results, the following proposition is needed to build our main results.

Proposition 7 *The following optimization problem* (21) *has a closed-form solution*

$$\boldsymbol{\mathcal{S}} \in \arg\min_{\boldsymbol{\mathcal{Y}}} \|\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{S}} - (\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{Y}}) \times_1 \mathbf{U}^{(1)}\mathbf{U}^{(1)\top} \times_2 \cdots \times_N \mathbf{U}^{(N)}\mathbf{U}^{(N)\top}\|_F^2. \tag{21}$$

Proof The objective function in (21) is a convex function of $\boldsymbol{\mathcal{Y}}$. Therefore, based on the first-order optimality condition, we have

$$\begin{aligned}\mathbf{0} =&2(\boldsymbol{\mathcal{Y}}-\boldsymbol{\mathcal{X}})\times_1 \mathbf{U}^{(1)}\mathbf{U}^{(1)\top}\times_2\cdots\times_N \mathbf{U}^{(N)}\mathbf{U}^{(N)\top}\\ &-2(\boldsymbol{\mathcal{S}}-\boldsymbol{\mathcal{X}})\times_1 \mathbf{U}^{(1)}\mathbf{U}^{(1)\top}\times_2\cdots\times_N \mathbf{U}^{(N)}\mathbf{U}^{(N)\top}\\ =&2(\boldsymbol{\mathcal{Y}}-\boldsymbol{\mathcal{S}})\times_1 \mathbf{U}^{(1)}\mathbf{U}^{(1)\top}\times_2\cdots\times_N \mathbf{U}^{(N)}\mathbf{U}^{(N)\top}.\end{aligned}$$

Thus, (21) is valid. ■

Now, our first main lemma about the sufficient decrease property of the iterative sequence $\{\mathbf{U}_k\mathbf{U}_k^\top,\boldsymbol{\mathcal{S}}^k,\boldsymbol{\mathcal{X}}^k\}_{k\geq 0}$ from Algorithm 2 is ready to be introduced.

Lemma 8 (Sufficient decrease property) *Given that $0<\rho<\infty$, $\{\mathbf{U}_k\mathbf{U}_k^\top,\boldsymbol{\mathcal{S}}^k,\boldsymbol{\mathcal{X}}^k\}_{k\geq 0}$ is the sequence generated from the proposed Algorithm 2, then the sequence satisfies*

$$\begin{aligned}&\rho(\|\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top-\mathbf{U}_k\mathbf{U}_k^\top\|_F^2+\|\boldsymbol{\mathcal{S}}^{k+1}-\boldsymbol{\mathcal{S}}^k\|_F^2+\|\boldsymbol{\mathcal{X}}^{k+1}-\boldsymbol{\mathcal{X}}^k\|_F^2)\\ \leq& G(\mathbf{U}_k\mathbf{U}_k^\top,\boldsymbol{\mathcal{S}}^k,\boldsymbol{\mathcal{X}}^k)-G(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top,\boldsymbol{\mathcal{S}}^{k+1},\boldsymbol{\mathcal{X}}^{k+1}).\end{aligned}\tag{22}$$

Proof At $(k+1)$-th iteration, we can get the optimal solution for (7) for all $n=1,2,3$. Thus,

$$\begin{aligned}&G(\{\mathbf{U}_{k+1}^{(i)}\mathbf{U}_{k+1}^{(i)\top}\}_{i\leq n},\{\mathbf{U}_{k}^{(i)}\mathbf{U}_{k}^{(i)\top}\}_{i>n},\boldsymbol{\mathcal{S}}^k,\boldsymbol{\mathcal{X}}^k)\\ \leq& G(\{\mathbf{U}_{k+1}^{(i)}\mathbf{U}_{k+1}^{(i)\top}\}_{i<n},\{\mathbf{U}_{k}^{(i)}\mathbf{U}_{k}^{(i)\top}\}_{i\geq n},\boldsymbol{\mathcal{S}}^k,\boldsymbol{\mathcal{X}}^k)-\rho\|\mathbf{U}_{k}^{(n)}\mathbf{U}_{k}^{(n)\top}-\mathbf{U}_{k+1}^{(n)}\mathbf{U}_{k+1}^{(n)\top}\|_F^2.\end{aligned}\tag{23}$$

The above inequality (23) implies the following inequality

$$\begin{aligned}&G(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top,\boldsymbol{\mathcal{S}}^k,\boldsymbol{\mathcal{X}}^k)-G(\mathbf{U}_k\mathbf{U}_k^\top,\boldsymbol{\mathcal{S}}^k,\boldsymbol{\mathcal{X}}^k)\\ =&\sum_{n=1}^{3}\Big(G(\{\mathbf{U}_{k+1}^{(i)}\mathbf{U}_{k+1}^{(i)\top}\}_{i\leq n},\{\mathbf{U}_{k}^{(i)}\mathbf{U}_{k}^{(i)\top}\}_{i>n},\boldsymbol{\mathcal{S}}^k,\boldsymbol{\mathcal{X}}^k)\\ &-G(\{\mathbf{U}_{k+1}^{(i)}\mathbf{U}_{k+1}^{(i)\top}\}_{i<n},\{\mathbf{U}_{k}^{(i)}\mathbf{U}_{k}^{(i)\top}\}_{i\geq n},\boldsymbol{\mathcal{S}}^k,\boldsymbol{\mathcal{X}}^k)\Big)\\ \leq&-\sum_{n=1}^{3}\rho\|\mathbf{U}_{k}^{(n)}\mathbf{U}_{k}^{(n)\top}-\mathbf{U}_{k+1}^{(n)}\mathbf{U}_{k+1}^{(n)\top}\|_F^2.\end{aligned}\tag{24}$$

The fact that $\boldsymbol{\mathcal{S}}^{k+1}$ is the optimal solution for problem (11) shows

$$\begin{aligned}&\|\boldsymbol{\mathcal{X}}^k-\boldsymbol{\mathcal{S}}^{k+1}-(\boldsymbol{\mathcal{X}}^k-\boldsymbol{\mathcal{S}}^k)\times_1\mathbf{U}_{k+1}^{(1)}\mathbf{U}_{k+1}^{(1)\top}\times_2\mathbf{U}_{k+1}^{(2)}\mathbf{U}_{k+1}^{(2)\top}\times_3\mathbf{U}_{k+1}^{(3)}\mathbf{U}_{k+1}^{(3)\top}\|_F^2\\ &+\lambda\|\boldsymbol{\mathcal{S}}^{k+1}\|_{TV1}+\rho\|\boldsymbol{\mathcal{S}}^{k+1}-\boldsymbol{\mathcal{S}}^k\|_F^2\leq G(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top,\boldsymbol{\mathcal{S}}^k,\boldsymbol{\mathcal{X}}^k).\end{aligned}\tag{25}$$

Based on Proposition 7, the following holds

$$\begin{aligned}&G(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top,\boldsymbol{\mathcal{S}}^{k+1},\boldsymbol{\mathcal{X}}^k)\\ =&\min_{\boldsymbol{\mathcal{S}}}\Big(\|\boldsymbol{\mathcal{X}}^k-\boldsymbol{\mathcal{S}}^{k+1}-(\boldsymbol{\mathcal{X}}^k-\boldsymbol{\mathcal{S}})\times_1\mathbf{U}_{k+1}^{(1)}\mathbf{U}_{k+1}^{(1)\top}\times_2\mathbf{U}_{k+1}^{(2)}\mathbf{U}_{k+1}^{(2)\top}\times_3\mathbf{U}_{k+1}^{(3)}\mathbf{U}_{k+1}^{(3)\top}\|_F^2\\ &+\lambda\|\boldsymbol{\mathcal{S}}^{k+1}\|_{TV1}\Big).\end{aligned}\tag{26}$$

Combining (25) and (26), we have

$$G(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top, \boldsymbol{\mathcal{S}}^{k+1}, \boldsymbol{\mathcal{X}}^k) \leq G(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top, \boldsymbol{\mathcal{S}}^{k}, \boldsymbol{\mathcal{X}}^k) - \rho\|\boldsymbol{\mathcal{S}}^{k+1} - \boldsymbol{\mathcal{S}}^k\|_F^2. \tag{27}$$

Since $\boldsymbol{\mathcal{X}}^{k+1}$ is the optimal solution for (18), the following holds

$$\begin{aligned}&\|\boldsymbol{\mathcal{X}}^{k+1} - \boldsymbol{\mathcal{S}}^{k+1} - (\boldsymbol{\mathcal{X}}^k - \boldsymbol{\mathcal{S}}^{k+1}) \times_1 \mathbf{U}_{k+1}^{(1)}\mathbf{U}_{k+1}^{(1)\top} \times_2 \mathbf{U}_{k+1}^{(2)}\mathbf{U}_{k+1}^{(2)\top} \times_3 \mathbf{U}_{k+1}^{(3)}\mathbf{U}_{k+1}^{(3)\top}\|_F^2 \\ &+ \lambda\|\boldsymbol{\mathcal{S}}^{k+1}\|_{TV1} + \rho\|\boldsymbol{\mathcal{X}}^{k+1} - \boldsymbol{\mathcal{X}}^k\|_F^2 \leq G(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top, \boldsymbol{\mathcal{S}}^{k+1}, \boldsymbol{\mathcal{X}}^k).\end{aligned} \tag{28}$$

Again, the following can be obtained through Proposition 7

$$\begin{aligned}&G(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top, \boldsymbol{\mathcal{S}}^{k+1}, \boldsymbol{\mathcal{X}}^{k+1}) \\ =&\min_{\boldsymbol{\mathcal{X}}} \Big(\|\boldsymbol{\mathcal{X}}^{k+1} - \boldsymbol{\mathcal{S}}^{k+1} - (\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{S}}^{k+1}) \times_1 \mathbf{U}_{k+1}^{(1)}\mathbf{U}_{k+1}^{(1)\top} \times_2 \mathbf{U}_{k+1}^{(2)}\mathbf{U}_{k+1}^{(2)\top} \times_3 \mathbf{U}_{k+1}^{(3)}\mathbf{U}_{k+1}^{(3)\top}\|_F^2 \\ &+ \lambda\|\boldsymbol{\mathcal{S}}^{k+1}\|_{TV1}\Big).\end{aligned} \tag{29}$$

Combining (28) and (29), we have

$$G(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top, \boldsymbol{\mathcal{S}}^{k+1}, \boldsymbol{\mathcal{X}}^{k+1}) \leq G(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top, \boldsymbol{\mathcal{S}}^{k+1}, \boldsymbol{\mathcal{X}}^k) - \rho\|\boldsymbol{\mathcal{X}}^{k+1} - \boldsymbol{\mathcal{X}}^k\|_F^2. \tag{30}$$

Adding (24), (27), and (30) together, we obtain our result

$$\begin{aligned}&\rho(\|\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top - \mathbf{U}_k\mathbf{U}_k^\top\|_F^2 + \|\boldsymbol{\mathcal{S}}^{k+1} - \boldsymbol{\mathcal{S}}^k\|_F^2 + \|\boldsymbol{\mathcal{X}}^{k+1} - \boldsymbol{\mathcal{X}}^k\|_F^2) \\ \leq& G(\mathbf{U}_k\mathbf{U}_k^\top, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k) - G(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top, \boldsymbol{\mathcal{S}}^{k+1}, \boldsymbol{\mathcal{X}}^{k+1}),\end{aligned}$$

which is the sufficient decrease property. ■

Our second main lemma is about the subgradient lower bound. To build the lemma, the following two propositions are needed: (1) Proposition 9 shows the iterative sequence $\{\mathbf{U}_k\mathbf{U}_k^\top, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k\}_{k\geq 0}$ from our Algorithm 2 is bounded; (2) Proposition 10 shows the Lipschitz continuity of the gradient of $\hat{F}(\cdot)$.

Proposition 9 *The sequences* $\{\boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k\}_{k\geq 0}$ *are bounded, where the bounds are determined by the initial values of* $(\mathbf{U}_0\mathbf{U}_0^\top, \boldsymbol{\mathcal{S}}^0, \boldsymbol{\mathcal{X}}^0)$*. Specifically, it follows*

$$\begin{aligned}\|\boldsymbol{\mathcal{S}}^k\|_F &\leq \|\boldsymbol{\mathcal{S}}^0\|_F + G(\mathbf{U}_0\mathbf{U}_0^\top, \boldsymbol{\mathcal{S}}^0, \boldsymbol{\mathcal{X}}^0)/\rho \\ \|\boldsymbol{\mathcal{X}}^k\|_F &\leq \|\boldsymbol{\mathcal{X}}^0\|_F + G(\mathbf{U}_0\mathbf{U}_0^\top, \boldsymbol{\mathcal{S}}^0, \boldsymbol{\mathcal{X}}^0)/\rho.\end{aligned} \tag{31}$$

Proof To start with,

$$\begin{aligned}\|\boldsymbol{\mathcal{S}}^k\|_F &= \|\sum_{i=1}^k(\boldsymbol{\mathcal{S}}^i - \boldsymbol{\mathcal{S}}^{i-1}) + \boldsymbol{\mathcal{S}}^0\|_F \\ &\leq \sum_{i=1}^k \|\boldsymbol{\mathcal{S}}^i - \boldsymbol{\mathcal{S}}^{i-1}\| + \|\boldsymbol{\mathcal{S}}^0\|_F & \text{(32a)} \\ &\leq \sum_{i=1}^k \Big(G(\mathbf{U}_{i-1}\mathbf{U}_{i-1}^\top, \boldsymbol{\mathcal{S}}^{i-1}, \boldsymbol{\mathcal{X}}^{i-1}) - G(\mathbf{U}_i\mathbf{U}_i^\top, \boldsymbol{\mathcal{S}}^i, \boldsymbol{\mathcal{X}}^i)\Big)/\rho + \|\boldsymbol{\mathcal{S}}^0\|_F & \text{(32b)}\end{aligned}$$

$$
\begin{aligned}
&= \Big(G(\mathbf{U}_0\mathbf{U}_0^\top, \boldsymbol{\mathcal{S}}^0, \boldsymbol{\mathcal{X}}^0) - G(\mathbf{U}_k\mathbf{U}_k^\top, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k)\Big)/\rho + \|\boldsymbol{\mathcal{S}}^0\|_F \\
&\leq G(\mathbf{U}_0\mathbf{U}_0^\top, \boldsymbol{\mathcal{S}}^0, \boldsymbol{\mathcal{X}}^0)/\rho + \|\boldsymbol{\mathcal{S}}^0\|_F, \qquad (32c)
\end{aligned}
$$

where inequality (32a) comes from the triangle inequality, inequality (32b) comes from (22) in Lemma 8, and the last inequality (32c) is due to the fact that $G(\mathbf{U}_k\mathbf{U}_k^\top, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k) \geq 0$. The same proof can also be applied to $\boldsymbol{\mathcal{X}}^k$. In practice, we can set $\boldsymbol{\mathcal{S}}^0 = 0$, $\boldsymbol{\mathcal{X}}^0 = \boldsymbol{\mathcal{P}}(\boldsymbol{\mathcal{F}})$, $\mathbf{U}_0\mathbf{U}_0^\top$ is the HOSVD of $\boldsymbol{\mathcal{X}}^0$. By doing so, $G(\mathbf{U}_0\mathbf{U}_0^\top, \boldsymbol{\mathcal{S}}^0, \boldsymbol{\mathcal{X}}^0)$ is bounded by the input data. That is, $G(\mathbf{U}_0\mathbf{U}_0^\top, \boldsymbol{\mathcal{S}}^0, \boldsymbol{\mathcal{X}}^0) \leq \|\boldsymbol{\mathcal{X}}^0\|_F^2$. ■

Proposition 10 *For bounded* $\|\boldsymbol{\mathcal{S}}\|_F^2$ *and* $\|\boldsymbol{\mathcal{X}}\|_F^2$*, there exists a constant* $L \coloneqq (\|\boldsymbol{\mathcal{X}}\|_F^2 + \|\boldsymbol{\mathcal{S}}\|_F^2)\frac{2^N(\prod_{i=1}^N r_i)}{\sqrt{N-1}}$ *such that*

$$
\| - \nabla_{\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}} \hat{F}(\hat{\mathbf{U}}\hat{\mathbf{U}}^\top, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{X}}) + \nabla_{\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}} \hat{F}(\tilde{\mathbf{U}}\tilde{\mathbf{U}}^\top, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{X}})\|_F \leq L\|\hat{\mathbf{U}}\hat{\mathbf{U}}^\top - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^\top\|_F \qquad (33)
$$

for any pair of $\hat{\mathbf{U}}\hat{\mathbf{U}}^\top$ *and* $\tilde{\mathbf{U}}\tilde{\mathbf{U}}^\top$*,* $\forall\ n \in [N]$*.*

Proof Define $\|\cdot\|$ as the 2-operator norm.

$$
\begin{aligned}
&\| - \nabla_{\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}} \hat{F}(\hat{\mathbf{U}}\hat{\mathbf{U}}^\top, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{X}}) + \nabla_{\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}} \hat{F}(\tilde{\mathbf{U}}\tilde{\mathbf{U}}^\top, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{X}})\|_F \\
=&\|(\mathbf{X}_{(n)} - \mathbf{S}_{(n)}) \cdot \hat{\mathbf{U}}_{\mathbf{\Psi}^{(n)}} \cdot \hat{\mathbf{U}}_{\mathbf{\Psi}^{(n)}}^\top \cdot (\mathbf{X}_{(n)} - \mathbf{S}_{(n)})^\top \\
&- (\mathbf{X}_{(n)} - \mathbf{S}_{(n)}) \cdot \tilde{\mathbf{U}}_{\mathbf{\Psi}^{(n)}} \cdot \tilde{\mathbf{U}}_{\mathbf{\Psi}^{(n)}}^\top \cdot (\mathbf{X}_{(n)} - \mathbf{S}_{(n)})^\top\|_F \\
=&\|(\mathbf{X}_{(n)} - \mathbf{S}_{(n)}) \cdot (\hat{\mathbf{U}}_{\mathbf{\Psi}^{(n)}} \cdot \hat{\mathbf{U}}_{\mathbf{\Psi}^{(n)}}^\top - \tilde{\mathbf{U}}_{\mathbf{\Psi}^{(n)}} \cdot \tilde{\mathbf{U}}_{\mathbf{\Psi}^{(n)}}^\top) \cdot (\mathbf{X}_{(n)} - \mathbf{S}_{(n)})^\top\|_F \\
\leq&\|(\mathbf{X}_{(n)} - \mathbf{S}_{(n)})(\mathbf{X}_{(n)} - \mathbf{S}_{(n)})^\top\| \cdot \|\hat{\mathbf{U}}_{\mathbf{\Psi}^{(n)}} \cdot \hat{\mathbf{U}}_{\mathbf{\Psi}^{(n)}}^\top - \tilde{\mathbf{U}}_{\mathbf{\Psi}^{(n)}} \cdot \tilde{\mathbf{U}}_{\mathbf{\Psi}^{(n)}}^\top\|_F \qquad (34a) \\
\leq&L\|\hat{\mathbf{U}}\hat{\mathbf{U}}^\top - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^\top\|_F, \qquad (34b)
\end{aligned}
$$

where inequality (34a) is due to the Frobenius norm and operator norm inequality, inequality (34b) is based on the definitions of $\hat{\mathbf{U}}_{\mathbf{\Psi}^{(n)}}$ and $\tilde{\mathbf{U}}_{\mathbf{\Psi}^{(n)}}$, and $L \coloneqq (\|\boldsymbol{\mathcal{X}}\|_F^2 + \|\boldsymbol{\mathcal{S}}\|_F^2)\frac{2^N(\prod_{i=1}^N r_i)}{\sqrt{N-1}}$.

Next, we will show how we get the value for L.

$$
\begin{aligned}
&\|\hat{\mathbf{U}}_{\mathbf{\Psi}^{(n)}} \cdot \hat{\mathbf{U}}_{\mathbf{\Psi}^{(n)}}^\top - \tilde{\mathbf{U}}_{\mathbf{\Psi}^{(n)}} \cdot \tilde{\mathbf{U}}_{\mathbf{\Psi}^{(n)}}^\top\|_F \\
=&\prod_{i\neq n} \|\hat{\mathbf{U}}^{(i)}\hat{\mathbf{U}}^{(i)\top} - \tilde{\mathbf{U}}^{(i)}\tilde{\mathbf{U}}^{(i)\top}\|_F \qquad (35a) \\
=&\frac{1}{N-1}\sum_{j\neq n}\prod_{i\neq n} \|\hat{\mathbf{U}}^{(i)}\hat{\mathbf{U}}^{(i)\top} - \tilde{\mathbf{U}}^{(i)}\tilde{\mathbf{U}}^{(i)\top}\|_F \\
=&\frac{2^{N-1}}{N-1}\sum_{j\neq n}(\prod_{i\neq n,j} r_i)\|\hat{\mathbf{U}}^{(j)}\hat{\mathbf{U}}^{(j)\top} - \tilde{\mathbf{U}}^{(j)}\tilde{\mathbf{U}}^{(j)\top}\|_F \\
\leq&\frac{2^{N-1}(\prod_{i=1}^N r_i)}{N-1}\sum_{j\neq n} \|\hat{\mathbf{U}}^{(j)}\hat{\mathbf{U}}^{(j)\top} - \tilde{\mathbf{U}}^{(j)}\tilde{\mathbf{U}}^{(j)\top}\|_F \qquad (35b)
\end{aligned}
$$

$$\leq \frac{2^{N-1}(\prod_{i=1}^{N} r_i)}{N-1}\sqrt{N-1}\sqrt{\sum_{j\neq n}\|\hat{\mathbf{U}}^{(j)}\hat{\mathbf{U}}^{(j)\top}-\tilde{\mathbf{U}}^{(j)}\tilde{\mathbf{U}}^{(j)\top}\|_F^2} \tag{35c}$$

$$\leq \frac{2^{N-1}(\prod_{i=1}^{N} r_i)}{\sqrt{N-1}}\|\hat{\mathbf{U}}\hat{\mathbf{U}}^\top-\tilde{\mathbf{U}}\tilde{\mathbf{U}}^\top\|_F, \tag{35d}$$

where equality (35a) is based on the definition of $\hat{\mathbf{U}}_{\mathbf{\Psi}^{(n)}}$, inequality (35b) holds since $r_i \geq 1$, inequality (35c) is from the Cauchy–Schwarz inequality, and inequality (35d) comes from the fact that $j \neq n$. There is another inequality that needs to be proved. That is,

$$\begin{aligned}
&\|(\mathbf{X}_{(n)}-\mathbf{S}_{(n)})(\mathbf{X}_{(n)}-\mathbf{S}_{(n)})^\top\|\\
\leq&\|(\mathbf{X}_{(n)}-\mathbf{S}_{(n)})(\mathbf{X}_{(n)}-\mathbf{S}_{(n)})^\top\|_F \qquad (36a)\\
=&\|\mathbf{X}_{(n)}-\mathbf{S}_{(n)}\|_F^2\\
\leq&2(\|\boldsymbol{\mathcal{X}}\|_F^2+\|\boldsymbol{\mathcal{S}}\|_F^2), \qquad (36b)
\end{aligned}$$

where inequality (36a) is based on the definition of the 2-operator norm and inequality (36b) is due to the Cauchy–Schwarz inequality. In addition, $\|\boldsymbol{\mathcal{X}}\|_F^2+\|\boldsymbol{\mathcal{S}}\|_F^2$ is bounded in our algorithm based on (31). Based on the above two inequalities (35) and (36), $L := (\|\boldsymbol{\mathcal{X}}\|_F^2 + \|\boldsymbol{\mathcal{S}}\|_F^2)\frac{2^N(\prod_{i=1}^N r_i)}{\sqrt{N-1}}$ can be derived. ■

Now our second main lemma is ready to be presented.

Lemma 11 (Subgradient lower bound) *Suppose that* $0<\rho<\infty$*,* $\{\mathbf{U}_k\mathbf{U}_k^\top, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k\}_{k\geq 0}$ *is the sequence generated from the proposed Algorithm 2, then the sequence satisfies*

$$\|\omega_{k+1}\|_F \leq \rho_1\sqrt{\|\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top-\mathbf{U}_k\mathbf{U}_k^\top\|_F^2+\|\boldsymbol{\mathcal{X}}^{k+1}-\boldsymbol{\mathcal{X}}^k\|_F^2+\|\boldsymbol{\mathcal{S}}^{k+1}-\boldsymbol{\mathcal{S}}^k\|_F^2},$$

where $\rho_1 = \max\{2\sqrt{3}\rho+3L, 4+2\rho+\kappa, 2+2\rho+\kappa\}$ *with* $L = \frac{2^N(\prod_{i=1}^N r_i)}{\sqrt{N-1}}\Big(2G(\mathbf{U}_0\mathbf{U}_0^\top, \boldsymbol{\mathcal{S}}^0, \boldsymbol{\mathcal{X}}^0)/\rho + \|\boldsymbol{\mathcal{S}}^0\|_F+\|\boldsymbol{\mathcal{X}}^0\|_F\Big)^2$ *and* $\kappa = 4G(\mathbf{U}_0\mathbf{U}_0^\top, \boldsymbol{\mathcal{S}}^0, \boldsymbol{\mathcal{X}}^0)/\rho + 2\|\boldsymbol{\mathcal{S}}^0\|_F + 2\|\boldsymbol{\mathcal{X}}^0\|_F$.

Proof According to the first-order optimality condition for each sub-problem (5) in k-th iteration of the proposed algorithm, we have

$$\begin{aligned}
\mathbf{0} \in&\nabla_{\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}}\hat{F}(\{\mathbf{U}_{k+1}^{(i)}\mathbf{U}_{k+1}^{(i)\top}\}_{i\leq n}, \{\mathbf{U}_k^{(i)}\mathbf{U}_k^{(i)\top}\}_{i>n}, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k)\\
&+2\rho(\mathbf{U}_{k+1}^{(n)}\mathbf{U}_{k+1}^{(n)\top}-\mathbf{U}_k^{(n)}\mathbf{U}_k^{(n)\top})+\partial_{\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}}\delta_{S_n}(\mathbf{U}_{k+1}^{(n)}\mathbf{U}_{k+1}^{(n)\top}),
\end{aligned}$$

which can be rewritten as

$$\begin{aligned}
&-2\rho(\mathbf{U}_{k+1}^{(n)}\mathbf{U}_{k+1}^{(n)\top}-\mathbf{U}_k^{(n)}\mathbf{U}_k^{(n)\top})-\nabla_{\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}}\hat{F}(\{\mathbf{U}_{k+1}^{(i)}\mathbf{U}_{k+1}^{(i)\top}\}_{i\leq n}, \{\mathbf{U}_k^{(i)}\mathbf{U}_k^{(i)\top}\}_{i>n}, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k)\\
&+\nabla_{\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}}\hat{F}(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k)-\nabla_{\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}}\hat{F}(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k)\\
&+\nabla_{\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}}\hat{F}(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top, \boldsymbol{\mathcal{S}}^{k+1}, \boldsymbol{\mathcal{X}}^k)-\nabla_{\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}}\hat{F}(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top, \boldsymbol{\mathcal{S}}^{k+1}, \boldsymbol{\mathcal{X}}^k)\\
&+\nabla_{\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}}\hat{F}(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top, \boldsymbol{\mathcal{S}}^{k+1}, \boldsymbol{\mathcal{X}}^{k+1})\\
&\in\nabla_{\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}}\hat{F}(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top, \boldsymbol{\mathcal{S}}^{k+1}, \boldsymbol{\mathcal{X}}^{k+1})+\partial_{\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}}\delta_{S_n}(\mathbf{U}_{k+1}^{(n)}\mathbf{U}_{k+1}^{(n)\top})\\
&:=\partial_{\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}}G(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top, \boldsymbol{\mathcal{S}}^{k+1}, \boldsymbol{\mathcal{X}}^{k+1}).
\end{aligned} \tag{37}$$

Based on the Definition 6, we have the following proposition.

Proposition 12 (*Subdifferentiability property* (Bolte et al., 2014)) *Given that $\Psi(\boldsymbol{x},\boldsymbol{y}) = H(\boldsymbol{x},\boldsymbol{y}) + f(\boldsymbol{x}) + g(\boldsymbol{y})$, if H is continuously differentiable, then for all $(\boldsymbol{x},\boldsymbol{y})$ we have*

$$\partial\Psi(\boldsymbol{x},\boldsymbol{y}) = (\nabla_{\boldsymbol{x}} H(\boldsymbol{x},\boldsymbol{y}) + \partial f(\boldsymbol{x}), \nabla_{\boldsymbol{y}} H(\boldsymbol{x},\boldsymbol{y}) + \partial g(\boldsymbol{y})).$$

Accordingly, based on the first-order optimality condition (well-known Fermat's rule) of (11),

$$\begin{aligned}\mathbf{0} \in\, & 2\Big(\boldsymbol{\mathcal{S}}^{k+1} - \boldsymbol{\mathcal{X}}^{k} + (\boldsymbol{\mathcal{X}}^{k} - \boldsymbol{\mathcal{S}}^{k}) \times_1 \mathbf{U}^{(1)}_{k+1}\mathbf{U}^{(1)\top}_{k+1} \times_2 \mathbf{U}^{(2)}_{k+1}\mathbf{U}^{(2)\top}_{k+1} \times_3 \mathbf{U}^{(3)}_{k+1}\mathbf{U}^{(3)\top}_{k+1}\Big) \\ & + \lambda\partial\|\boldsymbol{\mathcal{S}}^{k+1}\|_{TV1} + 2\rho(\boldsymbol{\mathcal{S}}^{k+1} - \boldsymbol{\mathcal{S}}^{k}).\end{aligned} \tag{38}$$

In addition,

$$\begin{aligned}& \partial_{\boldsymbol{\mathcal{S}}} G(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^{\top}, \boldsymbol{\mathcal{S}}^{k+1}, \boldsymbol{\mathcal{X}}^{k+1}) \\ =& 2(\boldsymbol{\mathcal{S}}^{k+1} - \boldsymbol{\mathcal{X}}^{k+1}) + 2(\boldsymbol{\mathcal{X}}^{k+1} - \boldsymbol{\mathcal{S}}^{k+1}) \times_1 \mathbf{U}^{(1)}_{k+1}\mathbf{U}^{(1)\top}_{k+1} \times_2 \mathbf{U}^{(2)}_{k+1}\mathbf{U}^{(2)\top}_{k+1} \times_3 \mathbf{U}^{(3)}_{k+1}\mathbf{U}^{(3)\top}_{k+1} \\ & + \lambda\partial\|\boldsymbol{\mathcal{S}}^{k+1}\|_{TV1}.\end{aligned} \tag{39}$$

Combining with (38) and (39),

$$\begin{aligned}& -2(\boldsymbol{\mathcal{X}}^{k+1} - \boldsymbol{\mathcal{X}}^{k}) + 2(\boldsymbol{\mathcal{X}}^{k+1} - \boldsymbol{\mathcal{X}}^{k}) \times_1 \mathbf{U}^{(1)}_{k+1}\mathbf{U}^{(1)\top}_{k+1} \times_2 \mathbf{U}^{(2)}_{k+1}\mathbf{U}^{(2)\top}_{k+1} \times_3 \mathbf{U}^{(3)}_{k+1}\mathbf{U}^{(3)\top}_{k+1} \\ & + 2(\boldsymbol{\mathcal{S}}^{k} - \boldsymbol{\mathcal{S}}^{k+1}) \times_1 \mathbf{U}^{(1)}_{k+1}\mathbf{U}^{(1)\top}_{k+1} \times_2 \mathbf{U}^{(2)}_{k+1}\mathbf{U}^{(2)\top}_{k+1} \times_3 \mathbf{U}^{(3)}_{k+1}\mathbf{U}^{(3)\top}_{k+1} + 2\rho(\boldsymbol{\mathcal{S}}^{k} - \boldsymbol{\mathcal{S}}^{k+1}) \\ & \in \partial_{\boldsymbol{\mathcal{S}}} G(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^{\top}, \boldsymbol{\mathcal{S}}^{k+1}, \boldsymbol{\mathcal{X}}^{k+1}).\end{aligned} \tag{40}$$

Accordingly, based on the first order optimality condition of (18),

$$\begin{aligned}\mathbf{0} \in\, & 2(\boldsymbol{\mathcal{X}}^{k+1} - \boldsymbol{\mathcal{S}}^{k+1}) - 2(\boldsymbol{\mathcal{X}}^{k} - \boldsymbol{\mathcal{S}}^{k+1}) \times_1 \mathbf{U}^{(1)}_{k+1}\mathbf{U}^{(1)\top}_{k+1} \times_2 \mathbf{U}^{(2)}_{k+1}\mathbf{U}^{(2)\top}_{k+1} \times_3 \mathbf{U}^{(3)}_{k+1}\mathbf{U}^{(3)\top}_{k+1} \\ & + 2\rho(\boldsymbol{\mathcal{X}}^{k+1} - \boldsymbol{\mathcal{X}}^{k}) + \partial\delta_S(\boldsymbol{\mathcal{X}}^{k+1}).\end{aligned} \tag{41}$$

In addition,

$$\begin{aligned}& \partial_{\boldsymbol{\mathcal{X}}} G(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^{\top}, \boldsymbol{\mathcal{S}}^{k+1}, \boldsymbol{\mathcal{X}}^{k+1}) \\ =& 2(\boldsymbol{\mathcal{X}}^{k+1} - \boldsymbol{\mathcal{S}}^{k+1}) - 2(\boldsymbol{\mathcal{X}}^{k+1} - \boldsymbol{\mathcal{S}}^{k+1}) \times_1 \mathbf{U}^{(1)}_{k+1}\mathbf{U}^{(1)\top}_{k+1} \times_2 \mathbf{U}^{(2)}_{k+1}\mathbf{U}^{(2)\top}_{k+1} \times_3 \mathbf{U}^{(3)}_{k+1}\mathbf{U}^{(3)\top}_{k+1} \\ & + \partial\delta_S(\boldsymbol{\mathcal{X}}^{k+1}).\end{aligned} \tag{42}$$

Combining with (41) and (42), we have

$$\begin{aligned}& -2(\boldsymbol{\mathcal{X}}^{k+1} - \boldsymbol{\mathcal{X}}^{k}) \times_1 \mathbf{U}^{(1)}_{k+1}\mathbf{U}^{(1)\top}_{k+1} \times_2 \mathbf{U}^{(2)}_{k+1}\mathbf{U}^{(2)\top}_{k+1} \times_3 \mathbf{U}^{(3)}_{k+1}\mathbf{U}^{(3)\top}_{k+1} + 2\rho(\boldsymbol{\mathcal{X}}^{k} - \boldsymbol{\mathcal{X}}^{k+1}) \\ \in& \partial_{\boldsymbol{\mathcal{X}}} G(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^{\top}, \boldsymbol{\mathcal{S}}^{k+1}, \boldsymbol{\mathcal{X}}^{k+1}).\end{aligned} \tag{43}$$

Thus

$$\begin{aligned}& \|\nabla_{\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}} \hat{F}(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^{\top}, \boldsymbol{\mathcal{S}}^{k}, \boldsymbol{\mathcal{X}}^{k}) - \nabla_{\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}} \hat{F}(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^{\top}, \boldsymbol{\mathcal{S}}^{k+1}, \boldsymbol{\mathcal{X}}^{k})\|_F \\ =& \|(\mathbf{X}^{k}_{(n)} - \mathbf{S}^{k}_{(n)}) \cdot \mathbf{U}_{\boldsymbol{\Psi}^{(n)}_{k+1}} \cdot \mathbf{U}_{\boldsymbol{\Psi}^{(n)}_{k+1}}{}^{\top} \cdot (\mathbf{X}^{k}_{(n)} - \mathbf{S}^{k}_{(n)})^{\top} \\ & - (\mathbf{X}^{k}_{(n)} - \mathbf{S}^{k+1}_{(n)}) \cdot \mathbf{U}_{\boldsymbol{\Psi}^{(n)}_{k+1}} \cdot \mathbf{U}_{\boldsymbol{\Psi}^{(n)}_{k+1}}{}^{\top} \cdot (\mathbf{X}^{k}_{(n)} - \mathbf{S}^{k+1}_{(n)})^{\top}\|_F\end{aligned}$$

$$\begin{aligned}
\leq &\|(\mathbf{X}^k_{(n)} - \mathbf{S}^k_{(n)}) \cdot \mathbf{U}_{\boldsymbol{\Psi}^{(n)}_{k+1}} \cdot \mathbf{U}_{\boldsymbol{\Psi}^{(n)}_{k+1}}{}^{\top} \cdot (\mathbf{X}^k_{(n)} - \mathbf{S}^k_{(n)})^{\top} \\
&- (\mathbf{X}^k_{(n)} - \mathbf{S}^k_{(n)}) \cdot \mathbf{U}_{\boldsymbol{\Psi}^{(n)}_{k+1}} \cdot \mathbf{U}_{\boldsymbol{\Psi}^{(n)}_{k+1}}{}^{\top} \cdot (\mathbf{X}^k_{(n)} - \mathbf{S}^{k+1}_{(n)})^{\top}\|_F \\
&+ \|(\mathbf{X}^k_{(n)} - \mathbf{S}^k_{(n)}) \cdot \mathbf{U}_{\boldsymbol{\Psi}^{(n)}_{k+1}} \cdot \mathbf{U}_{\boldsymbol{\Psi}^{(n)}_{k+1}}{}^{\top} \cdot (\mathbf{X}^k_{(n)} - \mathbf{S}^{k+1}_{(n)})^{\top} \\
&- (\mathbf{X}^k_{(n)} - \mathbf{S}^{k+1}_{(n)}) \cdot \mathbf{U}_{\boldsymbol{\Psi}^{(n)}_{k+1}} \cdot \mathbf{U}_{\boldsymbol{\Psi}^{(n)}_{k+1}}{}^{\top} \cdot (\mathbf{X}^k_{(n)} - \mathbf{S}^{k+1}_{(n)})^{\top}\|_F
\end{aligned} \tag{44a}$$

$$\begin{aligned}
= &\|(\mathbf{X}^k_{(n)} - \mathbf{S}^k_{(n)}) \cdot \mathbf{U}_{\boldsymbol{\Psi}^{(n)}_{k+1}} \cdot \mathbf{U}_{\boldsymbol{\Psi}^{(n)}_{k+1}}{}^{\top} \cdot (\mathbf{S}^{k+1}_{(n)} - \mathbf{S}^k_{(n)})^{\top}\|_F \\
&+ \|(\mathbf{X}^k_{(n)} - \mathbf{S}^{k+1}_{(n)}) \cdot \mathbf{U}_{\boldsymbol{\Psi}^{(n)}_{k+1}} \cdot \mathbf{U}_{\boldsymbol{\Psi}^{(n)}_{k+1}}{}^{\top} \cdot (\mathbf{S}^{k+1}_{(n)} - \mathbf{S}^k_{(n)})^{\top}\|_F
\end{aligned}$$

$$\leq \Big(\|(\mathbf{X}^k_{(n)} - \mathbf{S}^k_{(n)}) \cdot \mathbf{U}_{\boldsymbol{\Psi}^{(n)}_{k+1}} \cdot \mathbf{U}_{\boldsymbol{\Psi}^{(n)}_{k+1}}{}^{\top}\| + \|(\mathbf{X}^k_{(n)} - \mathbf{S}^{k+1}_{(n)}) \cdot \mathbf{U}_{\boldsymbol{\Psi}^{(n)}_{k+1}} \cdot \mathbf{U}_{\boldsymbol{\Psi}^{(n)}_{k+1}}{}^{\top}\|\Big)\|\boldsymbol{\mathcal{S}}^{k+1} - \boldsymbol{\mathcal{S}}^k\|_F \tag{44b}$$

$$\leq \Big(\|(\mathbf{X}^k_{(n)} - \mathbf{S}^k_{(n)}) \cdot \mathbf{U}_{\boldsymbol{\Psi}^{(n)}_{k+1}} \cdot \mathbf{U}_{\boldsymbol{\Psi}^{(n)}_{k+1}}{}^{\top}\|_F + \|(\mathbf{X}^k_{(n)} - \mathbf{S}^{k+1}_{(n)}) \cdot \mathbf{U}_{\boldsymbol{\Psi}^{(n)}_{k+1}} \cdot \mathbf{U}_{\boldsymbol{\Psi}^{(n)}_{k+1}}{}^{\top}\|_F\Big)\|\boldsymbol{\mathcal{S}}^{k+1} - \boldsymbol{\mathcal{S}}^k\|_F \tag{44c}$$

$$\leq \Big(\|\mathbf{X}^k_{(n)} - \mathbf{S}^k_{(n)}\|_F + \|\mathbf{X}^k_{(n)} - \mathbf{S}^{k+1}_{(n)}\|_F\Big)\|\boldsymbol{\mathcal{S}}^{k+1} - \boldsymbol{\mathcal{S}}^k\|_F \tag{44d}$$

$$\leq (2\|\boldsymbol{\mathcal{X}}^k\|_F + \|\boldsymbol{\mathcal{S}}^k\|_F + \|\boldsymbol{\mathcal{S}}^{k+1}\|_F)\|\boldsymbol{\mathcal{S}}^{k+1} - \boldsymbol{\mathcal{S}}^k\|_F \tag{44e}$$

$$\leq \kappa\|\boldsymbol{\mathcal{S}}^{k+1} - \boldsymbol{\mathcal{S}}^k\|_F, \tag{44f}$$

where the inequalities (44a) and (44e) are because of the triangle inequality, inequality (44b) comes from the Frobenius norm and the operator norm inequality, the inequality (44c) is based on the definition of the 2-operator norm, and the inequality (44d) is because $\mathbf{U}_{\boldsymbol{\Psi}^{(n)}_{k+1}}$ is a semi-orthogonal matrix. In (44f), $\kappa = 4G(\mathbf{U}_0\mathbf{U}_0^{\top}, \boldsymbol{\mathcal{S}}^0, \boldsymbol{\mathcal{X}}^0)/\rho + 2\|\boldsymbol{\mathcal{S}}^0\|_F + 2\|\boldsymbol{\mathcal{X}}^0\|_F \geq 2\|\boldsymbol{\mathcal{X}}^k\|_F + \|\boldsymbol{\mathcal{S}}^k\|_F + \|\boldsymbol{\mathcal{S}}^{k+1}\|_F$ is dependent on the upper bound for $\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{S}}$.

In the end, combining (37), (40), and (43), we have

$$\mathrm{d}(\mathbf{0}, \partial G(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^{\top}, \boldsymbol{\mathcal{S}}^{k+1}, \boldsymbol{\mathcal{X}}^{k+1})$$

$$\begin{aligned}
\leq &\sum_{n=1}^{3} \| - 2\rho(\mathbf{U}^{(n)}_{k+1}\mathbf{U}^{(n)\top}_{k+1} - \mathbf{U}^{(n)}_{k}\mathbf{U}^{(n)\top}_{k}) \\
&- \nabla_{\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}} \hat{F}(\{\mathbf{U}^{(i)}_{k+1}\mathbf{U}^{(i)\top}_{k+1}\}_{i\leq n}, \{\mathbf{U}^{(i)}_{k}\mathbf{U}^{(i)\top}_{k}\}_{i>n}, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k) \\
&+ \nabla_{\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}} \hat{F}(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^{\top}, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k) - \nabla_{\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}} \hat{F}(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^{\top}, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k) \\
&+ \nabla_{\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}} \hat{F}(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^{\top}, \boldsymbol{\mathcal{S}}^{k+1}, \boldsymbol{\mathcal{X}}^k) - \nabla_{\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}} \hat{F}(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^{\top}, \boldsymbol{\mathcal{S}}^{k+1}, \boldsymbol{\mathcal{X}}^k) \\
&+ \nabla_{\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}} \hat{F}(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^{\top}, \boldsymbol{\mathcal{S}}^{k+1}, \boldsymbol{\mathcal{X}}^{k+1})\|_F \\
&+ \| - 2(\boldsymbol{\mathcal{X}}^{k+1} - \boldsymbol{\mathcal{X}}^k) + 2(\boldsymbol{\mathcal{X}}^{k+1} - \boldsymbol{\mathcal{X}}^k) \times_1 \mathbf{U}^{(1)}_{k+1}\mathbf{U}^{(1)\top}_{k+1} \times_2 \mathbf{U}^{(2)}_{k+1}\mathbf{U}^{(2)\top}_{k+1} \times_3 \mathbf{U}^{(3)}_{k+1}\mathbf{U}^{(3)\top}_{k+1} \\
&+ 2(\boldsymbol{\mathcal{S}}^k - \boldsymbol{\mathcal{S}}^{k+1}) \times_1 \mathbf{U}^{(1)}_{k+1}\mathbf{U}^{(1)\top}_{k+1} \times_2 \mathbf{U}^{(2)}_{k+1}\mathbf{U}^{(2)\top}_{k+1} \times_3 \mathbf{U}^{(3)}_{k+1}\mathbf{U}^{(3)\top}_{k+1} + 2\rho(\boldsymbol{\mathcal{S}}^k - \boldsymbol{\mathcal{S}}^{k+1})\|_F \\
&+ \| - 2(\boldsymbol{\mathcal{X}}^{k+1} - \boldsymbol{\mathcal{X}}^k) \times_1 \mathbf{U}^{(1)}_{k+1}\mathbf{U}^{(1)\top}_{k+1} \times_2 \mathbf{U}^{(2)}_{k+1}\mathbf{U}^{(2)\top}_{k+1} \times_3 \mathbf{U}^{(3)}_{k+1}\mathbf{U}^{(3)\top}_{k+1} + 2\rho(\boldsymbol{\mathcal{X}}^k - \boldsymbol{\mathcal{X}}^{k+1})\|_F
\end{aligned} \tag{45a}$$

$$\begin{aligned}
\leq &\sum_{n=1}^{3} (2\rho\|\mathbf{U}^{(n)}_{k+1}\mathbf{U}^{(n)\top}_{k+1} - \mathbf{U}^{(n)}_{k}\mathbf{U}^{(n)\top}_{k}\|_F + L\|\mathbf{U}_{k+1}\mathbf{U}_{k+1}^{\top} - \mathbf{U}_k\mathbf{U}_k^{\top}\|_F) \\
&+ \|\nabla_{\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}} \hat{F}(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^{\top}, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k) - \nabla_{\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}} \hat{F}(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^{\top}, \boldsymbol{\mathcal{S}}^{k+1}, \boldsymbol{\mathcal{X}}^k)\|_F
\end{aligned} \tag{45b}$$

$$
\begin{aligned}
&+\|\nabla_{\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}}\hat{F}(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top,\boldsymbol{\mathcal{S}}^{k+1},\boldsymbol{\mathcal{X}}^{k})-\nabla_{\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}}\hat{F}(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top,\boldsymbol{\mathcal{S}}^{k+1},\boldsymbol{\mathcal{X}}^{k+1})\|_F\\
&+(4+2\rho)\|\boldsymbol{\mathcal{X}}^{k+1}-\boldsymbol{\mathcal{X}}^{k}\|_F+(2+2\rho)\|\boldsymbol{\mathcal{S}}^{k+1}-\boldsymbol{\mathcal{S}}^{k}\|_F\\
\leq&(2\sqrt{3}\rho+3L)\|\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top-\mathbf{U}_{k}\mathbf{U}_{k}^\top\|_F \qquad\qquad (45c)\\
&+(4+2\rho+\kappa)\|\boldsymbol{\mathcal{X}}^{k+1}-\boldsymbol{\mathcal{X}}^{k}\|_F+(2+2\rho+\kappa)\|\boldsymbol{\mathcal{S}}^{k+1}-\boldsymbol{\mathcal{S}}^{k}\|_F,
\end{aligned}
$$

where inequality (45a) and inequality (45b) are due to the triangle inequality, inequality (45c) is from (33) in Proposition 10 with $L=\frac{2^N(\prod_{i=1}^N r_i)}{\sqrt{N-1}}\Big(2G(\mathbf{U}_0\mathbf{U}_0^\top,\boldsymbol{\mathcal{S}}^0,\boldsymbol{\mathcal{X}}^0)/\rho+\|\boldsymbol{\mathcal{S}}^0\|_F+\|\boldsymbol{\mathcal{X}}^0\|_F\Big)^2$ and (44).

Therefore, the following holds

$$
\begin{aligned}
&\mathrm{d}(\mathbf{0},\partial G(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top,\boldsymbol{\mathcal{S}}^{k+1},\boldsymbol{\mathcal{X}}^{k+1}))\\
\leq&(2\sqrt{3}\rho+3L)\|\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top-\mathbf{U}_{k}\mathbf{U}_{k}^\top\|_F\\
&+(4+2\rho+\kappa)\|\boldsymbol{\mathcal{X}}^{k+1}-\boldsymbol{\mathcal{X}}^{k}\|_F+(2+2\rho+\kappa)\|\boldsymbol{\mathcal{S}}^{k+1}-\boldsymbol{\mathcal{S}}^{k}\|_F\\
\leq&\sqrt{3}\rho_1\sqrt{\|\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top-\mathbf{U}_{k}\mathbf{U}_{k}^\top\|_F^2+\|\boldsymbol{\mathcal{X}}^{k+1}-\boldsymbol{\mathcal{X}}^{k}\|_F^2+\|\boldsymbol{\mathcal{S}}^{k+1}-\boldsymbol{\mathcal{S}}^{k}\|_F^2},
\end{aligned}
\tag{46}
$$

where the first inequality is the same thing as inequality (45) and the second inequality is because of the Cauchy–Schwarz inequality with $\rho_1=\max\{2\sqrt{3}\rho+3L,4+2\rho+\kappa,2+2\rho+\kappa\}$. Thus, there exist $\omega_{k+1}\in\partial G(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top,\boldsymbol{\mathcal{S}}^{k+1},\boldsymbol{\mathcal{X}}^{k+1})$ such that

$$
\|\omega_{k+1}\|_F\leq\rho_1\sqrt{\|\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top-\mathbf{U}_{k}\mathbf{U}_{k}^\top\|_F^2+\|\boldsymbol{\mathcal{X}}^{k+1}-\boldsymbol{\mathcal{X}}^{k}\|_F^2+\|\boldsymbol{\mathcal{S}}^{k+1}-\boldsymbol{\mathcal{S}}^{k}\|_F^2},
$$

which is the property of subgradient lower bound. ■

Definition 13 (*Critical point* (Attouch and Bolte, 2009; Attouch et al., 2010)) *A necessary condition for $\boldsymbol{x}$ to be a minimizer of a proper and lower semicontinuous (PLSC) function f is that*

$$
\mathbf{0}\in\partial f(\boldsymbol{x}).\tag{47}
$$

A point that satisfies (47) *is called limiting-critical or simply critical.*

Based on Lemmas 8 and 11, the following theorem summarizes the theoretical property of the iterative sequence $\{\mathbf{U}_k\mathbf{U}_k^\top,\boldsymbol{\mathcal{S}}^k,\boldsymbol{\mathcal{X}}^k\}_{k\geq0}$ from our Algorithm 2.

Theorem 14 *Let $\{\mathbf{U}_k\mathbf{U}_k^\top,\boldsymbol{\mathcal{S}}^k,\boldsymbol{\mathcal{X}}^k\}_{k\geq0}$ denote the sequence generated from Algorithm 2 with $w(\mathbf{U}_0\mathbf{U}_0^\top,\boldsymbol{\mathcal{S}}^0,\boldsymbol{\mathcal{X}}^0)$ denoting the set of all its limit points, and let the set critG $=\{(\mathbf{U}\mathbf{U}^\top,\boldsymbol{\mathcal{S}},\boldsymbol{\mathcal{X}}):(\mathbf{U}\mathbf{U}^\top,\boldsymbol{\mathcal{S}},\boldsymbol{\mathcal{X}})$ is a critical point of* (20)$\}$. *Then*

(i) The sequence $\{G(\mathbf{U}_k\mathbf{U}_k^\top,\boldsymbol{\mathcal{S}}^k,\boldsymbol{\mathcal{X}}^k)\}_{k\geq0}$ is nonincreasing;

(ii) $\sum_{k\geq0}\|(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top,\boldsymbol{\mathcal{S}}^{k+1},\boldsymbol{\mathcal{X}}^{k+1})-(\mathbf{U}_k\mathbf{U}_k^\top,\boldsymbol{\mathcal{S}}^k,\boldsymbol{\mathcal{X}}^k)\|_F^2<+\infty$;

(iii) $w(\mathbf{U}_0\mathbf{U}_0^\top,\boldsymbol{\mathcal{S}}^0,\boldsymbol{\mathcal{X}}^0)\subseteq$ critG;

(iv) $w(\mathbf{U}_0\mathbf{U}_0^\top, \boldsymbol{\mathcal{S}}^0, \boldsymbol{\mathcal{X}}^0)$ *is a nonempty compact connected set, and*

$$\begin{aligned}&\mathrm{d}\Big((\mathbf{U}_k\mathbf{U}_k^\top, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k), w(\mathbf{U}_0\mathbf{U}_0^\top, \boldsymbol{\mathcal{S}}^0, \boldsymbol{\mathcal{X}}^0)\Big)\\ := &\inf_{(\mathbf{U}\mathbf{U}^\top, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{X}})\in w(\mathbf{U}_0\mathbf{U}_0^\top, \boldsymbol{\mathcal{S}}^0, \boldsymbol{\mathcal{X}}^0)} \|(\mathbf{U}\mathbf{U}^\top, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{X}}) - (\mathbf{U}_k\mathbf{U}_k^\top, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k)\|_F \to 0, \text{as } k \to +\infty;\end{aligned}$$

(v) $G(\cdot)$ *is finite and constant on* $w(\mathbf{U}_0\mathbf{U}_0^\top, \boldsymbol{\mathcal{S}}^0, \boldsymbol{\mathcal{X}}^0)$.

Proof The proof is split into five parts.

(i) It comes from Lemma 8;

(ii) Let $f_k := G(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top, \boldsymbol{\mathcal{S}}^{k+1}, \boldsymbol{\mathcal{X}}^{k+1})$. According to the *Monotone convergence theorem*, $\lim\limits_{k\to\infty} f_k$ exist since f_k is bounded from below, namely, $f_k \geq 0$. Let $\varepsilon_k := f_{k-1} - f_k$

$$f_k = f_0 - \sum_{i=1}^{k} \varepsilon_i.$$

Based on (22),

$$\rho\|(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top, \boldsymbol{\mathcal{S}}^{k+1}, \boldsymbol{\mathcal{X}}^{k+1}) - (\mathbf{U}_k\mathbf{U}_k^\top, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k)\|_F^2 \leq \varepsilon_{k+1},$$

which implies that

$$\sum_{k\geq 0} \|(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top, \boldsymbol{\mathcal{S}}^{k+1}, \boldsymbol{\mathcal{X}}^{k+1}) - (\mathbf{U}_k\mathbf{U}_k^\top, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k)\|_F^2 \leq (\sum_{k\geq 0} \varepsilon_{k+1})/\rho < +\infty$$

$$\lim_{k\to\infty} \|(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top, \boldsymbol{\mathcal{S}}^{k+1}, \boldsymbol{\mathcal{X}}^{k+1}) - (\mathbf{U}_k\mathbf{U}_k^\top, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k)\|_F = 0;$$

(iii) Based on (46) in Lemma 11,

$$\lim_{k\to\infty} \mathrm{d}(\mathbf{0}, \partial G(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top, \boldsymbol{\mathcal{S}}^{k+1}, \boldsymbol{\mathcal{X}}^{k+1}) = 0;$$

(iv) This can be derived from part *(ii)*;

(v) This result is based on part *(i)* and the function $G(\cdot)$ is nonnegative. ■

However, Theorem 14 cannot guarantee the global convergence of the iterative sequence $\{\mathbf{U}_k\mathbf{U}_k^\top, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k\}_{k\geq 0}$, which has the following definition.

Definition 15 (*Global convergence* (Petrovai, 2017; Xu, 2018)) *Any iterative algorithm for solving an optimization problem over a set X, is said to be **globally convergent** if for any starting point $\boldsymbol{x}_0 \in X$, the sequence generated by the algorithm always has an accumulation critical point.*

To build the global convergence of our iterative sequence $\{\mathbf{U}_k\mathbf{U}_k^\top, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k\}_{k\geq 0}$ based on Theorem 14, the function $G(\cdot)$ needs to have the KŁ property as follows

Definition 16 (*KŁ property* (Attouch et al., 2010, 2013; Bolte et al., 2014; Xu, 2018)) *A real function $f : \mathbb{R}^d \to (-\infty, +\infty]$ has the Kurdyka Łojasiewicz (KŁ) property if for any point $\bar{\boldsymbol{u}} \in \mathbb{R}^d$, in a neighborhood $N(\bar{\boldsymbol{u}}, \sigma)$, there exists a desingularizing function $\phi(s) = cs^{1-\theta}$ for some $c > 0$ and $\theta \in [0, 1)$ such that*

$$\phi'(|f(\boldsymbol{u}) - f(\bar{\boldsymbol{u}})|)\mathrm{d}(0, \partial f(\boldsymbol{u})) \geq 1 \text{ for any } \boldsymbol{u} \in N(\bar{\boldsymbol{u}}, \sigma) \text{ and } f(\boldsymbol{u}) \neq f(\bar{\boldsymbol{u}}).$$

The semi-algebraic set and semi-algebraic function, which are related to the Kurdyka Łojasiewicz (KŁ) property, are introduced below.

Definition 17 (*Semi-algebraic* (Attouch and Bolte, 2009; Bolte et al., 2014)) *A subset S of $\mathbb{R}^d$ is a real* ***semi-algebraic set*** *if there exists a finite number of real polynomial functions g_{ij}, h_{ij}: $\mathbb{R}^d \to \mathbb{R}$ such that*

$$S = \cup_{j=1}^{q} \cap_{i=1}^{p} \{\boldsymbol{u} \in \mathbb{R}^d : g_{ij}(\boldsymbol{u}) = 0 \text{ and } h_{ij}(\boldsymbol{u}) < 0\}.$$

In addition, a function $h : \mathbb{R}^{d+1} \to \mathbb{R} \cup +\infty$ is called ***semi-algebraic*** *if its graph*

$$\{(\boldsymbol{u}, t) \in \mathbb{R}^{d+1} : h(\boldsymbol{u}) = t\}$$

is a real semi-algebraic set.

After introducing these two definitions, the following lemma shows that the objective function $G(\mathbf{U}\mathbf{U}^\top, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{X}})$ of (20) has the so-called KŁ property.

Lemma 18 *The function $G(\cdot)$ has the KŁ property.*

Proof We will first prove that the sets $\{S_n\}_{n\in[3]}$ are semi-algebraic sets. In (Bolte et al., 2014; Lewis and Malick, 2008), the authors showed that the set of all matrices with the same rank is semi-algebraic. Therefore, for each $n \in [N]$,

$$T_n \coloneqq \{\mathbf{X} \in \mathbb{R}^{I_n \times I_n} : \operatorname{rank}(\mathbf{X}) = r_n\}$$

is semi-algebraic. In addition, we observe that

$$\begin{aligned} K_n &:= \{\mathbf{X} \in \mathbb{R}^{I_n \times I_n} : \text{eigenvalue of } \mathbf{X} \text{ is either 1 or } 0, \mathbf{X} = \mathbf{X}^\top\} \\ &= \{\mathbf{X} \in \mathbb{R}^{I_n \times I_n} : \mathbf{X}\mathbf{X}^\top = \mathbf{X}, \mathbf{X}^\top\mathbf{X} = \mathbf{X}\}, \end{aligned}$$

where all the equalities are quadratic. Clearly, the set K_n is semi-algebraic. Since $S_n = T_n \cap K_n$ and intersection of two semi-algebraic sets is still semi-algebraic (Bolte et al., 2014), S_n is a semi-algebraic set. Recall that

$$G(\mathbf{U}\mathbf{U}^\top, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{X}}) = \hat{F}(\mathbf{U}\mathbf{U}^\top, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{X}}) + \delta_S(\boldsymbol{\mathcal{X}}) + \sum_{n\in[3]} \delta_{S_n}(\mathbf{U}^{(n)}\mathbf{U}^{(n)\top}).$$

$\hat{F}(\mathbf{U}\mathbf{U}^\top, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{X}})$ is a function of summation of polynomial functions of all the elements in $\{\mathbf{U}\mathbf{U}^\top, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{X}}\}$. In addition, the characteristic function $\delta_A(\cdot)$ is a semi-algebraic function if the set A is semi-algebraic (Attouch et al., 2010; Bolte et al., 2014). Any finite sum of semi-algebraic functions is semi-algebraic; thus, the function $G(\mathbf{U}\mathbf{U}^\top, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{X}})$ is semi-algebraic.

A semi-algebraic real-valued function is a KŁ function based on the work of (Bolte et al., 2007, 2014). ■

Based on Lemmas 8, 11, and 18 and conclusions in (Bolte et al., 2014, Section 3.2), the following main theorem can be obtained.

Theorem 19 (Global Convergence) $\{\mathbf{U}_k\mathbf{U}_k^\top, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k\}_{k\geq 0}$ *is the sequence generated from the proposed Algorithm 2 with any initial point so that* $\boldsymbol{\mathcal{S}}^0, \boldsymbol{\mathcal{X}}^0$*, and* $G(\mathbf{U}_0\mathbf{U}_0^\top, \boldsymbol{\mathcal{S}}^0, \boldsymbol{\mathcal{X}}^0)$ *are bounded. Then, there exists* $(\mathbf{U}_*\mathbf{U}_*^\top, \boldsymbol{\mathcal{S}}^*, \boldsymbol{\mathcal{X}}^*)$ *such that*

(i) $(\mathbf{U}_k\mathbf{U}_k^\top, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k) \to (\mathbf{U}_*\mathbf{U}_*^\top, \boldsymbol{\mathcal{S}}^*, \boldsymbol{\mathcal{X}}^*)$*;*

(ii) $\mathbf{0} \in \partial G(\mathbf{U}_*\mathbf{U}_*^\top, \boldsymbol{\mathcal{S}}^*, \boldsymbol{\mathcal{X}}^*)$*;*

(iii) $\{\mathbf{U}_k\mathbf{U}_k^\top, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k\}_{k\geq 0}$ *has a finite length, namely,*

$$\sum_{k=0}^{+\infty} \|(\mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top, \boldsymbol{\mathcal{S}}^{k+1}, \boldsymbol{\mathcal{X}}^{k+1}) - (\mathbf{U}_k\mathbf{U}_k^\top, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k)\|_F < +\infty.$$

5. Numerical Studies

To evaluate the performance of the proposed SRTC, its performance on open-sourced video data is presented in this section. In Section 5.1, the empirical convergence of the proposed algorithm is illustrated to verify our theoretical results. The performances of the proposed algorithm for background reconstruction and foreground detection are presented in Sections 5.2 and 5.3, respectively. In Sections 5.2 and 5.3, MCOS[1] (Li et al., 2021), BFMNM[2] (Shang et al., 2017), HQ-ASD[3] (He et al., 2019), RTRC[4] (Huang et al., 2021), and HQ-TCASD[5] (He and Atia, 2020) are selected as benchmarks for comparison with the proposed SRTC, which are the state-of-the-art methods in the related area. The benchmarks have two categories:

1. MCOS, BFMNM, and HQ-ASD are the most advanced Robust Matrix Completion algorithms in the literature;
2. RTRC and HQ-TCASD are the state-of-the-art Robust Tensor Completion algorithms.

All results in this section are the average results of 20 repetitions for comparison. The codes of SRTC are implemented in Matlab 2021a. The CPU of the computer to conduct experiments in this paper is an Intel® Xeon® Processor E-2286M (8-cores 2.40-GHz Turbo, 16 MB).

Performance evaluation indices and parameter tuning: For the task of background reconstruction, the peak signal-to-noise ratio (PSNR) and the structural similarity

[1] `https://github.com/ZihengLi6321/MCOS`

[2] `https://1drv.ms/u/s!Aur7I-sQwed-gjE_15tuofYp2n7x?e=xN57mw`

[3] `https://github.com/he1c/robust-matrix-completion`

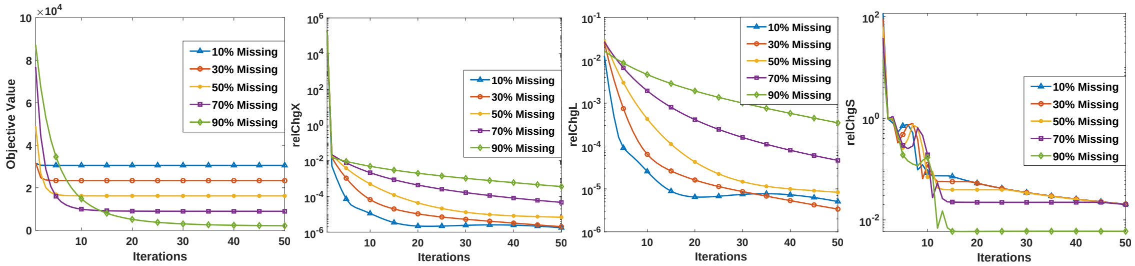
[4] `https://github.com/HuyanHuang/Robust-Low-rank-Tensor-Ring-Completion`

[5] `https://github.com/he1c/robust-tensor-completion`

index (SSIM) are used to measure the recovery accuracy. PSNR and SSIM commonly measure the similarity of two images in intensity and structure, respectively. Specifically, PSNR is defined as: PSNR $= 10 \times \log_{10} \frac{255^2}{\|\mathbf{I}-\hat{\mathbf{I}}\|_F^2}$, where $\mathbf{I}$ and $\hat{\mathbf{I}}$ are respectively the original and recovered background. SSIM measures the structural similarity of two images; see (Wang et al., 2004) for details. Average PSNR and SSIM over all image frames in the video are used to evaluate the recovery performance of the video background. For the task of foreground detection, F-measure is applied to assess the foreground detection performance. Average F-measure over all image frames in the video is applied to evaluate the detection performance of the video foreground. 20 repetitions are sufficient to represent the performance of each method since each repetition is the average performance of multiple image frames. For these performance indices PSNR, SSIM, and F-measure, higher values indicate better performance.

5.1 Convergence Analysis

The video data set *Caviar2* from SBI data set[6] (Maddalena and Petrosino, 2015) is used in this subsection. In total, this video data set has 460 image frames, where the size of each grayscale image is 256×384. For simplicity, the first 80 image frames in the sequence are used for experiments. Therefore, the tensor size is $256 \times 384 \times 80$. One image frame from *Caviar2* in this experiment is shown in Figure 6c. In the video, there are people entering and leaving a store, standing only for a few frames. For each image, some pixels are randomly selected as missing pixels, and the positions of the missing pixels are unknown (one example with 50% missing pixels is shown in Figure 6f). To evaluate the convergence of the proposed algorithm, the relative change relChgA $= \frac{\|\boldsymbol{\mathcal{A}}^k - \boldsymbol{\mathcal{A}}^{k-1}\|_F}{\max(1, \|\boldsymbol{\mathcal{A}}^{k-1}\|_F)}$ and the relative error relErrA $= \frac{\|\boldsymbol{\mathcal{A}}^k - \boldsymbol{\mathcal{A}}^*\|_F}{\max(1, \|\boldsymbol{\mathcal{A}}^*\|_F)}$ are applied as the assessment indices of the algorithm convergence, where $\boldsymbol{\mathcal{A}}^k$ is the result in k-th iteration and $\boldsymbol{\mathcal{A}}^*$ is the ground truth. The ground truth of the static video background is provided in the first column of Figure 7. For the case of orthogonal matrices $\mathbf{U}$, the relative change has the following representation relChgU $= \frac{\|\mathbf{U}_k\mathbf{U}_k^\top - \mathbf{U}_{k-1}\mathbf{U}_{k-1}^\top\|_F}{\max(1, \|\mathbf{U}_{k-1}\mathbf{U}_{k-1}^\top\|_F)}$.

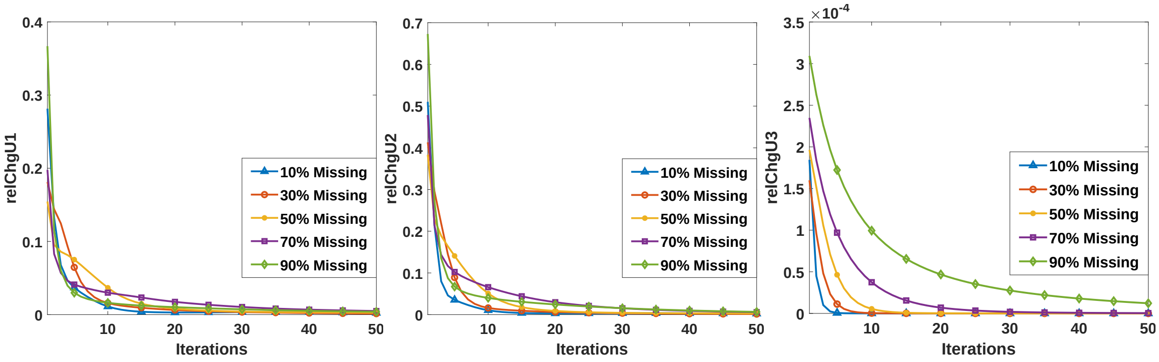


(a) Objective value (b) Relative change of $\boldsymbol{\mathcal{X}}$ (c) Relative change of $\boldsymbol{\mathcal{L}}$ (d) Relative change of $\boldsymbol{\mathcal{S}}$

Figure 3: The empirical convergence analysis of tenPAM algorithm with different ratios of missing pixels (a) Objective value in (3); (b) Relative change of $\boldsymbol{\mathcal{X}}$; (c) Relative change of $\boldsymbol{\mathcal{L}}$; (d) Relative change of $\boldsymbol{\mathcal{S}}$.

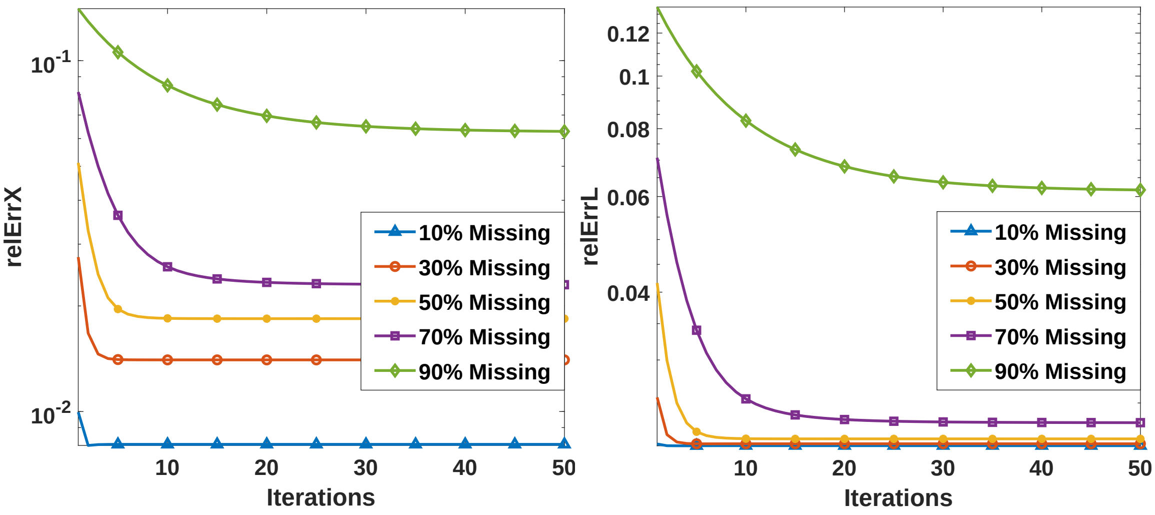
[6] https://sbmi2015.na.icar.cnr.it/SBIdataset.html

In this experiment, parameter λ is set to a value of 0.5, and the ratio of missing pixels can be selected from $\{10\%, 30\%, 50\%, 70\%, 90\%\}$. The curves of the objective value in (3), the relative change of the full video $\boldsymbol{\mathcal{X}}$, the relative change of the video background $\boldsymbol{\mathcal{L}}$, and the relative change of the video foreground $\boldsymbol{\mathcal{S}}$ are shown in Figure 3. Figure 3a illustrates the monotone decreasing trends for the curve of the objective value in (3), which verify the theoretical results in Theorem 14. Meanwhile, it also shows that a bigger ratio of missing pixels implies a smaller objective value with a slower convergence speed because it has fewer data to learn. From Figures 3b, 3c, and 3d, the relative changes of $\boldsymbol{\mathcal{X}}$, $\boldsymbol{\mathcal{L}}$, and $\boldsymbol{\mathcal{S}}$ converge to zero when the number of iterations keeps increasing. Figure 4 illustrates that the relative changes of $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$, and $\mathbf{U}^{(3)}$ converge to zero very fast for different ratios of missing



(a) Relative change of $\mathbf{U}^{(1)}$ (b) Relative change of $\mathbf{U}^{(2)}$ (c) Relative change of $\mathbf{U}^{(3)}$

Figure 4: The empirical convergence analysis of tenPAM algorithm with different ratios of missing pixels: (a) Relative change of $\mathbf{U}^{(1)}$; (b) Relative change of $\mathbf{U}^{(2)}$; (c) Relative change of $\mathbf{U}^{(3)}$.



(a) Relative error of $\boldsymbol{\mathcal{X}}$ (b) Relative change of $\boldsymbol{\mathcal{L}}$

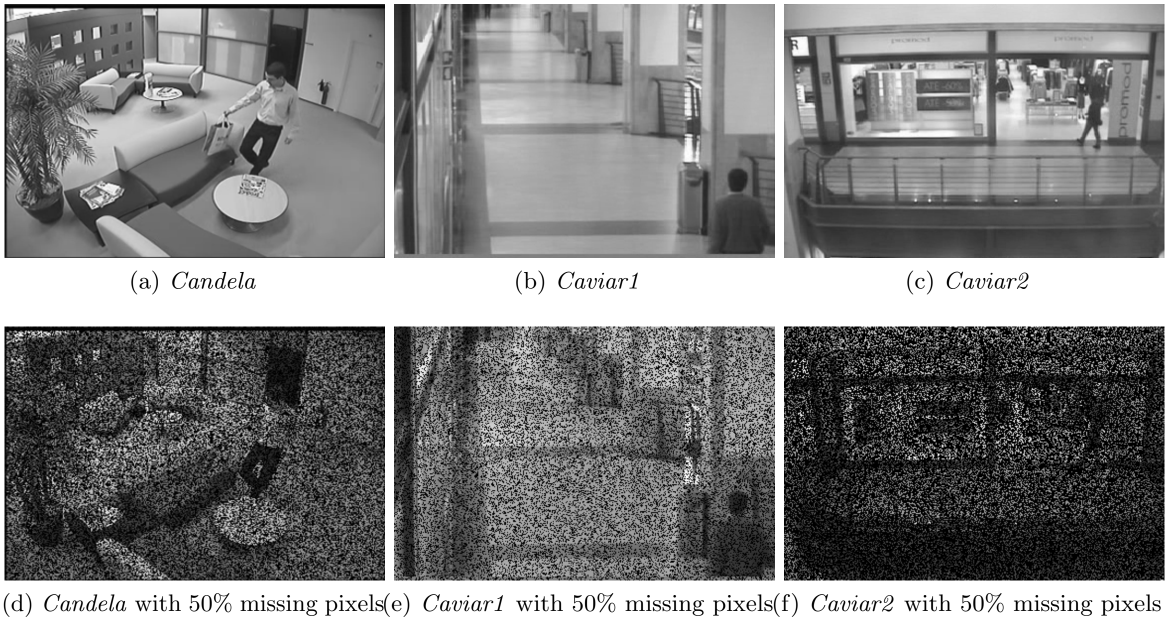
Figure 5: The empirical convergence analysis of tenPAM algorithm with different ratios of missing pixels: (a) Relative error of $\boldsymbol{\mathcal{X}}$; (b) Relative error of $\boldsymbol{\mathcal{L}}$.

pixels. Figures 3 and 4 illustrate that the convergence results in Theorems 14 and 19 are empirically verified.

In addition, the ground truth of full video $\boldsymbol{\mathcal{X}}$ and video background $\boldsymbol{\mathcal{L}}$ is known to us. The curves of the relative error of $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{A}}$ are shown in Figure 5. The curve of the relative error of the video foreground $\boldsymbol{\mathcal{S}}$ is not provided since the ground truth video foreground for actual data is unknown. From Figures 5a and 5b, the relative errors of the full video $\boldsymbol{\mathcal{X}}$ and video background $\boldsymbol{\mathcal{L}}$ gradually decrease to a stable value. In general, the results in this experiment show that our proposed tenPAM algorithm 2 can converge within 50 iterations.

5.2 Background Reconstruction

In this subsection, the proposed method is applied to background reconstruction. The video data set *Caviar2* used in Section 5.1, *Candela*, and *Caviar1* from the SBI data set (Maddalena and Petrosino, 2015) are used for experiments. Figures 6a, 6b, and 6c show three image frames from the three video data sets. In the data set of *Candela*, there is a man entering and leaving a room, abandoning a bag. In the data set of *Caviar1*, there are people slowly walking along a corridor with mild shadows. For all the videos, the first 80 image frames are used for experiments. Thus, the tensor data size is $256 \times 384 \times 80$. The background in each video data set is static, which is provided as the ground truth for comparison. There are people walking in the video, which are treated as the smooth foreground. In each video, some pixels in each image are treated as missing pixels, and the


(a) *Candela* (b) *Caviar1* (c) *Caviar2*

(d) *Candela* with 50% missing pixels(e) *Caviar1* with 50% missing pixels(f) *Caviar2* with 50% missing pixels

Figure 6: Video data sets for background reconstruction: (a) *Candela*; (b) *Caviar1*; (c) *Caviar2*; (d) *Candela* with 50% missing pixels; (e) *Caviar1* with 50% missing pixels; (f) *Caviar2* with 50% missing pixels.

positions of the missing pixels are unknown. The corresponding images with 50% missing pixels from the three video data sets are shown in the second row of Figure 6.

Table 1: Background reconstruction results comparison on different video data sets with different missing ratios

Videos	Missing Ratio	Indices	MCOS (Li et al., 2021)	BFMNM (Shang et al., 2017)	HQ-ASD (He et al., 2019)	RTRC (Huang et al., 2021)	HQ-TCASD (He and Atia, 2020)	SRTC
Candela	10%	PSNR	33.10	40.25	38.55	26.47	24.90	**43.23**
		SSIM	0.8342	0.8784	0.8824	0.7879	0.5656	**0.9011**
	30%	PSNR	33.06	40.23	37.38	26.25	24.68	**43.33**
		SSIM	0.8241	0.8761	0.8573	0.7459	0.5258	**0.9034**
	50%	PSNR	32.90	40.10	35.27	25.69	23.95	**43.34**
		SSIM	0.8062	0.8714	0.8097	0.6840	0.4669	**0.9048**
	70%	PSNR	17.65	39.35	31.82	24.54	22.46	**43.29**
		SSIM	0.3225	0.8596	0.7164	0.5863	0.3839	**0.9044**
	90%	PSNR	8.74	15.95	23.07	21.75	19.18	**41.14**
		SSIM	0.0942	0.3442	0.4568	0.3857	0.2337	**0.8555**
Caviar1	10%	PSNR	29.07	29.91	31.83	24.92	24.63	**36.68**
		SSIM	0.7769	0.7703	0.7762	0.7201	0.6898	**0.7992**
	30%	PSNR	23.04	34.46	31.61	25.03	24.65	**36.74**
		SSIM	0.3736	0.7779	0.7584	0.7225	0.6288	**0.8028**
	50%	PSNR	17.88	35.32	30.64	25.14	24.60	**36.75**
		SSIM	0.2032	0.7762	0.7296	0.7211	0.5490	**0.8051**
	70%	PSNR	15.54	31.22	28.98	25.23	24.24	**36.78**
		SSIM	0.1681	0.7497	0.6488	0.7072	0.4560	**0.8068**
	90%	PSNR	8.27	14.84	22.20	24.89	22.21	**36.62**
		SSIM	0.0463	0.1869	0.3636	0.6079	0.3025	**0.7901**
Caviar2	10%	PSNR	36.28	43.72	43.34	33.87	31.49	**44.65**
		SSIM	0.9443	0.9358	0.9485	0.8884	0.8721	**0.9493**
	30%	PSNR	27.04	43.64	41.03	34.09	31.11	**44.79**
		SSIM	0.6157	0.9357	0.9300	0.8921	0.8312	**0.9509**
	50%	PSNR	21.59	43.46	36.66	33.66	29.52	**44.88**
		SSIM	0.4452	0.9338	0.8941	0.8915	0.7460	**0.9518**
	70%	PSNR	18.55	42.48	31.62	31.82	26.96	**44.85**
		SSIM	0.3681	0.9276	0.8241	0.8752	0.6541	**0.9508**
	90%	PSNR	9.24	14.43	22.44	26.42	22.49	**43.31**
		SSIM	0.0914	0.3248	0.5415	0.7419	0.4838	**0.9362**

Note: The bold red numbers are the best performance for each case.

In this experiment, the cases of 10%, 30%, 50%, 70%, and 90% missing pixels are studied to show the performance of background reconstruction under different missing ratios. The quantitative results of all benchmark methods with different missing ratios on simulated *Candela*, *Caviar1*, and *Caviar2* are summarized in Table 1 regarding PSNR and SSIM. For all cases, our method achieves the best performance in terms of PSNR and SSIM. When the missing ratio increases, our proposed SRTC is the most consistent among the benchmark methods. Specifically, our approach has a very small variation for different missing ratios and is the only method that performs well for the case of 90% missing pixels. Meanwhile, the performances of MCOS, BFMNM, and HQ-ASD degrade significantly when the missing ratio increases. RTRC and HQ-TCASD perform poorly for all cases. These results show that the proposed method has the best performance in terms of accuracy due to the advantage of low-rank Tucker decomposition for the static background model over the nuclear norm.

The background reconstruction results from the case of 50% missing ratio on all three video data sets are visualized in Figure 7. Our proposed approach generally produces the cleanest background for all three video data sets. For the backgrounds generated using RTRC and HQ-TCASD, people still remain in the background even though the image is recovered. For the results from BFMNM and HQ-ASD, their performance is very close to the proposed method but still poor for the *Caviar1*, where a shadow of people is left in

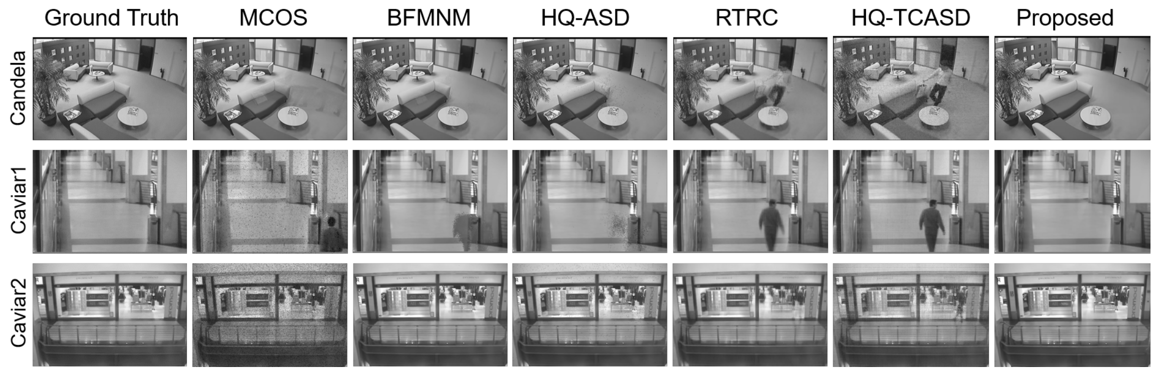
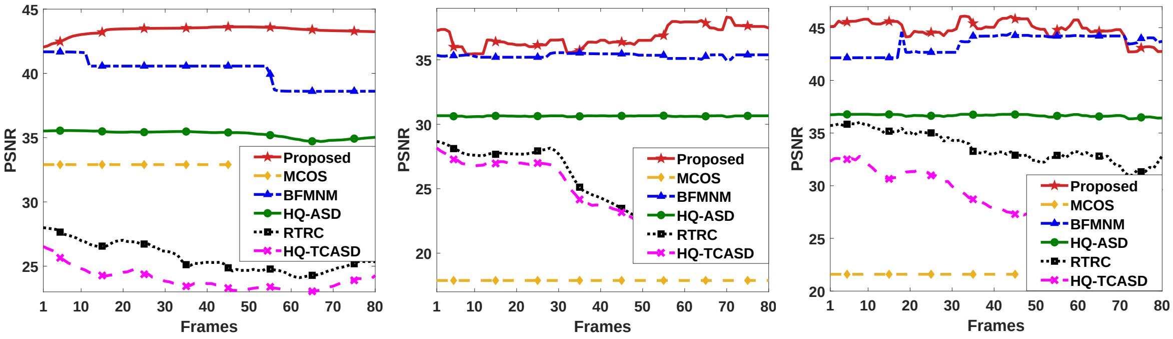


Figure 7: Visualization results of different methods on a frame of *Candela*, *Caviar1*, and *Caviar2* for background reconstruction with 50% missing pixels.

the background. For MOCS, it cannot recover the video background where there are still missing pixels in *Caviar1* and *Caviar2*.



(a) *Candela* with 50% missing pixels(b) *Caviar1* with 50% missing pixels(c) *Caviar2* with 50% missing pixels

Figure 8: Background reconstruction performance comparison by frames for different methods: (a) *Candela* with 50% missing pixels; (b) *Caviar1* 50% missing pixels; (c) *Caviar2* 50% missing pixels.

Since the quantitative results in Table 1 are the average performance of each algorithm for 80 frames in the video, the performance variability for each frame needs to be analyzed. For the case of 50% missing ratio, Figure 8 shows the PSNR of each frame in videos of *Candela*, *Caviar1*, and *Caviar2*. This result shows that the proposed method is very stable since the variation among different frames is quite small. Besides, our proposed method achieves the best performance for almost every frame, except for few frames in *Caviar2*. BFMNM is the only comparable method with the second-best performance, while other benchmark methods perform very poorly.

5.3 Foreground Detection

In this subsection, our proposed method is applied to foreground detection. The video data sets, namely, *Highway*, *Office*, and *Pedestrians* from CDnet data set[7] (Wang et al., 2014), are used for experiments. Figures 9a, 9b, and 9c show three image frames from the three video data sets, respectively. In the data set of *Highway*, there are vehicles moving along the highway, where the size of each grayscale image is 240 × 320. In the data set of *Office*, there is a man entering, staying, and leaving the room, where the size of each grayscale image is 240 × 360. In the data set of *Pedestrians*, there are people walking on the road, where the size of each grayscale image is 240 × 360. For all the videos, the first 80 image frames are used for experiments for the sake of computational time. The background in each video data set is static. The binary masks for the video foreground are provided as ground truth for comparison. In each video, some pixels in each image are set as missing pixels, and the positions of the missing pixels are unknown. The corresponding images with 70% missing pixels from the three video data sets are shown in the second row of Figure 9.

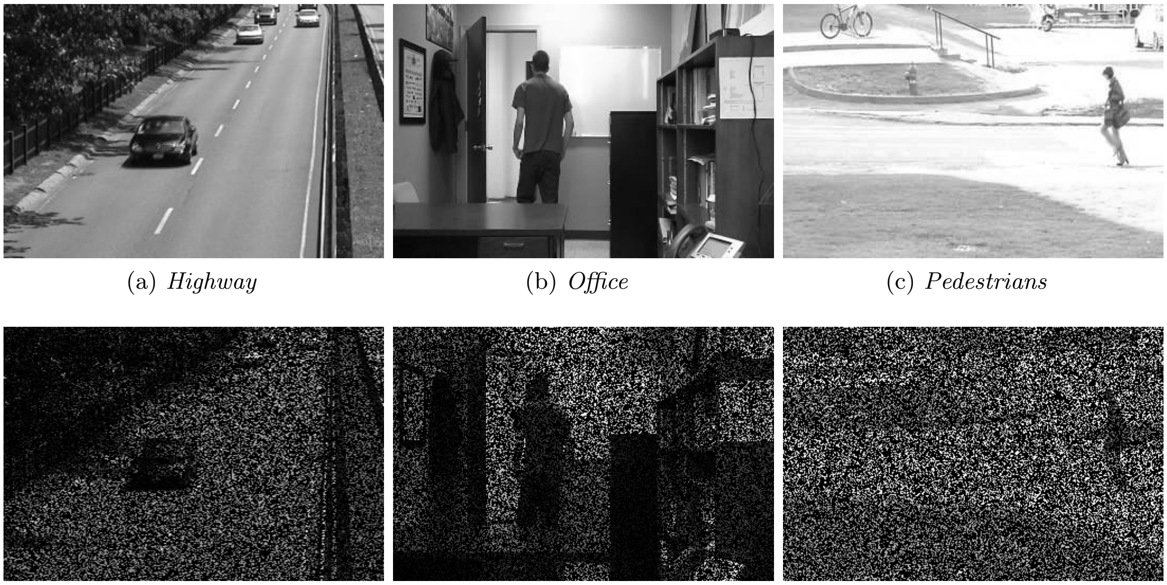

(a) *Highway*
(b) *Office*
(c) *Pedestrians*
(d) *Highway* with 70% missing pixels
(e) *Office* with 70% missing pixels
(f) *Pedestrians* with 70% missing pixels

Figure 9: Video data sets for foreground detection: (a) *Highway*; (b) *Office*; (c) *Pedestrians*; (d) *Highway* with 70% missing pixels; (e) *Office* with 70% missing pixels; (f) *Pedestrians* with 70% missing pixels.

In this experiment, the cases of 10%, 30%, 50%, 70%, and 90% missing pixels are investigated to show the performance of foreground detection under different missing ratios. The quantitative results of all benchmark methods with different missing ratios of simulated *Highway*, *Office*, and *Pedestrians* are summarized in Table 2 in terms of Precision, Recall, and F-measure. In terms of precision, our method achieves the best performance, except in

[7] `http://jacarini.dinf.usherbrooke.ca/dataset2014/`

Table 2: Foreground detection results comparison on different video data sets with different missing ratios

Videos	Missing Ratio	Indexs	MCOS (Li et al., 2021)	BFMNM (Shang et al., 2017)	HQ-ASD (He et al., 2019)	RTRC (Huang et al., 2021)	HQ-TCASD (He and Atia, 2020)	SRTC
Highway	10%	Precision	0.5966	**0.7881**	0.6944	0.5276	0.1917	0.7810
		Recall	0.7968	0.2350	0.8089	0.6418	0.2237	**0.9443**
		F-measure	0.6794	0.3557	0.7457	0.5786	0.2044	**0.8546**
	30%	Precision	0.5901	0.7733	0.5507	0.4219	0.1476	**0.7864**
		Recall	0.7965	0.1029	0.7928	0.5968	0.1657	**0.9453**
		F-measure	0.6756	0.1791	0.6480	0.4934	0.1539	**0.8583**
	50%	Precision	0.4871	0.4646	0.3986	0.3099	0.1085	**0.7919**
		Recall	0.2953	0.0891	0.7371	0.5319	0.0988	**0.9456**
		F-measure	0.3644	0.1453	0.5155	0.3906	0.0987	**0.8616**
	70%	Precision	0.3042	0.4922	0.2416	0.1903	0.1313	**0.8068**
		Recall	0.1481	0.0838	0.6430	0.4375	0.0567	**0.9368**
		F-measure	0.1951	0.1412	0.3495	0.2642	0.0727	**0.8666**
	90%	Precision	0.4234	0.4768	0.0797	0.0653	0.0902	**0.7775**
		Recall	0.0695	0.0559	0.3935	0.2663	0.0379	**0.9315**
		F-measure	0.1003	0.0936	0.1314	0.1035	0.0520	**0.8466**
Office	10%	Precision	0.6602	0.6725	0.6373	0.3291	0.2844	**0.8539**
		Recall	0.6033	0.2673	0.5988	0.2669	0.1307	**0.9443**
		F-measure	0.6158	0.3787	0.6017	0.2926	0.1761	**0.8954**
	30%	Precision	0.5971	0.7364	0.5026	0.2721	0.2383	**0.8580**
		Recall	0.4194	0.1417	0.5414	0.2568	0.1247	**0.9415**
		F-measure	0.4779	0.2348	0.5170	0.2616	0.1601	**0.8963**
	50%	Precision	0.5258	0.8091	0.3738	0.2051	0.2796	**0.8584**
		Recall	0.2197	0.1022	0.5049	0.2481	0.0834	**0.9450**
		F-measure	0.2844	0.1787	0.4250	0.2211	0.1229	**0.8982**
	70%	Precision	0.6565	0.7505	0.2274	0.1272	0.2904	**0.8601**
		Recall	0.0850	0.0867	0.4403	0.2435	0.0665	**0.9483**
		F-measure	0.1496	0.1536	0.2965	0.1643	0.1063	**0.9008**
	90%	Precision	0.8400	0.8042	0.0790	0.0446	0.0992	**0.8490**
		Recall	0.0720	0.0782	0.2646	0.2442	0.0657	**0.9397**
		F-measure	0.1303	0.1404	0.1211	0.0746	0.0772	**0.8908**
Pedestrians	10%	Precision	0.6252	**0.8364**	0.6931	0.6042	0.2750	0.7737
		Recall	0.8304	0.1653	0.8708	0.7026	0.4153	**0.9626**
		F-measure	0.7115	0.2758	0.7713	0.6490	0.3291	**0.8577**
	30%	Precision	0.6197	**0.8421**	0.5429	0.4823	0.2073	0.7849
		Recall	0.8256	0.0697	0.7956	0.6728	0.3385	**0.9617**
		F-measure	0.7062	0.1238	0.6446	0.5612	0.2553	**0.8642**
	50%	Precision	0.3066	0.2665	0.3913	0.3462	0.1232	**0.7972**
		Recall	0.1251	0.0868	0.6984	0.6392	0.2251	**0.9626**
		F-measure	0.1667	0.1209	0.5007	0.4484	0.1570	**0.8719**
	70%	Precision	0.5241	0.3147	0.2416	0.2105	0.0834	**0.8159**
		Recall	0.0320	0.0864	0.5373	0.5668	0.1306	**0.9583**
		F-measure	0.0586	0.1299	0.3329	0.3063	0.0990	**0.8811**
	90%	Precision	0.2561	0.8010	0.0818	0.0721	0.0703	**0.8448**
		Recall	0.0324	0.0272	0.2427	0.3659	0.0375	**0.9283**
		F-measure	0.0539	0.0486	0.1220	0.1201	0.0428	**0.8838**

Note: The bold red numbers are the best performance for each case.

one case of *Highway* with 10% missing pixels. In terms of recall and F-measure, the proposed SRTC can achieve the best performance in all cases. Our proposed SRTC is the most consistent one while all the benchmark methods show significant performance degradation when the missing ratio increases. Specifically, the performances of our method with 10%, 30%, 50%, 70%, and 90% missing ratios are almost the same. MCOS and HQ-ASD have good performance when the missing ratio is small. BFMNM and HQ-TCASD perform very poorly for all cases. Overall, our proposed approach shows the best performance due to the advantage of smoothness modeling of the video foreground instead of sparsity.

The case of 70% noise ratio is used to show the foreground detection results. Specifically, the visualizations of one frame from each video data set for different methods are shown in Figure 10. Our method can generally detect the most accurate foreground among all benchmarks even though the video is missing. For the foreground masks subtracted from

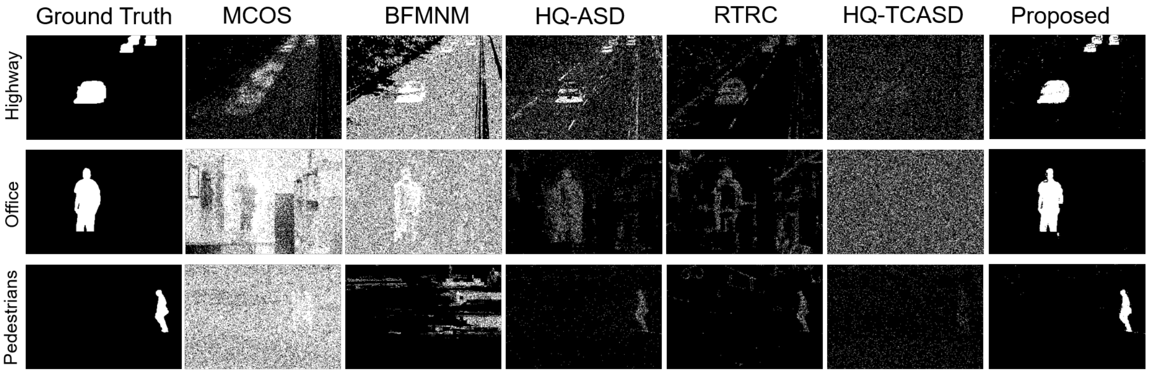
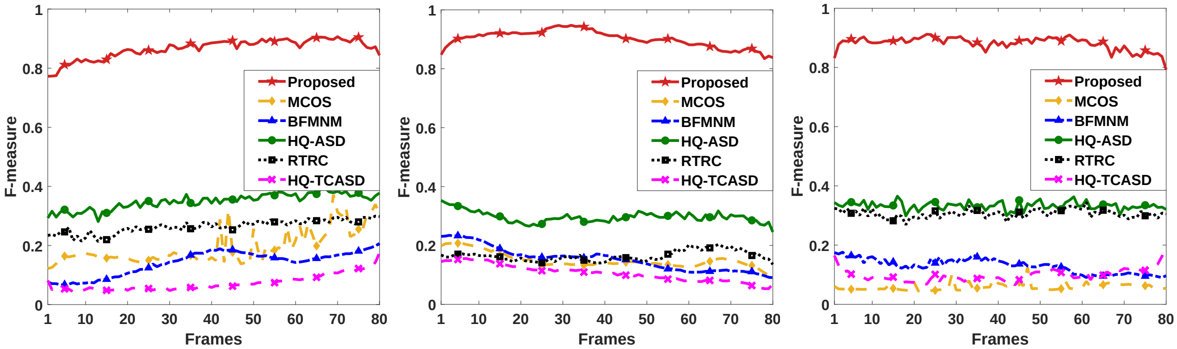


Figure 10: Visualization results of different methods on a frame of *Highway*, *Office*, and *Pedestrians* for the foreground with 70% missing pixels.



(a) *Highway* with 70% missing pixels (b) *Office* with 70% missing pixels (c) *Pedestrians* with 70% missing pixels

Figure 11: Foreground detection performance comparison by frames for different methods: (a) *Highway* with 70% missing pixels; (b) *Office* with 70% missing pixels; (c) *Pedestrians* with 70% missing pixels.

MCOS, BFMNM, and HQ-TCASD, a lot of noise caused by the missing pixels remains in the foreground, where the moving objects are not well detected either. The results from HQ-ASD and RTRC can detect the shape of the moving objects. However, the detected foreground is not as complete as our result.

Since the quantitative results in Table 2 are the average performance of each algorithm for 80 frames in the video, the performance variability for each frame needs further analysis. For the case of 70% missing ratio, Figure 11 shows the F-measure of each frame in videos of *Highway*, *Office*, and *Pedestrians*. This result shows that the proposed method is very stable since the variation among different frames is quite small. In addition, our proposed method achieves much better performance than all other benchmark methods for all frames.

6. Conclusion

In this article, a new smooth robust tensor completion is developed for background/foreground separation with missing pixels. The proposed SRTC simultaneously recovers the video data and decomposes it into low-rank and smooth components, respectively. To obtain the solutions efficiently, a tensor proximal alternating minimization (tenPAM) algorithm is implemented for SRTC. The global convergence of the tenPAM algorithm has been established under very mild conditions. The empirical convergence experiment shows that the proposed SRTC can converge and run efficiently in practice. The background reconstruction and foreground detection results on simulated video data demonstrate that our method outperforms MCOS, BFMNM, HQ-ASD, RTRC, and HQ-TCASD, which are current state-of-the-art algorithms in the literature. These results illustrate the effectiveness of Tucker decomposition for the low-rank tensor and total variation regularization for the smooth tensor.

Acknowledgments

Dr. Shen and Dr. Kong are supported by the Office of Naval Research under Award Number N00014-18-1-2794, and the Department of Defense under Award Number N00014-19-1-2728. Dr. Xie has been supported in part by the National Science Foundation grants 2046426 and 2153607.

Appendix A. Additional Theoretical Results

Appendix A1. Convergence of U

In our main paper, the convergence of the sequence $\{\mathbf{U}_k\mathbf{U}_k^\top\}_{k>0}$ from our Algorithm 2 can be derived. The following theorem shows how we can get the convergent sequence of $\{\mathbf{U}_{k,new}\}_{k>0}$ from $\{\mathbf{U}_k\mathbf{U}_k^\top\}_{k>0}$.

Theorem 20 *Let $\mathbf{U}_*\mathbf{U}_*^\top = \{\mathbf{U}_*^{(1)}\mathbf{U}_*^{(1)\top}, \mathbf{U}_*^{(2)}\mathbf{U}_*^{(2)\top}, \dots, \mathbf{U}_*^{(N)}\mathbf{U}_*^{(N)\top}\}$ be the limit point of Algorithm 2. By SVD, we have $\mathbf{U}_k^{(n)\top}\mathbf{U}_*^{(n)} = \mathbf{W}\Sigma\mathbf{V}$. Now let $\mathbf{D}_k^{(n)} = \mathbf{W}\mathbf{V}$ and $\mathbf{U}_{k,new}^{(n)} = \mathbf{U}_k^{(n)}\mathbf{D}_k^{(n)}$, then*

$$\lim_{k\to\infty} \mathbf{U}_{k,new}^{(n)} = \mathbf{U}_*^{(n)}.$$

Proof For any n, we have that $\lim_{k\to\infty} \mathbf{U}_k^{(n)}\mathbf{U}_k^{(n)\top} = \mathbf{U}_*^{(n)}\mathbf{U}_*^{(n)\top}$, therefore, $\lim_{k\to\infty} \|\mathbf{U}_k^{(n)}\mathbf{U}_k^{(n)\top} - \mathbf{U}_*^{(n)}\mathbf{U}_*^{(n)\top}\|_F = 0$. We first claim that $\mathbf{D}_k^{(n)}$ is the optimal solution of the Orthogonal Procrustes problem as below

$$\mathbf{D}_k^{(n)} \in \operatorname*{arg\,min}_{\mathbf{D}\in\mathbb{R}^{r_n\times r_n}} \|\mathbf{U}_*^{(n)} - \mathbf{U}_k^{(n)}\mathbf{D}\|_F,$$

where $\mathbf{D}$ is an orthogonal matrix. The objective function in the above problem can be rewritten as

$$\|\mathbf{U}_*^{(n)} - \mathbf{U}_k^{(n)}\mathbf{D}\|_F^2 = \mathrm{Tr}\big((\mathbf{U}_*^{(n)} - \mathbf{U}_k^{(n)}\mathbf{D})(\mathbf{U}_*^{(n)} - \mathbf{U}_k^{(n)}\mathbf{D})^\top\big)$$

$$
\begin{aligned}
&= \mathrm{Tr}(\mathbf{U}_*^{(n)}\mathbf{U}_*^{(n)\top}) + \mathrm{Tr}(\mathbf{U}_k^{(n)}\mathbf{U}_k^{(n)\top}) - 2\mathrm{Tr}(\mathbf{U}_k^{(n)\top}\mathbf{U}_*^{(n)}\mathbf{D}^\top)\\
&= 2r_n - 2\mathrm{Tr}(\mathbf{U}_k^{(n)\top}\mathbf{U}_*^{(n)}\mathbf{D}^\top).
\end{aligned}
$$

Equivalently,

$$
\mathbf{D}_k^{(n)} \in \operatorname*{arg\,max}_{\mathbf{D}\in\mathbb{R}^{r_n\times r_n}} \mathrm{Tr}(\mathbf{U}_k^{(n)\top}\mathbf{U}_*^{(n)}\mathbf{D}^\top).
$$

By SVD, $\mathbf{U}_k^{(n)\top}\mathbf{U}_*^{(n)} = \mathbf{W}\Sigma\mathbf{V}$. $\mathbf{D}_k^{(n)} = \mathbf{W}\mathbf{V}$ is the optimal solution for the above problem, but it may not be a unique solution. Next, the optimal objective function value can be obtained

$$
\begin{aligned}
\max_{\mathbf{D}} \mathrm{Tr}(\mathbf{U}_k^{(n)\top}\mathbf{U}_*^{(n)}\mathbf{D}^\top) &= \|\mathbf{U}_k^{(n)\top}\mathbf{U}_*^{(n)}\|_*\\
&= \sum_{i=1}^{r_n}\sigma_i(\mathbf{U}_k^{(n)\top}\mathbf{U}_*^{(n)}),
\end{aligned}
$$

where $\sigma_i(\cdot)$ is the i-th largest singular value of a matrix. We claim that $\forall k, n$, $\sigma_i(\mathbf{U}_k^{(n)\top}\mathbf{U}_*^{(n)}) \leq 1$ because

$$
\begin{aligned}
\sigma_i(\mathbf{U}_k^{(n)\top}\mathbf{U}_*^{(n)}) &\leq \sqrt{\lambda_{\max}(\mathbf{U}_k^{(n)\top}\mathbf{U}_*^{(n)}\mathbf{U}_*^{(n)\top}\mathbf{U}_k^{(n)})}\\
&= \sqrt{\max_{\|\boldsymbol{x}\|_2=1}\mathrm{Tr}(\boldsymbol{x}^\top\mathbf{U}_k^{(n)\top}\mathbf{U}_*^{(n)}\mathbf{U}_*^{(n)\top}\mathbf{U}_k^{(n)}\boldsymbol{x})}\\
&\leq \sqrt{\max_{\|\boldsymbol{y}\|_2=1}\mathrm{Tr}(\boldsymbol{y}^\top\mathbf{U}_*^{(n)}\mathbf{U}_*^{(n)\top}\boldsymbol{y})}\\
&= \sqrt{\lambda_{\max}(\mathbf{U}_*^{(n)}\mathbf{U}_*^{(n)\top})} = 1
\end{aligned}
$$

where $\lambda_{\max}(\cdot)$ is the maximum eigenvalue of a positive semidefinite matrix. The second inequality is because $\boldsymbol{y} = \mathbf{U}_k^{(n)}\boldsymbol{x}$ may not span the entire $\mathbb{R}^{r_n}$ space. Finally,

$$
\begin{aligned}
\|\mathbf{U}_*^{(n)} - \mathbf{U}_k^{(n)}\mathbf{D}_k^{(n)}\|_F^2 &= 2r_n - 2\sum_{i=1}^{r_n}\sigma_i(\mathbf{U}_k^{(n)\top}\mathbf{U}_*^{(n)})\\
&\leq 2r_n - 2\sum_{i=1}^{r_n}\sigma_i(\mathbf{U}_k^{(n)\top}\mathbf{U}_*^{(n)})^2\\
&= 2r_n - 2\sum_{i=1}^{r_n}\lambda_i(\mathbf{U}_k^{(n)\top}\mathbf{U}_*^{(n)}\mathbf{U}_*^{(n)\top}\mathbf{U}_k^{(n)})\\
&= 2r_n - 2\mathrm{Tr}(\mathbf{U}_k^{(n)\top}\mathbf{U}_*^{(n)}\mathbf{U}_*^{(n)\top}\mathbf{U}_k^{(n)})\\
&= \|\mathbf{U}_k^{(n)}\mathbf{U}_k^{(n)\top} - \mathbf{U}_*^{(n)}\mathbf{U}_*^{(n)\top}\|_F^2,
\end{aligned}
$$

where $\lambda_i(\cdot)$ is the i-th largest eigenvalue of a matrix. The first inequality is due to $x \geq x^2, \forall x \in [0,1]$. Now let $\mathbf{U}_{k,new}^{(n)} = \mathbf{U}_k^{(n)}\mathbf{D}_k^{(n)}$, then $\lim_{k\to\infty}\mathbf{U}_{k,new}^{(n)} = \mathbf{U}_*^{(n)}$. ■

Appendix A2. Rate of Convergence of tenPAM Algorithm 2

The following result shows that if we can specify the θ value in Definition 16 of the function $G(\cdot)$, then we are able to establish the rate of convergence of tenPAM algorithm.

Theorem 21 (*Rate of Convergence* (Attouch and Bolte, 2009; Bolte et al., 2014)) *Suppose that the desingularizing function of the function $G(\cdot)$ is in the form of $\phi(s) = cs^{1-\theta}$ for some positive constant $c > 0$. Then one of the following results must hold*

(i) If $\theta = 0$, then the sequence $\{\mathbf{U}_k\mathbf{U}_k^\top, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k\}_{k\geq 0}$ converges after a finite number of iterations;

(ii) If $\theta \in (0, 1/2]$, then there exist $\eta > 0$ and $\tau \in [0, 1)$ such that $\|(\mathbf{U}_k\mathbf{U}_k^\top, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k) - (\mathbf{U}_*\mathbf{U}_*^\top, \boldsymbol{\mathcal{S}}^*, \boldsymbol{\mathcal{X}}^*)\|_F \leq \eta\tau^k$; and

(iii) If $\theta \in (1/2, 1)$, then there exist $\eta > 0$ such that

$$\|(\mathbf{U}_k\mathbf{U}_k^\top, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k) - (\mathbf{U}_*\mathbf{U}_*^\top, \boldsymbol{\mathcal{S}}^*, \boldsymbol{\mathcal{X}}^*)\|_F \leq \eta k^{-\frac{1-\theta}{2\theta-1}},$$

where $(\mathbf{U}_*\mathbf{U}_*^\top, \boldsymbol{\mathcal{S}}^*, \boldsymbol{\mathcal{X}}^*)$ is the limit point of the sequence $\{\mathbf{U}_k\mathbf{U}_k^\top, \boldsymbol{\mathcal{S}}^k, \boldsymbol{\mathcal{X}}^k\}_{k\geq 0}$.

However, it is difficult to specify the θ of the function $G(\cdot)$. Thus, the rate of convergence of tenPAM algorithm is unknown in general. This could be one of our future research directions.

References

Hedy Attouch and Jérôme Bolte. On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. *Mathematical Programming*, 116(1-2):5–16, 2009.

Hédy Attouch, Jérôme Bolte, Patrick Redont, and Antoine Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality. *Mathematics of Operations Research*, 35(2):438–457, 2010.

Hedy Attouch, Jérôme Bolte, and Benar Fux Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods. *Mathematical Programming*, 137(1-2):91–129, 2013.

Jérôme Bolte, Aris Daniilidis, and Adrian Lewis. The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM Journal on Optimization*, 17(4):1205–1223, 2007.

Jérôme Bolte, Shoham Sabach, and Marc Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, 146(1-2): 459–494, 2014.

Thierry Bouwmans, Lucia Maddalena, and Alfredo Petrosino. Scene background initialization: A taxonomy. *Pattern Recognition Letters*, 96:3–11, 2017a.

Thierry Bouwmans, Andrews Sobral, Sajid Javed, Soon Ki Jung, and El-Hadi Zahzah. Decomposition into low-rank plus additive matrices for background/foreground separation: A review for a comparative evaluation with a large-scale dataset. *Computer Science Review*, 23:1–71, 2017b.

Emmanuel J Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772, 2009.

Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust Principal Component Analysis? *Journal of the ACM (JACM)*, 58(3):1–37, 2011.

Wenfei Cao, Yao Wang, Jian Sun, Deyu Meng, Can Yang, Andrzej Cichocki, and Zongben Xu. Total variation regularized tensor RPCA for background subtraction from compressive measurements. *IEEE Transactions on Image Processing*, 25(9):4075–4090, 2016.

Xiaochun Cao, Liang Yang, and Xiaojie Guo. Total variation regularized RPCA for irregularly moving object detection under dynamic background. *IEEE Transactions on Cybernetics*, 46(4):1014–1027, 2015.

Stanley H Chan, Ramsin Khoshabeh, Kristofor B Gibson, Philip E Gill, and Truong Q Nguyen. An augmented lagrangian method for total variation video restoration. *IEEE Transactions on Image Processing*, 20(11):3097–3111, 2011.

Shixiang Chen, Shiqian Ma, Anthony Man-Cho So, and Tong Zhang. Proximal gradient method for nonsmooth optimization over the Stiefel manifold. *SIAM Journal on Optimization*, 30(1):210–239, 2020.

Yudong Chen, Huan Xu, Constantine Caramanis, and Sujay Sanghavi. Matrix completion with column manipulation: Near-optimal sample-robustness-rank tradeoffs. *IEEE Transactions on Information Theory*, 62(1):503–526, 2015.

Patrick L Combettes and Jean-Christophe Pesquet. Proximal splitting methods in signal processing. In *Fixed-point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer, 2011.

David L Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627, 1995.

Haiyan Fan, Gangyao Kuang, and Linbo Qiao. Fast Tensor Principal Component Analysis via proximal alternating direction method with vectorized technique. *Applied Mathematics*, 8(01):77, 2017.

Ferenc Firtha, András Fekete, Tímea Kaszab, Bíborka Gillay, Médea Nogula-Nagy, Zoltán Kovács, and David B Kantor. Methods for improving image quality and reducing data load of NIR hyperspectral images. *Sensors*, 8(5):3287–3298, 2008.

Belmar Garcia-Garcia, Thierry Bouwmans, and Alberto Jorge Rosales Silva. Background subtraction in real applications: Challenges, current models and future directions. *Computer Science Review*, 35:100204, 2020.

Tom Goldstein and Stanley Osher. The split bregman method for $l1$-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.

Yicong He and George K Atia. Robust low-tubal-rank tensor completion based on tensor factorization and maximum correntopy criterion. *arXiv preprint arXiv:2010.11740*, 2020.

Yicong He, Fei Wang, Yingsong Li, Jing Qin, and Badong Chen. Robust matrix completion via maximum correntropy criterion and half-quadratic optimization. *IEEE Transactions on Signal Processing*, 68:181–195, 2019.

Mingyi Hong and Zhi-Quan Luo. On the linear convergence of the alternating direction method of multipliers. *Mathematical Programming*, 162(1):165–199, 2017.

Minhui Huang, Shiqian Ma, and Lifeng Lai. Robust low-rank matrix completion via an alternating manifold proximal gradient continuation method. *IEEE Transactions on Signal Processing*, 69:2639–2652, 2021.

Sajid Javed, Thierry Bouwmans, and Soon Ki Jung. SBMI-LTD: Stationary background model initialization based on low-rank tensor decomposition. In *Proceedings of the Symposium on Applied Computing*, pages 195–200, 2017.

Ibrahim Kajo, Nidal Kamel, Yassine Ruichek, and Aamir Saeed Malik. SVD-based tensor-completion technique for background initialization. *IEEE Transactions on Image Processing*, 27(6):3114–3126, 2018.

Ibrahim Kajo, Nidal Kamel, and Yassine Ruichek. Self-motion-assisted tensor completion method for background initialization in complex video sequences. *IEEE Transactions on Image Processing*, 29:1915–1928, 2019.

Misha E Kilmer and Carla D Martin. Factorization strategies for third-order tensors. *Linear Algebra and its Applications*, 435(3):641–658, 2011.

Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

Adrian S Lewis and Jérôme Malick. Alternating projections on manifolds. *Mathematics of Operations Research*, 33(1):216–234, 2008.

Xiao Peng Li and Hing Cheung So. Robust low-rank tensor completion based on tensor ring rank via $\ell_{p,\epsilon}$-norm. *IEEE Transactions on Signal Processing*, 69:3685–3698, 2021.

Ziheng Li, Zhanxuan Hu, Feiping Nie, Rong Wang, and Xuelong Li. Matrix completion with column outliers and sparse noise. *Information Sciences*, 573:125–140, 2021.

Ji Liu, Przemyslaw Musialski, Peter Wonka, and Jieping Ye. Tensor completion for estimating missing values in visual data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):208–220, 2012.

Yipeng Liu, Longxi Chen, and Ce Zhu. Improved robust tensor Principal Component Analysis via low-rank core matrix. *IEEE Journal of Selected Topics in Signal Processing*, 12(6):1378–1389, 2018.

Canyi Lu, Jiashi Feng, Yudong Chen, Wei Liu, Zhouchen Lin, and Shuicheng 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*, pages 5249–5257, 2016.

Lucia Maddalena and Alfredo Petrosino. Towards benchmarking scene background initialization. In *International Conference on Image Analysis and Processing*, pages 469–476. Springer, 2015.

Diana Mărginean Petrovai. The global convergence of the algorithms described by multifunctions. *Procedia Engineering*, 181:924–927, 2017.

Pu Ren, Xinyu Chen, Lijun Sun, and Hao Sun. Incremental bayesian matrix/tensor learning for structural monitoring data imputation and response forecasting. *Mechanical Systems and Signal Processing*, 158:107734, 2021.

Fanhua Shang, James Cheng, Yuanyuan Liu, Zhi-Quan Luo, and Zhouchen Lin. Bilinear factor matrix norm minimization for robust PCA: Algorithms and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(9):2066–2080, 2017.

Bo Shen, Rakesh R Kamath, Hahn Choo, and Zhenyu Kong. Robust tensor decomposition based background/foreground separation in noisy videos and its applications in additive manufacturing. *TechRXiv*, 2021.

Andrews Sobral and El-hadi Zahzah. Matrix and tensor completion algorithms for background model initialization: A comparative evaluation. *Pattern Recognition Letters*, 96:22–33, 2017.

Andrews Sobral, Thierry Bouwmans, and El-hadi Zahzah. Lrslibrary: Low-rank and sparse tools for background modeling and subtraction in videos. *Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing*, 2016.

G. Strang. *Introduction to Linear Algebra.* Wellesley-Cambridge Press, 2016. ISBN 9780980232776. URL `https://books.google.com/books?id=efbxjwEACAAJ`.

Bokun Wang, Shiqian Ma, and Lingzhou Xue. Riemannian stochastic proximal gradient methods for nonsmooth optimization over the Stiefel manifold. *arXiv preprint arXiv:2005.01209*, 2020.

Yao Wang, Jiangjun Peng, Qian Zhao, Yee Leung, Xi-Le Zhao, and Deyu Meng. Hyperspectral image restoration via total variation regularized low-rank tensor decomposition. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11 (4):1227–1243, 2017.

Yi Wang, Pierre-Marc Jodoin, Fatih Porikli, Janusz Konrad, Yannick Benezeth, and Prakash Ishwar. Cdnet 2014: An expanded change detection benchmark dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 387–394, 2014.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13 (4):600–612, 2004.

Yangyang Xu. On the convergence of Higher-order Orthogonal Iteration. *Linear and Multilinear Algebra*, 66(11):2247–2265, 2018.

Jing-Hua Yang, Xi-Le Zhao, Teng-Yu Ji, Tian-Hui Ma, and Ting-Zhu Huang. Low-rank tensor train for Tensor Robust Principal Component Analysis. *Applied Mathematics and Computation*, 367:124783, 2020.

Hongyan Zhang, Lu Liu, Wei He, and Liangpei Zhang. Hyperspectral image denoising with total variation regularization and nonlocal low-rank tensor decomposition. *IEEE Transactions on Geoscience and Remote Sensing*, 58(5):3071–3084, 2019.

Zemin Zhang, Gregory Ely, Shuchin Aeron, Ning Hao, and Misha Kilmer. Novel methods for multilinear data completion and de-noising based on tensor-SVD. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3842–3849, 2014.

Pan Zhou, Canyi Lu, Zhouchen Lin, and Chao Zhang. Tensor factorization for low-rank tensor completion. *IEEE Transactions on Image Processing*, 27(3):1152–1163, 2017.

Xiaowei Zhou, Can Yang, and Weichuan Yu. Moving object detection by detecting contiguous outliers in the low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(3):597–610, 2012.