applied sciences

MDPI

Article

Tensor Robust Principal Component Analysis via Non-Convex Low Rank Approximation

Shuting Cai [1], Qilun Luo [2], Ming Yang [3,*], Wen Li [4] and Mingqing Xiao [2,*]

[1] School of Automation, Guangdong University of Technology, Guangzhou 510006, China; shutingcai@gdut.edu.cn
[2] Department of Mathematics, Southern Illinois University, Carbondale, IL 62901, USA; qilun.luo@siu.edu
[3] Department of Computer Science, Southern Illinois University, Carbondale, IL 62901, USA
[4] School of Mathematical Sciences, South China Normal University, Guangzhou 510631, China; liwen@scnu.edu.cn
* Correspondence: yangmingmath@gmail.com (M.Y.); mxiao@siu.edu (M.X.)

Received: 25 January 2019; Accepted: 29 March 2019; Published: 3 April 2019

Abstract: Tensor Robust Principal Component Analysis (TRPCA) plays a critical role in handling high multi-dimensional data sets, aiming to recover the low-rank and sparse components both accurately and efficiently. In this paper, different from current approach, we developed a new t-Gamma tensor quasi-norm as a non-convex regularization to approximate the low-rank component. Compared to various convex regularization, this new configuration not only can better capture the tensor rank but also provides a simplified approach. An optimization process is conducted via tensor singular decomposition and an efficient augmented Lagrange multiplier algorithm is established. Extensive experimental results demonstrate that our new approach outperforms current state-of-the-art algorithms in terms of accuracy and efficiency.

Keywords: tensor robust principal component analysis; t-SVD; non-convex tensor norm

1. Introduction

An array of numbers arranged on a regular grid with a variable number of axes is described as a tensor. As the dimensionality reaches beyond 3D, conventional data representations such as vector-based and matrix-based become insufficient, and highly dimensional data sets are usually formulated as tensors [1–3].

Robust Principal Component Analysis (RPCA) [4–8], aiming to recover the low-rank and sparse matrices both accurately and efficiently, has been widely studied in data compressed sensing and computer vision. It is of great use when we try to solve the problem of subspace learning or Principal Component Analysis (PCA) in the presence of outliers [9]. Especially RPCA is a very useful tool in foreground detection, which is the first step in video surveillance system to detect moving objects [10,11]. Tensor Robust Principal Component Analysis (TRPCA) extends the RPCA from matrix to the tensor case, that is, to recover a low-rank tensor and a sparse (noise entries) tensor.

Tensor rank [12] essentially characterizes the intrinsic "dimension" or "degree of freedom" for highly dimensional data sets. In order to make the rank characteristic be more transparent, one can make use of tensor Singular Value Decomposition (t-SVD) developed recently by Kilmer and Martin [13], rendering the tensor rank approximation approachable. The key for their framework is the well-defined t-product that leads to the 3rd-order tensor space to be an inner product space so that approximation becomes more effective such as in image processing [14,15]. t-SVD allows to define tensor tubal-rank, which appears to be much more approachable such as in the study of image analysis

than previous rank definitions [16–18]. The characteristic of a low tensor tubal-rank implies that a 3rd-order tensor can be significantly compressed (Theorem 4.3 of [13]), which is particularly important in applications.

Mathematically, TRPCA based on t-SVD [19,20] aims to recover the low tubal rank component $\mathcal{L}_0$ and sparse component $\mathcal{E}_0$ from noisy observations $\mathcal{X} = \mathcal{L}_0 + \mathcal{E}_0$ of size $N_1 \times N_2 \times N_3$ by convex optimization

$$\min_{\mathcal{L},\mathcal{E}} \quad \|\mathcal{L}\|_{TNN} + \lambda\|\mathcal{E}\|_{\ell_1} \tag{1}$$

$$\text{s.t.} \quad \mathcal{X} = \mathcal{L} + \mathcal{E}. \tag{2}$$

λ is a Lagrange multiplier, $\mathcal{E}$ is sparse if the tensor ℓ_1 norm $\|\mathcal{E}\|_{\ell_1}$ is small. It is well-known that the t-SVD-based tensor nuclear norm (TNN, III.B of [21]) has been proven to be the tightest convex relaxation to ℓ_1-norm of the tensor multi-rank (Theorem 2.4.1 in [15] or Theorem 3.1 in [19]). However, TNN treats the singular values in t-SVD equally to pursue the convexity of the objective function, while the singular values associated with a practical image may emphasize its different properties and should be handled differently, in particular, when some singular values are very large. Due to the gap between the tensor rank function and its lower convex envelop, the adoption of tensor nuclear norm may lead to the approximation of the corresponding tensor tubal-rank being inaccurate. Therefore, a new development of tensor tubal-rank approximation becomes necessary to reflect these fundamental characteristics.

To make the the necessity of a new tensor tubal-rank approximation instead of TNN more transparent, we consider the the matrix RPCA [4] first. Now the goal becomes to recover a low-rank matrix L from highly corrupted observations $X = L + S$, by minimizing $\|L\|_* + \lambda\|S\|_{\ell_1}$. Here $\|L\|_*$ is the summation of the singular values of L. It is easy to see that

$$1 + \frac{\sigma_2}{\sigma_1} + \cdots + \frac{\sigma_N}{\sigma_1} = \frac{1}{\sigma_1}\|L\|_* \leq \text{rank}(L).$$

Hence, when σ_1 is very large, for example, and the rest are relatively small, we may have $\frac{1}{\sigma_1}\|L\|_* << \text{rank}(L)$. Then the minimization of $\|L\|_*$ will attempt to minimize σ_1 that leads to over-penalizes σ_1, and consequently it generates undesirable solution L with small σ_1. Recall that our goal is to find a low-rank L, whose largest singular value is not necessarily to be small. In addition, it is worth mentioning that in practice the matrix RPCA problem is very sensitive to small perturbations and this problem can be addressed by relaxing the equality constraints. The authors in [22,23] propose an alternative optimization approach that appears to be suitable to deal with robustness issues in the "Sparse Plus Low-rank" decomposition problem. It has long been noticed that the nuclear norm is not always satisfactory [24], since it over-penalizes large singular values, and consequently may only find a much biased solution. Since the the nuclear norm is the tightest convex relaxation of the rank function [25], a further improvement under the convex setting in general is limited, if it is still possible, in particular when their gap is not very small. In order to alleviate the above problem, the Gamma quasi-norm for *matrices* is introduced in [26] in RPCA, which results in a tighter approximation of the rank function. For $L \in \mathbb{R}^{m\times n}$ and the singular values of L, denoted by $\sigma_1(L), \ldots, \sigma_{\min\{m,n\}}(L)$, the Gamma quasi-norm of L is defined as

$$||L||_\gamma = \sum_{i=1}^{\min\{m,n\}} \frac{(1+\gamma)\sigma_i(L)}{\gamma + \sigma_i(L)}, \gamma > 0.$$

It is not difficult to see

$$\frac{(1+\gamma)\sigma_i(L)}{\gamma + \sigma_i(L)} < \frac{1+\gamma}{\gamma + M}\sigma_i(L),$$

when $\sigma_i(L) > M > 0$ and $||L||_\gamma \leq (1+\gamma)\,\mathrm{rank}(L)$. So it overcomes the imbalanced penalization by different singular values in convex nuclear norm. In addition, when γ goes to 0, the limit of the Gamma quasi-norm is exactly the matrix rank. The Gamma quasi-norm has nice algebraic properties, i.e., unitarily invariant and positive definiteness. Thanks to the unitarily invariance [27], the Gamma quasi-norm minimization problem for $A \in \mathbb{R}^{m\times n}$:

$$\min_{Z\in\mathbb{R}^{m\times n}} ||Z||_\gamma + \frac{\mu}{2}\|Z-A\|_F^2$$

can be solved very effectively (Theorem 1 of [26]). Moreover, experimental results in [26] show that Gamma quasi-norma-based RPCA outperforms several other convex/non-convex norma-based RPCA in both accuracy and efficiency.

In this paper, inspired by the success of the non-convex rank approximation [26], we proposed tensor Gamma quasi-norm via t-SVD to seek a closer approximation and to alleviate the above-mentioned limitations resulted from the tensor nuclear norm, which has not been used for TRPCA yet in literature.

The main contributions of this paper are illustrated as follows:

1. We develop new non-convex tensor rank approximation functions that can be used as desirable regularization for the optimization process, which appears to be more effective than current existing approaches in literature.
2. An efficient algorithm based on the alternating direction method of multipliers (ADMM) is developed to solve the equivalent optimization problem in the Fourier domain, which is also suitable for general TRPCA problem. In addition, we provided the convergent analysis of the augmented Lagrange multiplier based optimization algorithm.
3. Extensive evaluation of the proposed approach on several benchmark data sets has been conducted, and detailed comparison with most recent approaches are provided. Experimental results demonstrate that our proposed approach yields superior performance for image recovery and video background modeling.

The rest of the paper is organized as follows. In Section 2, we roughly describe some related works. Then Section 3 provides factorization strategies for 3rd-order tensors and its relationship to the corresponding tensor approximation problem. And then we present the main results on t-Gamma quasi-norm of 3rd-order tensors. Section 4 introduces our non-convex TRPCA framework and the convergent analysis of the augmented Lagrange multiplier based optimization algorithm. Experimental analysis and results are shown in Section 5 to verify the proposed algorithm. Finally, the paper ends with concluding remarks in Section 6.

2. Related Works

Recently, many authors observed the limitations of t-SVD-based TRPCA [15,19]. It was pointed out in [28] that the classical TRPCA methods based on t-SVD fails to utilize the low rank structure in the third mode. To further exploit the low-rank structures in multiway data, they defined a new TNN that extends TNN with core matrix and proposed a creative algorithm to deal with TRPCA problems. In addition, the authors of [29] presented that there are some weak points in the existing low-rank approximation approaches, for example, people need to predefine the rank values or fail to consider local information of data (e.g., spatial or spectral smooth structure). So they proposed a novel TNN-based low-rank approximation with total variation regularization, which uses the TNN to encode the global low-rank prior of tensor data and the total variation regularization to preserve the spatial-spectral continuity in a unified framework.

It has been noticed by others that TNN is not always satisfactory. For instance, to seek a closer approximation of tensor rank, Kong et al. in [30] introduced t-Schatten-p tensor quasi-norm to replace the TNN. The t-Schatten-p norm is convex when $p \geq 1$, but non-convex when $0 < p < 1$. In addition, they proved multi-Schatten-p norm surrogate theorem to convert the non-convex

t-Schatten-p tensor quasi-norm for $0 < p < 1$ into a weighted sum of convex functions. Then based on the multi-Schatten-p norm surrogate, they proposed a novel convex tensor RPCA model.

Besides above, a recent paper [31] developed randomized algorithms for tensor low-rank approximation and decomposition, together with tensor multiplication. Their proposed tubal focused algorithms employ a small number of lateral and/or horizontal slices of the underlying 3rd order tensor, that come with relative error guarantees for the quality of the obtained solutions.

Besides the above-mentioned methods based on t-SVD, there are many other TRPCA methods mainly applied CP (CANDECOMP/PARAFAC) or Tucker decompositions methods in their optimization models. We need to mention that recently Driggs et al. [32] studied TRPCA using a non-convex formulation of the CP based tensor atomic-norm and identified a class of local minima of this non-convex program that are globally optimal.

Presently, background subtraction become more and more important for visual surveillance systems and several subspace learning algorithms based on matrix and tensor tools have been used to perform the background modeling of the scenes. Recently, Sobral et al. [33] developed an incremental tensor subspace learning that uses only a small part of the entire data and updates the low-rank model incrementally when new data arrives and in [34] the authors proposed an online stochastic framework for tensor decomposition of multispectral video sequences. In particular, it is worth mentioning that Bouwmans et al. [35] gave a systematic and complete review of the "low-rank plus sparse matrix decomposition" framework of separating moving objects from the background.

However, all those above mentioned t-SVD-based TRPCA algorithms solve the TRPCA optimization problem under convex regularization. In addition, most analyses of ADMM algorithms are in the convex setting [36]. In our work, we explore the TRPCA method based on the non-convex t-Gamma quasi-norm of 3rd-order tensors. Due to the non-convex nature of our TRPCA model, the convergent analysis of algorithm is difficult to justify. In addition, the analysis of non-convex augmented Lagrangians is not easy, but this compels us to make efforts of clarity and pedagogy. Also it should be noted that there are few works on t-SVD-based TRPCA achieve background modeling. Thus experiments are provided in our paper on two computer vision applications, which are image recovery and video background modeling.

3. Notations and Definitions

Throughout this paper, all boldface Euler script letters ($\mathcal{X}$) denote tensors, uppercase characters (X) denote matrices, and lowercase letters (a) denote scalars. The field of real numbers is denoted as $\mathbb{R}$. We denote $N_i, i = 1, 2, 3$ as the index set associated with a given tensor. For 3rd-order tensors $\mathcal{X} = [a_{ijk}] \in \mathbb{R}^{N_1 \times N_2 \times N_3}$, we use $\mathcal{X}(i, :, :)$, $\mathcal{X}(:, i, :)$, $\mathcal{X}(:, :, i)$ to denote the i-th horizontal, lateral, and frontal slice, respectively. For simplicity, we denote the frontal slice $\mathcal{X}(:, :, i)$ as X_i. The inner product between $\mathcal{X}$ and $\mathcal{Y}$ of size $N_1 \times N_2 \times N_3$ is defined as $\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{i=1}^{N_3} \mathrm{tr}(X_i^T Y_i)$. The Frobenius norm of a tensor is defined as $\|\mathcal{X}\|_F = \sqrt{\sum_{i,j,k} \left| a_{ijk} \right|^2}$. The ℓ_1 norm of a tensor is defined as $\|\mathcal{X}\|_{\ell_1} = \sum_{i,j,k} \left| a_{ijk} \right|$. In this paper, the adopted basic definitions and notations follow the references [13,21,37] for the purpose of an easier comparison.

Definition 1 (Transpose p. 10 of [13]). *For $\mathcal{X} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$, the transpose tensor $\mathcal{X}^{\mathrm{T}}$ is an $N_2 \times N_1 \times N_3$ tensor obtained by transposing each frontal slice of $\mathcal{X}$ and then reversing the order of the transposed frontal slices 2 through N_3.*

Similar to matrices whose elements can be grouped by rows or by columns, higher dimensional tensors can be viewed forming by slices along the third mode. For $\mathcal{X} = [X_1 | \cdots | X_{N_3}] \in \mathbb{R}^{N_1 \times N_2 \times N_3}$ with $N_1 \times N_2$ frontal slices $X_1, \ldots, X_{N_3}$. Then the block circulant is created from the frontal slices in a form of

$$\text{bcirc}(\mathcal{X}) = \begin{bmatrix} X_1 & X_{N_3} & X_{N_3-1} & \cdots & X_2 \\ X_2 & X_1 & X_{N_3} & \cdots & X_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{N_3} & X_{N_3-1} & \cdots & X_2 & X_1 \end{bmatrix},$$

where $X_i = \mathcal{X}(:,:,i)$ for $i = 1, \ldots, N_3$. It is important to notice that the block circulant matrix $\text{bcirc}(\mathcal{X})$ keeps the order of frontal slices of $\mathcal{X}$ in an appropriate way, and thus it better maintains $\mathcal{X}$'s structure in terms of the frontal direction, compared with the direct tensor unfolding along the third mode.

Definition 2 (t-product p. 11 of [13]). *Let $\mathcal{X} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$, we define the matvec and fold operators as $matvec(\mathcal{X}) = [X_1; X_2; \cdots; X_{N_3}]$ is an $N_1 N_3 \times N_2$ matrix and $fold(matvec(\mathcal{X})) = \mathcal{X}$.*

Let $\mathcal{A} \in \mathbb{R}^{N_1 \times N_4 \times N_3}$, and $\mathcal{B} \in \mathbb{R}^{N_4 \times N_2 \times N_3}$. Then the t-product of 3rd-order tensors is defined as

$$\mathcal{A} * \mathcal{B} = fold(bcirc(\mathcal{A}) matvec(\mathcal{B}))$$

which is a tensor of size $N_1 \times N_2 \times N_3$.

Definition 3 (Identity, orthogonal and diagonal tensor p.10 of [13]). *The identity tensor $\mathcal{I} \in \mathbb{R}^{N_1 \times N_1 \times N_3}$ is a tensor whose first frontal slice is the $N_1 \times N_1$ identity matrix and all other frontal slices are zero. A tensor $\mathcal{Q} \in \mathbb{R}^{N_1 \times N_1 \times N_3}$ is orthogonal if $\mathcal{Q}^{\mathrm{T}} * \mathcal{Q} = \mathcal{Q} * \mathcal{Q}^{\mathrm{T}} = \mathcal{I}$, where $*$ is the t-product. A tensor is called f-diagonal if each of its frontal slices is diagonal matrix.*

It is well known that circulant matrices can be diagonalized by the FFT (21.11 of [38]), similarly, block-circulant matrices can also block-diagonalized as

$$\begin{aligned} &\left(\widetilde{\mathbf{F}}_{N_3} \otimes \mathbf{I}_{N_1}\right) \cdot \text{bcirc}(\mathcal{X}) \cdot \left(\widetilde{\mathbf{F}}^*_{N_3} \otimes \mathbf{I}_{N_2}\right) \\ =&\text{bdiag}(\hat{\mathcal{X}}) = \begin{bmatrix} \hat{X}_1 & & \\ & \ddots & \\ & & \hat{X}_{N_3} \end{bmatrix}, \end{aligned}$$

where $\widetilde{\mathbf{F}}_{N_3} = \frac{1}{\sqrt{N_3}} \mathbf{F}_{N_3}$, $\mathbf{F}_{N_3}$ denotes the $N_3 \times N_3$ fast Fourier transform matrix, $\mathbf{F}^*_{N_3}$ denotes its conjugate, and the notation $\otimes$ denotes the standard Kronecker product. In addition, $\mathcal{X} = \mathcal{A} * \mathcal{B}$ if and only if $\hat{X}_{n_3} = \hat{A}_{n_3} \hat{B}_{n_3}$ for $1 \le n_3 \le N_3$ (p.4 of [19]).

Since $\widetilde{\mathbf{F}}_{N_3} \otimes \mathbf{I}_{N_1}, \widetilde{\mathbf{F}}_{N_3} \otimes \mathbf{I}_{N_2}$ are unitary matrices, we have

$$\| \text{ bdiag}(\hat{\mathcal{X}}) \|_F^2 = \text{tr bcirc}(\mathcal{X})^T \text{bcirc}(\mathcal{X}) = N_3 ||\mathcal{X}||_F^2.$$

Then we have $\| \mathcal{X} \|_F = \frac{1}{\sqrt{N_3}} \| \text{ bdiag}(\hat{\mathcal{X}}) \|_F$.

Definition 4 (Tensor multi-rank p. 10 of [37]). *For tensor $\mathcal{X} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$ with $N_1 \times N_2$ frontal slices $X_1, \ldots, X_{N_3}$, the multi-rank of $\mathcal{X}$ is a N_3-vector $[rank(\hat{X}_1), \ldots, rank(\hat{X}_{N_3})]$, consisting of the ranks of all the $\hat{X}_1, \ldots, \hat{X}_{N_3}$.*

Please note that the matrix SVD [39] can be performed on each frontal slice of $\hat{\mathcal{X}}$, The matrix $\hat{X}_{n_3}$ has the following Singular Value Decomposition (SVD) $\hat{X}_{n_3} = \hat{U}_{n_3} \hat{\Sigma}_{n_3} \hat{V}^{\mathrm{T}}_{n_3}$, $\hat{U}_{n_3} \in \mathbb{O}(N_1), \hat{V}_{n_3} \in \mathbb{O}(N_2)$ (here $\mathbb{O}(N_i)$ stands for the set of orthogonal N_i-by-N_i matrices) and $\hat{\Sigma}_{n_3}$ is an $N_1 \times N_2$ diagonal matrix with diagonal entries $\sigma_1(\hat{X}_{n_3}), \ldots, \sigma_{\min\{N_1, N_2\}}(\hat{X}_{n_3})$ in descending order.

Definition 5 (t-SVD, Theorem 4.1 in [13]). *The tensor Singular Value Decomposition (t-SVD) of $\mathcal{X}$ is given by $\mathcal{X} = \mathcal{U} * \Sigma * \mathcal{V}^{\mathrm{T}}$, where $\mathcal{U}$ and $\mathcal{V}$ are orthogonal tensors of size $N_1 \times N_1 \times N_3$ and $N_2 \times N_2 \times N_3$ respectively. Σ is an f-diagonal tensor of size $N_1 \times N_2 \times N_3$, and $*$ denotes the t-product.*

Definition 6 (Tensor tubal rank III.B of [21]). *The tensor tubal rank is defined as the number of nonzero singular tubes of f-diagonal tensor Σ, which is the maximum of the number of nonzero $\{\hat{\Sigma}_{n,n,n_3},\ 1 \leq n \leq \min(N_1, N_2)\}$ for $1 \leq n_3 \leq N_3$.*

Definition 7 (Tensor nuclear norm III.B of [21]). *The nuclear norm for a given 3rd-order tensor $\mathcal{X} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$ is defined to be*

$$\| \mathcal{X} \|_{TNN} = \frac{1}{N_3} \sum_{n_3=1}^{N_3} ||\hat{X}_{n_3}||_*,$$

which is equal to

$$\sum_{n=1}^{\min(N_1,N_2)} \sum_{n_3=1}^{N_3} \frac{1}{N_3} \hat{\Sigma}_{n,n,n_3}.$$

Here $\hat{X}_{n_3}$ is the n_3-th frontal slice of $\hat{\mathcal{X}} = \mathrm{fft}(\mathcal{X}, [\,], 3) = [\hat{X}_1 | \cdots | \hat{X}_{N_3}] \in \mathbb{R}^{N_1 \times N_2 \times N_3}$ obtained by applying FFT on $\hat{\mathcal{X}}$ along the third mode.

In this paper, to overcome the disadvantages of the tensor nuclear norm, we propose smooth but more effective, though being non-convex, t-SVD-based non-convex rank approximation to the data sets represented by 3rd-order tensor. This will make the larger singular values be shrunk less in order to preserve their characteristics. Moreover, under our framework, we establish explicit solutions to the corresponding non-convex tensor norm minimization problems that is not available in current literature.

Definition 8 (Gamma quasi-norm for matrix [26]). *Let $X \in \mathbb{R}^{m \times n}$ and the singular values of X, denoted by $\sigma_1(X), \ldots, \sigma_{\min\{m,n\}}(X)$. We could introduce the Gamma norm of X, which is defined as $||X||_\gamma = \sum_{i=1}^{\min m,n} \frac{(1+\gamma)\sigma_i(X)}{\gamma + \sigma_i(X)}, \gamma > 0$.*

Definition 9 (t-Gamma tensor quasi-norm). *The Gamma tensor quasi-norm for a given 3rd-order tensor $\mathcal{X} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$ is defined to be*

$$||\mathcal{X}||_{t\text{-}\gamma} = \frac{1}{N_3} \sum_{n_3=1}^{N_3} ||\hat{X}_{n_3}||_\gamma,$$

which is equal to

$$\frac{1}{N_3} \sum_{n_3=1}^{N_3} \left(\sum_{n=1}^{\min\{N_1,N_2\}} \frac{(1+\gamma)\hat{\Sigma}_{n,n,n_3}}{\gamma + \hat{\Sigma}_{n,n,n_3}} \right).$$

Here $\hat{X}_{n_3}$ is the n_3-th frontal slice of $\hat{\mathcal{X}} = \mathrm{fft}(\mathcal{X}, [\,], 3) = [\hat{X}_1 | \cdots | \hat{X}_{N_3}] \in \mathbb{R}^{N_1 \times N_2 \times N_3}$ obtained by applying FFT on $\hat{\mathcal{X}}$ along the third mode.

Remark 1. *It is not hard to see the t-Gamma tensor quasi-norm is unitarily invariant, that is $||\mathcal{X}||_{t\text{-}\gamma} = ||\mathcal{U} * \mathcal{X} * \mathcal{V}^T||_{t\text{-}\gamma}$, where $\mathcal{U}$ and $\mathcal{V}$ are orthogonal tensors of size $N_1 \times N_1 \times N_3$ and $N_2 \times N_2 \times N_3$ respectively. In addition,*

$$\lim_{\gamma \to 0} ||\mathcal{X}||_{t\text{-}\gamma} = \text{tubal rank of } \mathcal{X}; \ \lim_{\gamma \to \infty} ||\mathcal{X}||_{t\text{-}\gamma} = ||\mathcal{X}||_{TNN}.$$

Lemma 1 (Theorem 1 of [26]). *Let $A = U\Sigma_A V^T$ be the SVD of $A \in \mathbb{R}^{m\times n}$ and $\Sigma_A = diag(\sigma_A)$ are singular values of A. Let $F(Z) = f \circ \sigma(Z)$ be a unitary invariant function and $\mu > 0$. Then an optimal solution to the following problem*

$$\min_Z F(Z) + \frac{\mu}{2}\|Z - A\|_F^2, \tag{3}$$

is $Z^* = U\Sigma_Z^* V^{\mathrm{T}}$, where $\Sigma_Z^* = diag(\sigma^*)$ and $\sigma^* = \mathrm{prox}_{f,\mu}(\sigma_A)$. Here $\mathrm{prox}_{f,\mu}(\sigma_A)$ is the Moreau-Yosida operator [40], defined as

$$\mathrm{prox}_{f,\mu}(\sigma_A) := \operatorname*{argmin}_{\sigma\geq 0} f(\sigma) + \frac{\mu}{2}\|\sigma - \sigma_A\|_2^2. \tag{4}$$

Now let us consider the tensor t-Gamma quasi-norm minimization problem:

Theorem 1 (t-Gamma tensor rank minimization). *Let us consider a 3rd-order tensor $\mathcal{X} \in \mathbb{R}^{N_1\times N_2\times N_3}$, which has t-SVD $\mathcal{X} = \mathcal{U} * \Sigma * \mathcal{V}^{\mathrm{T}}$ as in Definition 5. If $\mathcal{Z}^*$ minimizes*

$$\frac{1}{2}\|\mathcal{X} - \mathcal{Z}\|_F^2 + \tau \parallel \mathcal{Z} \parallel_{t\text{-}\gamma},$$

then

$$\mathcal{Z}^* = \mathcal{U} * \Omega_{\mathcal{Z}} * \mathcal{V}^{\mathrm{T}},$$

where $\Omega_{\mathcal{Z}} \in \mathbb{R}^{N_1\times N_2\times N_3}$ is a f-diagonal tensor. For $1 \leq n \leq \min(N_1, N_2), 1 \leq n_3 \leq N_3$, let $\hat{\Sigma} = [\hat{\Sigma}_{n,n,n_3}] = \mathrm{fft}(\Sigma, [\,], 3)$, $\hat{\Omega}_{\mathcal{Z}} = [\hat{\Omega}_{n,n,n_3}] = \mathrm{fft}(\Omega, [\,], 3)$, then $\hat{\Omega}_{n,n,n_3}$ is the limit point of Fixed Point Iteration

$$\hat{\Omega}_{n,n,n_3}^{(k+1)} = (\hat{\Sigma}_{n,n,n_3} - \tau\frac{\gamma(1+\gamma)}{(\gamma + \hat{\Omega}_{n,n,n_3}^{(k)})^2})_+.$$

Remark 2. *Theorem 1 provides the connection of the optimal solution between the original domain and the Fourier domain. More importantly, it shows that the minimization problem under our framework is tractable. After the original tensor $\mathcal{X}$ is transformed to $\hat{\mathcal{X}}$ in the Fourier domain, an equivalent minimization problem could be solved by using Lemma 1 for each slice $\hat{X}_{n_3}, 1 \leq n_3 \leq N_3$ in the Fourier domain. Then the optimal solution $\mathcal{Z}^*$ in the original domain can be obtained by the inverse Fourier transform $\hat{\mathcal{Z}}^* \to \mathcal{Z}^*$, whose detailed approach can be found from the proof in Theorem 1.*

Proof of Theorem 1. Let $f(\mathcal{Z}) = \frac{1}{2}\|\mathcal{X} - \mathcal{Z}\|_F^2 + \tau||\mathcal{Z}||_{t\text{-}\gamma}$, then it can be reformulated as matrix objective function in Fourier domain:

$$\hat{f}(\hat{\mathcal{Z}}) = \frac{1}{2}\|\mathrm{bdiag}(\hat{\mathcal{X}} - \hat{\mathcal{Z}})\|_F^2 + \tau \parallel \mathrm{bdiag}(\hat{\mathcal{Z}}) \parallel_\gamma = \sum_{n_3=1}^{N_3} \frac{1}{2}\|\hat{X}_{n_3} - \hat{Z}_{n_3}\|_F^2 + \tau \parallel \hat{Z}_{n_3} \parallel_\gamma .$$

Without loss of generality, for each diagonal block $\hat{Z}_{n_3}$ of $\hat{\mathcal{Z}}$, $1 \leq n_3 \leq N_3$, let us consider

$$\hat{Z}_{n_3}^* = \arg\min_{\hat{Z}_{n_3}} \frac{1}{2}\|\hat{X}_{n_3} - \hat{Z}_{n_3}\|_F^2 + \tau \parallel \hat{Z}_{n_3} \parallel_\gamma .$$

From Lemma 1, $\hat{Z}_{n_3}^* = \hat{U}_{\mathcal{X},n_3}\hat{\Omega}_{\mathcal{Z},n_3}\hat{V}_{\mathcal{X},n_3}^{\mathrm{T}}$, and

$$\hat{\Omega}_{n,n,n_3} = \arg\min_{\omega>0} \frac{1}{2}(\omega - \hat{\Sigma}_{n,n,n_3})^2 + \tau\frac{(1+\gamma)\omega}{\gamma+\omega}.$$

Following the DC programming algorithm used in Theorem 1 of [26] (similar method has been used in Section III of [41]), we can solve it efficiently using a local minimization method.

Let $f(\omega) = \frac{(1+\gamma)\omega}{\gamma+\omega}$, it can be approximated using first order Taylor expansion (with Big "O" truncation error) as

$$f(\omega) = f(\hat{\Omega}^{(k)}_{n,n,n_3}) + \frac{\gamma(1+\gamma)}{(\gamma+\hat{\Omega}^{(k)}_{n,n,n_3})^2}(\omega - \hat{\Omega}^{(k)}_{n,n,n_3}) + O((\omega - \hat{\Omega}^{(k)}_{n,n,n_3})^2).$$

Now we consider

$$\hat{\Omega}^{(k+1)}_{n,n,n_3} = \arg\min_{\omega>0} \frac{1}{2}(\omega - \hat{\Sigma}_{n,n,n_3})^2 + \tau\frac{\gamma(1+\gamma)}{(\gamma+\hat{\Omega}^{(k)}_{n,n,n_3})^2}|\omega|,$$

and it admits a closed-form solution

$$\hat{\Omega}^{(k+1)}_{n,n,n_3} = (\hat{\Sigma}_{n,n,n_3} - \tau\frac{\gamma(1+\gamma)}{(\gamma+\hat{\Omega}^{(k)}_{n,n,n_3})^2})_+,$$

where $(x)_+ = \max\{x, 0\}$. Here $\hat{\Omega}^{(k+1)}_{n,n,n_3}$ is the solution obtained in the $(k+1)$-th iteration. Let

$$g(\hat{\Omega}_{n,n,n_3}) = (\hat{\Sigma}_{n,n,n_3} - \tau\frac{\gamma(1+\gamma)}{(\gamma+\hat{\Omega}_{n,n,n_3})^2})_+,\ 0 \le \hat{\Omega}_{n,n,n_3} < \hat{\Sigma}_{n,n,n_3},$$

it can be easily seen that $|g'(\hat{\Omega}_{n,n,n_3})| < 1$, when $2\tau(1+\gamma) < \gamma^2$. The Fixed Point Iteration $\hat{\Omega}^{(k+1)}_{n,n,n_3} = g(\hat{\Omega}^{(k)}_{n,n,n_3})$ converge to the local minimum $\hat{\Omega}_{n,n,n_3}$ after several iterations (Chapter 1 of [42]), and the proof is thus completed. □

4. Tensor Robust Principal Component Analysis with Non-Convex Regularization

A TRPCA model using non-convex regularization can be formulated as

$$\begin{aligned} \min_{\mathcal{L},\mathcal{E}} \quad & \|\mathcal{L}\|_{t\text{-}\gamma} + \lambda\|\mathcal{E}\|_{\ell_1} \\ \text{s.t.} \quad & \mathcal{X} = \mathcal{L} + \mathcal{E}. \end{aligned} \tag{5}$$

Its augmented Lagrangian function is $L_\beta(\mathcal{L}, \mathcal{E}, \mathcal{M}) = \|\mathcal{L}\|_{t\text{-}\gamma} + \lambda\|\mathcal{E}\|_{\ell_1} + \frac{\beta}{2}\|\mathcal{X} - \mathcal{L} - \mathcal{E} - \frac{\mathcal{M}}{\beta}\|_F^2 + \mathcal{C}$, where $\mathcal{M}$ is the Lagrangian multiplier, β is the penalty parameter for the violation of the linear constraints, and $\mathcal{C}$ is a constant.

Then, the problem $\arg\min_{\mathcal{L},\mathcal{E},\mathcal{M}} L_\beta(\mathcal{L}, \mathcal{E}, \mathcal{M})$ can be updated as:

$$\begin{cases} \mathcal{L}_{k+1} = \arg\min_{\mathcal{L}} \|\mathcal{X} - \mathcal{E}_k - \frac{\mathcal{M}_k}{\beta} - \mathcal{L}\|_F^2 + \frac{2}{\beta}\|\mathcal{L}\|_{t\text{-}\gamma}, \\ \mathcal{E}_{k+1} = \mathcal{S}_{\frac{\lambda}{\beta}}\left(\mathcal{X} - \mathcal{L}_{k+1} + \frac{\mathcal{M}^k}{\beta}\right), \\ \mathcal{M}_{k+1} = \mathcal{M}_k + \beta(\mathcal{X} - \mathcal{L}_{k+1} - \mathcal{E}_{k+1}), \end{cases} \tag{6}$$

where the tensor non-negative **soft-thresholding operator** $\mathcal{E}_v(\cdot)$ [43] is defined as

$$\mathcal{S}_v(\mathcal{B}) = \bar{\mathcal{B}}$$

with

$$\bar{b}_{n_1n_2n_3} = \begin{cases} sgn(b_{n_1n_2n_3})(|b_{n_1n_2n_3}| - v), & |b_{n_1n_2n_3}| > v, \\ 0, & |b_{n_1n_2n_3}| \le v. \end{cases}$$

Algorithm 1 shows the pseudocode for the proposed non-convex tensor robust component analysis method.

Algorithm 1 Solve the non-convex TRPCA model (5) by ADMM.

Input: The observed tensor $\mathcal{X} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$, the set of index of observed entries Ω, parameter λ, stopping criterion ϵ.

Initialize: $\mathcal{L}^0_{\omega>0} = \mathcal{X}_{\omega>0}$, $\mathcal{L}^0_{\Omega^C} = \text{rand}(N_1 \times N_2 \times N_3)_{\Omega^C}$, $\mathcal{E}^0 = \mathcal{M}^0 = \text{zeros}(N_1 \times N_2 \times N_3)$, $\beta = 10^{-4}$, $\beta_{max} = 10^{10}$, $\eta = 1.2$.

1: **while** not converged **do**

2: $\mathcal{L}_{k+1} \leftarrow \mathcal{U}_{\mathcal{X}-\mathcal{E}_k-\frac{\mathcal{M}_k}{\beta_k}} * \Omega_{\mathcal{L}} * \mathcal{V}^{\mathrm{T}}_{\mathcal{X}-\mathcal{E}_k-\frac{\mathcal{M}_k}{\beta_k}}$ using the local minimum in Theorem 1 with $\tau = \frac{1}{\beta_k}$

3: $\mathcal{E}_{k+1} \leftarrow \mathcal{S}_{\frac{\lambda}{\beta_k}}\left(\mathcal{X} - \mathcal{L}_{k+1} + \frac{\mathcal{M}_k}{\beta_k}\right)$

4: $\mathcal{M}_{k+1} \leftarrow \mathcal{M}_k + \beta_k(\mathcal{X} - \mathcal{L}_{k+1} - \mathcal{E}_{k+1})$

5: $\beta_{k+1} \leftarrow \min\{\beta_k\eta, \beta_{max}\}$

6: Check the convergence conditions $\|\mathcal{L}_{k+1} - \mathcal{L}_k\|_\infty \leq \epsilon$, $\|\mathcal{E}_{k+1} - \mathcal{E}_k\|_\infty \leq \epsilon$, $\|\mathcal{L}_{k+1} + \mathcal{E}_{k+1} - \mathcal{X}\|_\infty \leq \epsilon$

7: **end while**

Output: The low rank tensor $\mathcal{L}$ and the sparse tensor $\mathcal{E}$

Theorem 2 (Convergent Analysis). *Suppose that $\mathcal{P}_k = \{\mathcal{L}_k, \mathcal{E}_k, \mathcal{M}_k\}$ is generated from the proposed Algorithm 1, then $\{\mathcal{P}_k\}$ is bounded. In addition, the accumulation point $\{\mathcal{L}^*, \mathcal{E}^*, \mathcal{M}^*\}$ of $\{\mathcal{P}_k\}_{k=1}^{+\infty}$ is a KKT stationary point, which satisfy the KKT conditions $\mathcal{M}^* \in \partial||\mathcal{L}||_{t\text{-}\gamma}$, $\mathcal{L}^* + \mathcal{E}^* = \mathcal{X}$, $\mathcal{M}^* \in \partial||\mathcal{E}||_{\ell_1}$.*

Proof of Theorem 2. $\mathcal{E}_{k+1}$ satisfies the first-order necessary local optimality condition,

$$\begin{aligned} &0 \in \partial_{\mathcal{E}} L_{\beta_k}\left(\mathcal{L}_{k+1}, \mathcal{E}, \mathcal{M}_k\right)|_{\mathcal{E}_{k+1}} \\ =&\partial\left(\lambda\|\mathcal{E}\|_{\ell_1}\right)|_{\mathcal{E}_{k+1}} + \mathcal{M}_k + \beta_k\left(\mathcal{L}_{k+1} - \mathcal{X} + \mathcal{E}_{k+1}\right) \\ =&\partial\left(\lambda\|\mathcal{E}\|_{\ell_1}\right)|_{\mathcal{E}_{k+1}} + \mathcal{M}_{k+1}. \end{aligned} \tag{7}$$

For $\|\mathcal{E}\|_{\ell_1}$ case, since $||\mathcal{E}||_{\ell_1}$ is nonsmooth at $\mathcal{E}_{ijk} = 0$, we redefine subgradient $\left[\partial\|\mathcal{E}\|_{\ell_1}\right]_{ijk} = 0$ if $\mathcal{E}_{ijk} = 0$. Then $0 \leq \|\partial\|\mathcal{E}\|_{\ell_1}\|_F^2 \leq N_1N_2N_3$, hence $\partial(\lambda\|\mathcal{E}\|_{\ell_1})|_{\mathcal{E}_{k+1}}$ is bounded. Thus $\{\mathcal{M}_k\}$ is bounded.

With some algebra, we have the following equality

$$\begin{aligned} &L_{\beta_k}\left(\mathcal{L}_k, \mathcal{E}_k, \mathcal{M}_k\right) \\ &= L_{\beta_{k-1}}\left(\mathcal{L}_k, \mathcal{E}_k, \mathcal{M}_{k-1}\right) + \frac{\beta_k - \beta_{k-1}}{2}\|\mathcal{L}_k - \mathcal{X} + \mathcal{E}_k\|_F^2 + tr[(\mathcal{M}_k - \mathcal{M}_{k-1})(\mathcal{L}_k - \mathcal{X} + \mathcal{E}_k)] \\ &= L_{\beta_{k-1}}\left(\mathcal{L}_k, \mathcal{E}_k, \mathcal{M}_{k-1}\right) + \frac{\beta_k + \beta_{k-1}}{2(\beta_{k-1})^2}\|\mathcal{M}_k - \mathcal{M}_{k-1}\|_F^2. \end{aligned}$$

Then, $L_{\beta_k}\left(\mathcal{L}_{k+1}, \mathcal{E}_{k+1}, \mathcal{M}_k\right) \leq L_{\beta_k}(\mathcal{L}_{k+1}, \mathcal{E}_k, \mathcal{M}_k) \leq L_{\beta_k}(\mathcal{L}_k, \mathcal{E}_k, \mathcal{M}_k) \leq L_{\beta_{k-1}}\left(\mathcal{L}_k, \mathcal{E}_k, \mathcal{M}_{k-1}\right) + \frac{\beta_k+\beta_{k-1}}{2(\beta_{k-1})^2}\|\mathcal{M}_k - \mathcal{M}_{k-1}\|_F^2$.

By induction, we could obtain

$$L_{\beta_k}(\mathcal{L}_{k+1}, \mathcal{E}_{k+1}, \mathcal{M}_k) \leq L_{\beta^0}\left(\mathcal{L}^1, \mathcal{E}^1, \mathcal{M}^0\right) + \sum_{i=1}^{k}\frac{\beta_i + \beta_{i-1}}{2(\beta_{i-1})^2}\|\mathcal{M}_i - \mathcal{M}_{i-1}\|_F^2.$$

Since $\frac{\beta_i+\beta_{i-1}}{2(\beta_{i-1})^2}\|\mathcal{M}_i-\mathcal{M}_{i-1}\|_F^2$ is bounded, it is not hard to see $L_{\beta_k}(\mathcal{L}_{k+1},\mathcal{E}_{k+1},\mathcal{M}_k)$ is upper bounded. Meanwhile,

$$\begin{aligned}&L_{\beta_k}(\mathcal{L}_{k+1},\mathcal{E}_{k+1},\mathcal{M}_k)+\frac{1}{2\beta_k}\|\mathcal{M}_k\|_F^2\\&=||\mathcal{L}_{k+1}||_{t\text{-}\gamma}+\lambda\|\mathcal{E}_{k+1}\|_{\ell_1}+\frac{\beta_k}{2}\|\mathcal{L}_{k+1}-\mathcal{X}+\mathcal{E}_{k+1}+\frac{\mathcal{M}_k}{\beta_k}\|_F^2.\end{aligned}$$

Since each term on the right-hand side is bounded, $\mathcal{E}_{k+1}$ is bounded. By the last term on the right-hand, $\mathcal{L}_{k+1}$ is bounded. Therefore, $\{\mathcal{L}_k\}$ and $\{\mathcal{E}_k\}$ are both bounded.

From Bolzano-Weierstrass theorem [39], there must be at least one accumulation point of the sequence $\{\mathcal{P}_k\}_{k=1}^{+\infty}$. We denote one of the points $\mathcal{P}^*=\{\mathcal{L}^*,\mathcal{E}^*,\mathcal{M}^*\}$. Without loss of generality, we assume $\{\mathcal{P}_k\}_{k=1}^{+\infty}$ converge to $\mathcal{P}^*$.

Since

$$(\mathcal{M}_{k+1}-\mathcal{M}_k)/\beta_k=\mathcal{L}_{k+1}+\mathcal{E}_{k+1}-\mathcal{X},$$

we have

$$\lim_{k\to\infty}(\mathcal{L}_{k+1}+\mathcal{E}_{k+1}-\mathcal{X})=\lim_{k\to\infty}(\mathcal{M}_{k+1}-\mathcal{M}_k)/\beta_k=0.$$

Then $\mathcal{L}^*+\mathcal{E}^*=\mathcal{X}$ is obtained.

Since $\mathcal{L}_{k+1}$ is optimally obtained by minimizing $L_{\beta_k}(\mathcal{L},\mathcal{E}_k,\mathcal{M}_k)$ according to its definition, we have

$$0\in\partial\|\mathcal{L}\|_{t\text{-}\gamma}|_{\mathcal{L}_{k+1}}+\mathcal{M}_k+\beta_k(\mathcal{L}_{k+1}+\mathcal{E}_k-\mathcal{X}).$$

From Theorem 3 of [26], we know

$$\nabla_{\hat{\mathcal{L}}_{n_3}}||\hat{\mathcal{L}}_{n_3}||_\gamma=\hat{U}_{\mathcal{L},n_3}Diag(\frac{\gamma(1+\gamma)}{(\gamma+\hat{\Omega}_{n,n,n_3})^2})\hat{V}_{\mathcal{L},n_3}^T$$

and $\frac{\gamma(1+\gamma)}{(\gamma+\hat{\Omega}_{n,n,n_3})^2}\le\frac{1+\gamma}{\gamma}\Longrightarrow\nabla_{\hat{\mathcal{L}_{n_3}}}||\hat{\mathcal{L}}_{n_3}||_\gamma$ is bounded.

Therefore,

$$\frac{\partial\|\mathcal{L}\|_{t\text{-}\gamma}}{\partial\hat{\mathcal{L}}}=[\frac{\partial||\hat{\mathcal{L}}_1||_\gamma}{\partial\hat{L}_1}|\cdots|\frac{\partial||\hat{\mathcal{L}}_{N_3}||_\gamma}{\partial\hat{L}_{N_3}}]$$

is bounded.

From the fact that $\hat{\mathcal{L}}$ is equivalent to Tucker product (for its definition see Chapter 3 of [44])

$$\hat{\mathcal{L}}=\mathcal{L}\times_3\widetilde{\mathbf{F}}_{N_3}$$

(see the proof of Proposition 3 of [30] and Remark 2.3. of [45]) and using the chain rule [39], we can deduce that

$$\nabla_{\mathcal{L}}\|\mathcal{L}\|_{t\text{-}\gamma}=\frac{\partial\|\mathcal{L}\|_{t\text{-}\gamma}}{\partial\hat{\mathcal{L}}}\times_3\widetilde{\mathbf{F}}_{N_3}^*$$

is bounded.

In addition, it is not hard to see

$$0\in\partial\|\mathcal{L}\|_{t\text{-}\gamma}|_{\mathcal{L}_{k+1}}+\mathcal{M}_{k+1}-\beta_k(\mathcal{E}_{k+1}-\mathcal{E}_k)$$

Please note that $\beta_k<\beta_{max}=10^{10}$ in Algorithm 1, then $k\to\infty$, $\beta_k(\mathcal{E}_{k+1}-\mathcal{E}_k)\to0\Longrightarrow-\mathcal{M}^*\in\partial\|\mathcal{L}^*\|_{t\text{-}\gamma}$.

Similarly, since $\mathcal{E}_{k+1}$ is the minimum of the subproblem $L_{\beta_k}(\mathcal{L}_k,\mathcal{E},\mathcal{M}_k)$, we have

$$0\in\partial||\mathcal{E}||_{\ell_1}|_{\mathcal{E}_{k+1}}+\mathcal{M}_k+\beta_k(\mathcal{L}_k+\mathcal{E}_{k+1}-\mathcal{X}).$$

Thus $\mathcal{M}_{k+1} \to \mathcal{M}^* \in \partial||\mathcal{E}^*||_{\ell_1}$. Therefore $(\mathcal{L}^*, \mathcal{E}^*, \mathcal{M}^*)$ satisfies the the Karush-Kuhn-Tuker (KKT) conditions of the Lagrange function $L_\beta(\mathcal{L}, \mathcal{E}, \mathcal{M})$ (Chapter 9 of [46]). We thus complete the proof. □

5. Experiments

In this section, we conduct several applications to demonstrate the performance of our algorithm. All experiments are implemented in Matlab 2017b on Ubuntu 16.04 LTS with 12 cores, 2.40 GHz CPU and 64 GB RAM.

5.1. Image Recovery

In this part, we focus on noisy color image recovery. A color image with size $N_1 \times N_2$ and 3 channel, i.e., red, green and blue, can be regarded as a third order tensor $\mathcal{X}^{N_1 \times N_2 \times N_3}$ and each frontal slice of $\mathcal{X}$ is the corresponding channel of the color image. Please note that each channel of the color image is usually not a low rank matrix, but it is observed that the top singular values dominate the main information [19,47]. Thus we can reconstruct the noisy image in low rank structure.

We perform the experiments on Berkeley Segmentation Dataset [48], which contains 200 color image for training. For each image, we randomly set 10% of pixels to the random values in [0, 255], that is, we corrupt the image on 3 channels simultaneously which is more challenging than make corruptions on only one channel. We compare our t-Gamma algorithms with RPCA [4], Gamma [26], SNN [49] and TRPCA [20], in which RPCA minimizes a weighted combination of the nuclear norm and of the l_1 norm in matrix form, Gamma replaces the nuclear norm by matrix Gamma quasi-norm, SNN uses the sum of the nuclear norm for all mode-i matricization of the given tensor, and TRPCA extends the RPCA from the matrix to tensor case. For RPCA and Gamma, we conduct the algorithms on each channel, i.e., regarding each channel of the image as a matrix, and then combine the channel recovery result into a color image. The parameter λ in RPCA and Gamma is set to $\lambda = \frac{1}{\sqrt{\max\{N_1, N_2\}}}$ as suggested in [4]. However, when the parameter γ in Gamma is set to 0.01 as stated in [26], it cannot perform well. We empirically set the $\gamma = \sqrt{\min\{N_1, N_2\}}$ in Gamma and t-Gamma. For SNN, we set $\lambda = [15, 15, 1.5]$ as suggested in [20]. The parameters λ in TRPCA and t-Gamma is set to $\lambda = \frac{1}{\sqrt{\max\{N_1, N_2\} N_3}}$. By using the Peak Signal to Noise Ratio (PSNR) to evaluate the recovery results and PSNR is defined as

$$PSNR = 10 \log_{10} \left(\frac{\|\mathcal{M}\|_\infty^2}{\frac{1}{N_1 N_2 N_3} \|\hat{\mathcal{X}} - \mathcal{M}\|_F^2} \right).$$

In Figure 1, we give the comparison of PSNR value on 200 images and each subfigure contains 50 results. The sample recovery performance are shown in Figure 2. From Figure 1, we can observe that the proposed t-Gamma algorithm perform better on almost every image and can achieve around 2 dB improvement in many cases. Table 1 shows the average recovery performance obtained by these five algorithms on Berkeley Segmentation Dataset. In Figure 2 we can visually conclude that our algorithm yields a better result than the compare algorithms which can illustrate the merit of the our approach.

Table 1. Average PSNR obtained by five Algorithms on 200 images.

Algorithm	RPCA	Gamma	SNN	TRPCA	t-Gamma
Average PSNR	25.5843	27.0762	27.6329	29.1157	30.7896

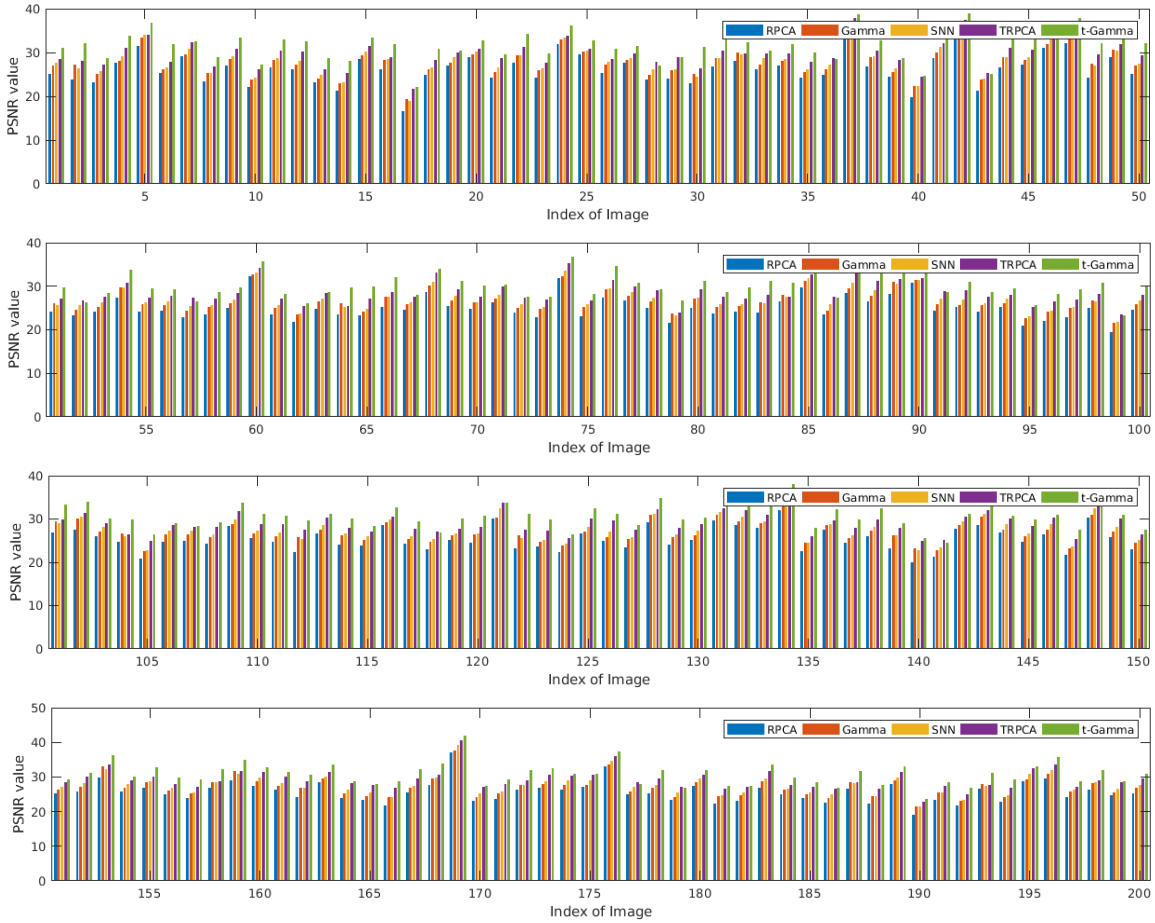
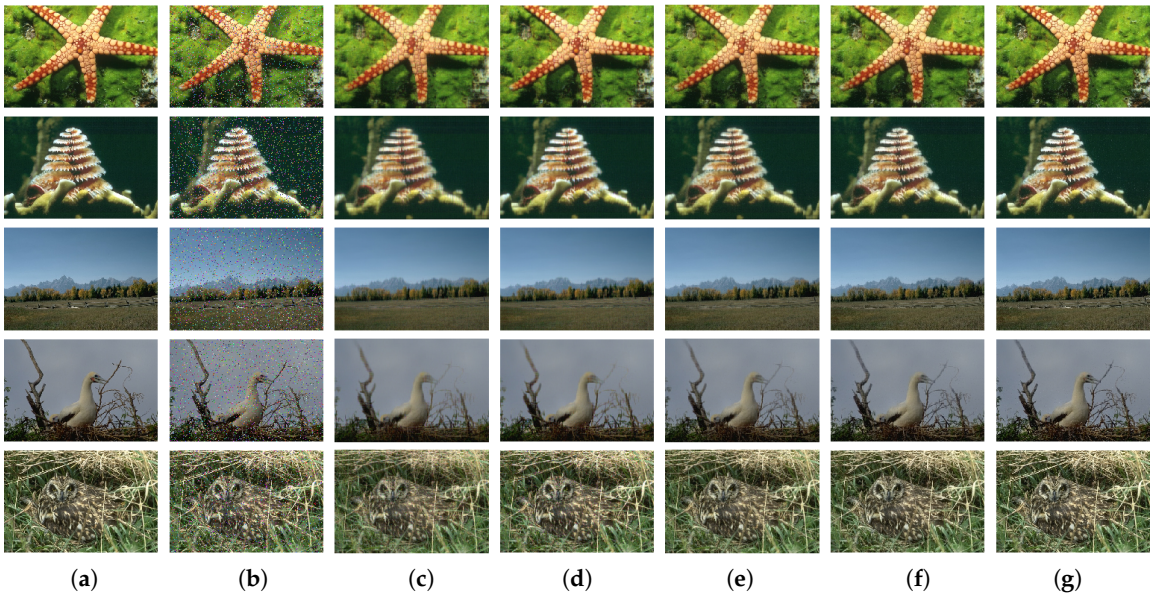


Figure 1. Comparison of the PSNR values with RPCA [4], Gamma [26], SNN [49], TRPCA [20] and t-Gamma.



(**a**) (**b**) (**c**) (**d**) (**e**) (**f**) (**g**)

Figure 2. Recovery results on 5 sample images. The first two columns are original and corrupted images reps. The third to seventh columns are recovery images obtained by RPCA, Gamma, SNN, TRPCA and t-Gamma resp. (**a**) Original; (**b**) Corrupted; (**c**) RPCA [4]; (**d**) Gamma [26]; (**e**) SNN [49]; (**f**) TRPCA [20]; (**g**) t-Gamma.

5.2. Video Background Modeling

5.2.1. Qualitative Experiments

In this part, we consider the background modeling problem, an essential task in video surveillance, which aims to separate the foreground objects from the background. The frames of background in video surveillance are highly correlated and thus it is reasonable to consider it as a low rank tensor. While the objects in the foreground usually only occupy a small part of the image in some frames and hence can regard it as the sparse part.

To solve this problem, we test our algorithm on four example color videos, *Hall*, *MovedObject*, *Escalator* and *Lobby* http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html, and compare it with RPCA [4], Gamma [26], SNN [49] and TRPCA [20]. We extract $N_2 = 200$ frames from these four videos. Suppose the size of frames in the given video is $h \times w$, we stack each frame in each color channel as a column vector of size $N_1 \times 1$, where $N_1 = h \times w$, and then all column vectors into a matrix of size $3N_1 \times N_2$ for RPCA and Gamma, and into a tensor of size $N_1 \times N_2 \times 3$ for SNN, TRPCA and t-Gamma. The parameter *lambda* is set to $\lambda = \frac{1}{\max\{3N_1, N_2\}}$ for RPCA and Gamma, and $\lambda = \frac{1}{\max\{N_1, N_2\}3}$ for SNN, TRPCA and t-Gamma. The parameter γ in Gamma and t-Gamma is set to $\gamma = \frac{1}{\sqrt{\min\{3N_1, N_2\}}}$ and $\gamma = \frac{3}{\sqrt{\min\{N_1, N_2\}}}$ respectively. As suggested in [20], λ in SNN is set to $[200, 2, 20]$.

The background separated results are depicted in Figures 3–6. In Figures 3 and 4, we can see that our method could reconstruct the background with fewer ghosting effects and more completely. From Figures 5 and 6, it could be observed that though all the method separate the background from the foreground effectively, the proposed algorithm are more robust to the dynamic foreground and illumination variation, e.g., only Gamma and t-Gamma can remove the steps of moving escalator in Figure 5, while the method Gamma cannot adapt to change of illumination in Figure 6.

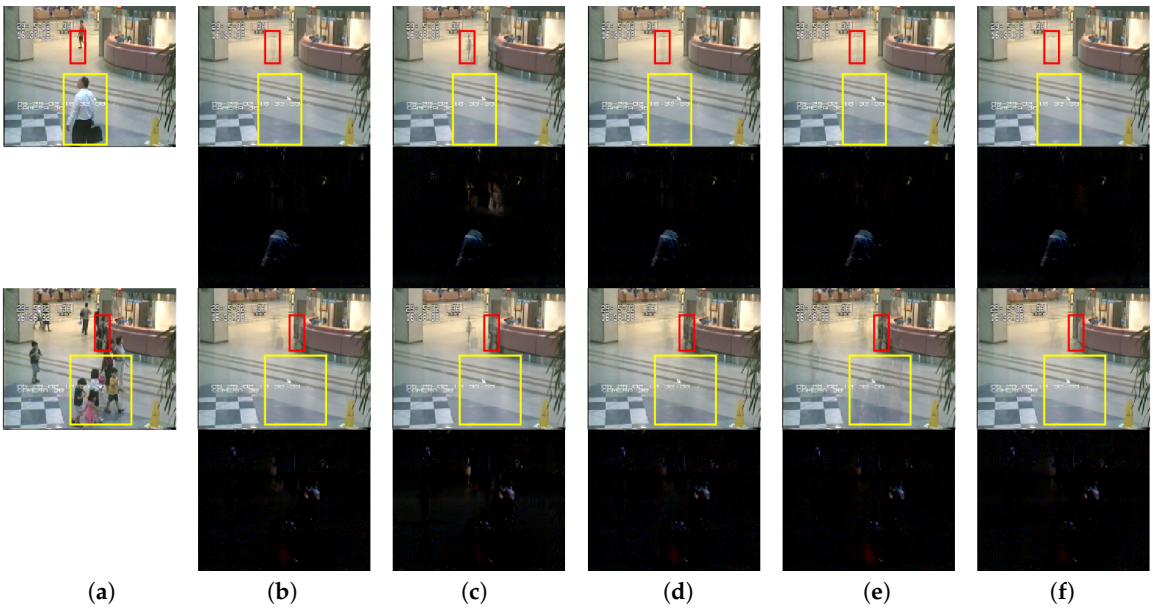
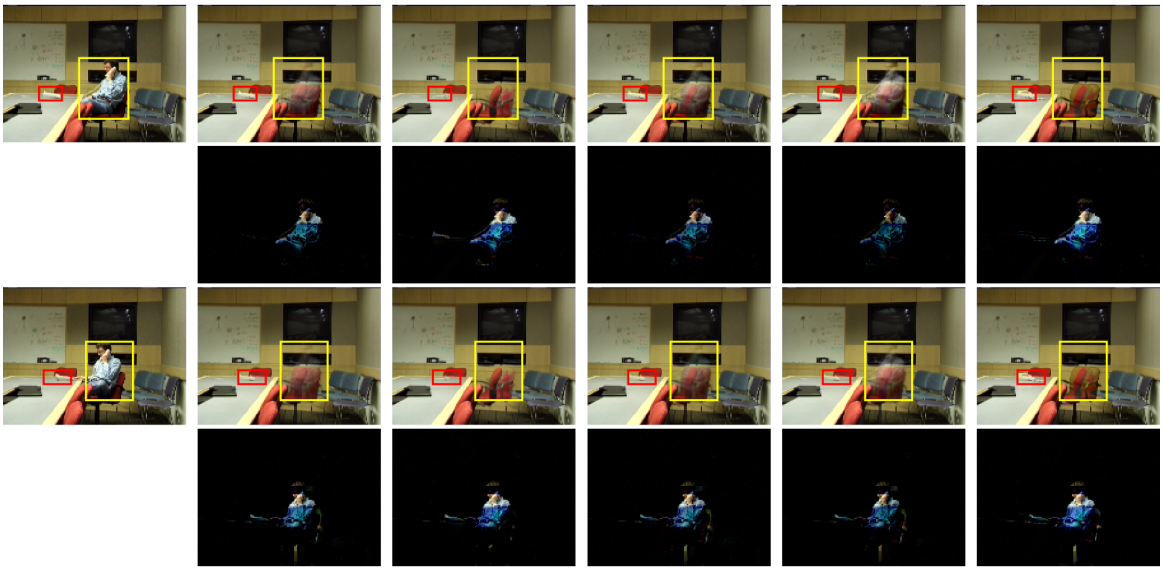
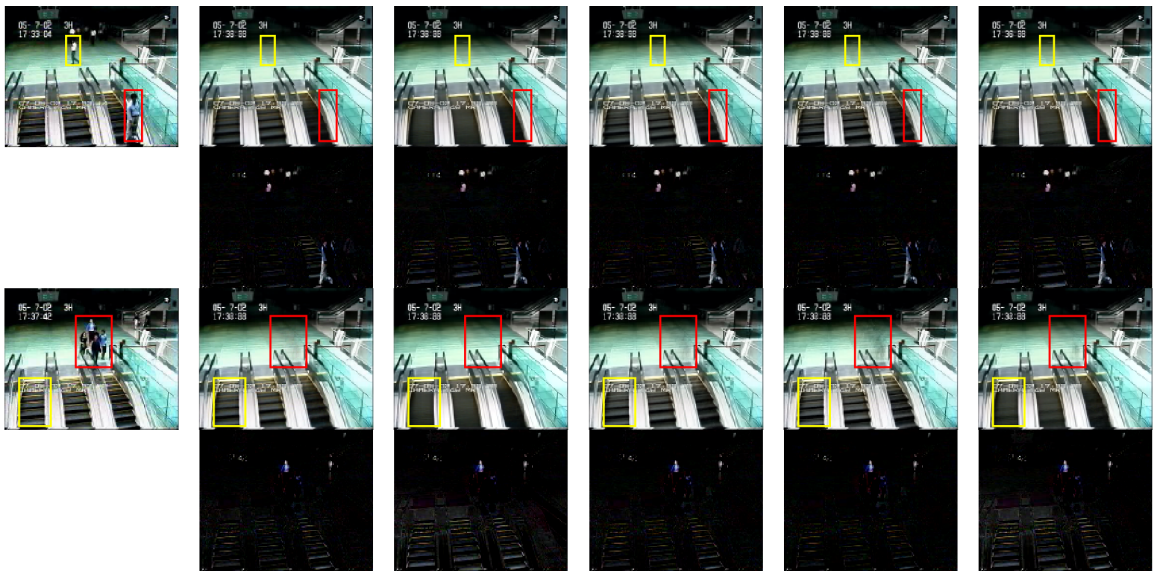


Figure 3. Background modeling: two example frames from Hall sequences are shown. (**a**) Original; (**b**) RPCA [4]; (**c**) Gamma [26]; (**d**) SNN [49]; (**e**) TRPCA [20]; (**f**) t-Gamma.



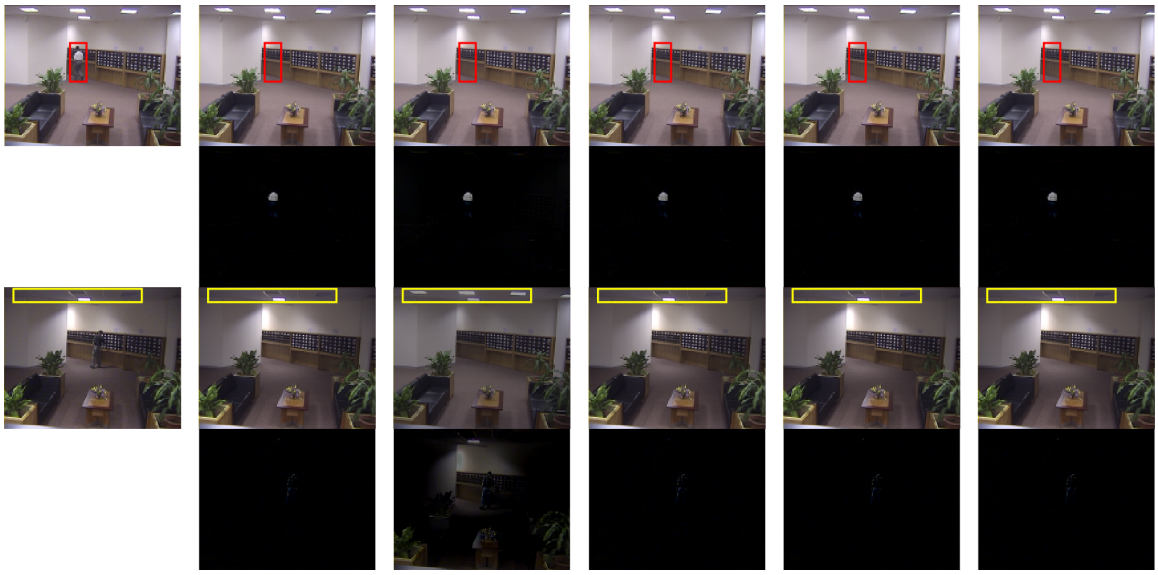
(**a**) (**b**) (**c**) (**d**) (**e**) (**f**)

Figure 4. Background modeling: two example frames from MovedObject sequences are shown. (**a**) Original; (**b**) RPCA [4]; (**c**) Gamma [26]; (**d**) SNN [49]; (**e**) TRPCA [20]; (**f**) t-Gamma.



(**a**) (**b**) (**c**) (**d**) (**e**) (**f**)

Figure 5. Background modeling: two example frames from Escalator sequences are shown. (**a**) Original; (**b**) RPCA [4]; (**c**) Gamma [26]; (**d**) SNN [49]; (**e**) TRPCA [20]; (**f**) t-Gamma.



(**a**) (**b**) (**c**) (**d**) (**e**) (**f**)

Figure 6. Background modeling: two example frames from Lobby sequences are shown. (**a**) Original; (**b**) RPCA [4]; (**c**) Gamma [26]; (**d**) SNN [49]; (**e**) TRPCA [20]; (**f**) t-Gamma.

5.2.2. Quantitative Experiments

Since the datasets in Section 5.2.1 do not include the ground truth background for each frame of the video, we perform the algorithms on the ChangeDetection.net(CDNet) dataset2014 [50] which consists of visual surveillance and smart environments. However, only part of frames are accompanied with ground truth background and thus we only take out some of frames of the video for testing. That is, we select the frames from 900 to 1000 of the dataset *highway*, *pedestrians* and *PETS2006* in the baseline category. Hence the size of the video will be $h \times w \times 101$, where $h \times w$ is the resolution of each frame, and 101 is the number of frames. We compare our algorithm with RPCA [4], Gamma [26], SNN [49] and TRPCA [20]. The parameters setting is conducted from Section 5.2.1.

To evaluate the performance of the experiment results, we use the three metrics for the quantitative measure, that is, Recall, Precision and F-Measure. They are defined as:

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}},$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}},$$

$$\text{F-Measure} = 2\frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}},$$

where TP, FP and FN are true positives numbers, false positives numbers and false negatives numbers respectively. We primarily use F-Measure for overall evaluation since it combines the metrics Recall and Precision and also makes a balance between Recall and Precision which can be viewed as a good indicators for comparison.

From Table 2, we can see that the proposed algorithm achieves better segmentation performance among the compare methods with respect to the F-Measure. Some of the frames of the segmentation results are shown in Figure 7, it can be observed that among the three dataset in CDNet our method can usually get the clearer and more accurate background segmentation. These quantitative comparison illustrates the superior of t-Gamma tensor quasi-norm as the surrogate of the rank of tensor.

Table 2. Quantitative Evaluation for algorithm on CDNet Datasets

Dataset	RPCA			Gamma			SNN			TNN			t-Gamma		
	Recall	Precision	F-Measure	Recall	Precision	F-Measure	Recall	Precision	F-Measure	Recall	Precision	F-Measure	Recall	Precision	F-Measure
highway	0.8642	0.9109	0.8869	0.8837	0.9067	0.895	0.6982	0.9041	0.7879	0.8235	0.902	0.861	0.8835	0.909	**0.8961**
pedestrians	0.9983	0.9419	0.9692	0.9982	0.9275	0.9616	0.9955	0.9285	0.9608	0.9983	0.9258	0.9607	0.9982	0.9309	**0.9634**
PETS2006	0.8151	0.7131	0.7607	0.8326	0.7678	0.7989	0.288	0.6974	0.4077	0.8033	0.6853	0.7396	0.8301	0.7805	**0.8045**

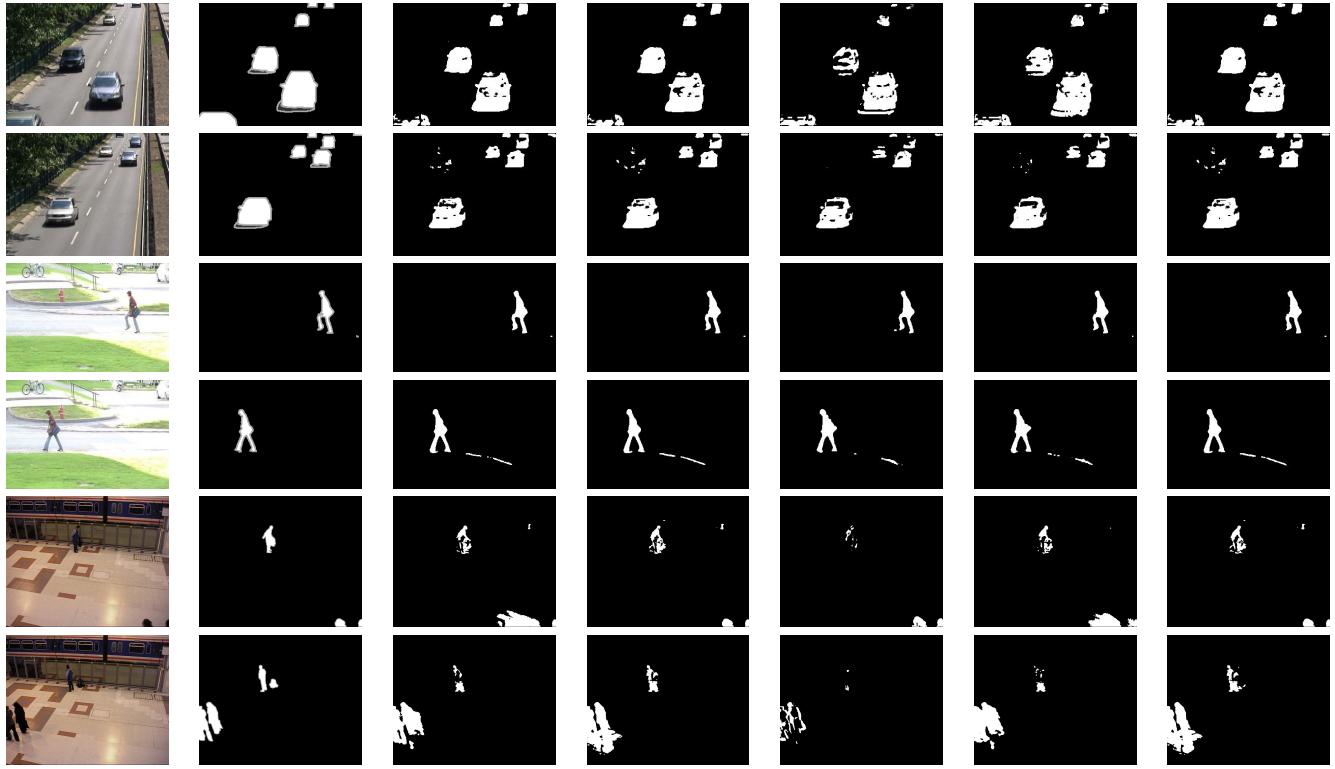


Figure 7. Visual results for CDNet dataset. Row 1 and row 2 are samples of *highway*; row 3 and row 4 are samples of *pedestrians*; row 5 and row 6 are samples of *PETS2006*. (**a**) Original; (**b**) Ground True; (**c**) RPCA [4]; (**d**) Gamma [26]; (**e**) SNN [49]; (**f**) TRPCA [20]; (**g**) t-Gamma

6. Concluding Remarks

In this paper, we conduct the TRPCA with a new t-Gamma tensor quasi-norm used as the regularization for tensor rank function. The t-Gamma tensor quasi-norm is developed via t-SVD in the Fourier domain, which simplifies the required optimization process. Compared with the convex tensor nuclear norm, our regularization approach appears to be more effective in approximating the tensor tubal rank. We incorporate the alternating direction method of multipliers with the t-Gamma tensor quasi-norm to efficiently solve the non-convex TRPCA problem, yielding an approachable framework for non-convex optimization. Numerical experimental results indicate that our approach outperforms the existing methods in distinguishable ways.

Author Contributions: Conceptualization, M.Y.; Data curation, Q.L.; Formal analysis, M.Y.; Funding acquisition, S.C.; Investigation, M.Y. and M.X.; Methodology, Q.L., M.Y. and M.X.; Project administration, S.C., W.L. and M.X.; Resources, M.X.; Software, Q.L.; Supervision, S.C., W.L. and M.X.; Validation, S.C.; Visualization, Q.L.; Writing—original draft, M.Y.; Writing—review & editing, M.X.

Funding: This research was funded by the National Science Foundation Division of Mathematical Sciences of the United States (Grant No. 140928), the National Natural Science Foundation of China (Grant Nos. 11671158, U181164, 61201392), the Natural Science Foundation of Guangdong Province, China (Grant No. 2015A030313497) and the Science and Technology Planning Project of Guangdong Province, China (Grant Nos. 2017B090909004, 2017B010124003, 2017B090909001).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Landsberg, J. *Tensors: Geometry and Applications*; Graduate Studies in Mathematics; American Mathematical Society: Providence, RI, USA, 2012.
2. Bro, R. Multi-Way Analysis in the Food Industry: Models, Algorithms, and Applications. Ph.D. Thesis, University of Amsterdam (NL), Amsterdam, The Netherlands, 1998.
3. Yang, M.; Li, W.; Xiao, M. On identifiability of higher order block term tensor decompositions of rank $L_r\otimes$ rank-1. *Linear Multilinear Algebra* **2018**, 1–23. [CrossRef]
4. Candès, E.J.; Li, X.; Ma, Y.; Wright, J. Robust principal component analysis? *J. ACM* **2011**, *58*, 11. [CrossRef]
5. Wright, J.; Ganesh, A.; Rao, S.; Peng, Y.; Ma, Y. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, USA, 7–10 December 2009; pp. 2080–2088.
6. Kang, Z.; Pan, H.; Hoi, S.C.H.; Xu, Z. Robust Graph Learning From Noisy Data. *IEEE Trans. Cybern.* **2019**, 1–11. [CrossRef]
7. Peng, C.; Kang, Z.; Cai, S.; Cheng, Q. Integrate and Conquer: Double-Sided Two-Dimensional k-Means Via Integrating of Projection and Manifold Construction. *ACM Trans. Intell. Syst. Technol.* **2018**, *9*, 57. [CrossRef]
8. Sun, P.; Qin, J. Speech enhancement via two-stage dual tree complex wavelet packet transform with a speech presence probability estimator. *J. Acoust. Soc. Am.* **2017**, *141*, 808–817. [CrossRef] [PubMed]
9. Vaswani, N.; Bouwmans, T.; Javed, S.; Narayanamurthy, P. Robust subspace learning: Robust PCA, robust subspace tracking, and robust subspace recovery. *IEEE Signal Process. Mag.* **2018**, *35*, 32–55. [CrossRef]
10. Bouwmans, T.; Zahzah, E.H. Robust PCA via principal component pursuit: A review for a comparative evaluation in video surveillance. *Comput. Vis. Image Underst.* **2014**, *122*, 22–34. [CrossRef]
11. Bouwmans, T.; Javed, S.; Zhang, H.; Lin, Z.; Otazo, R. On the applications of robust PCA in image and video processing. *Proc. IEEE* **2018**, *106*, 1427–1457. [CrossRef]
12. Cichocki, A.; Lee, N.; Oseledets, I.; Phan, A.H.; Zhao, Q.; Mandic, D.P. Tensor networks for dimensionality reduction and large-scale optimization: Part 1 low-rank tensor decompositions. *Found. Trends® Mach. Learn.* **2016**, *9*, 249–429. [CrossRef]
13. Kilmer, M.E.; Martin, C.D. Factorization strategies for third-order tensors. *Linear Algebra Appl.* **2011**, *435*, 641–658. [CrossRef]
14. Hao, N.; Kilmer, M.E.; Braman, K.; Hoover, R.C. Facial recognition using tensor-tensor decompositions. *SIAM J. Imaging Sci.* **2013**, *6*, 437–463. [CrossRef]

15. Zhang, Z.; Ely, G.; Aeron, S.; Hao, N.; Kilmer, M. 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, Columbus, OH, USA, 23–28 June 2014; pp. 3842–3849.
16. Signoretto, M.; De Lathauwer, L.; Suykens, J.A. Nuclear norms for tensors and their use for convex multilinear estimation. *Linear Algebra Appl.* **2010**, *43*. Available online: ftp://ftp.esat.kuleuven.be/sista/signoretto/Signoretto_nucTensors2.pdf (accessed on 3 April 2019).
17. Cichocki, A. Era of big data processing: A new approach via tensor networks and tensor decompositions. *arXiv* **2014**, arXiv:1403.2048.
18. Grasedyck, L.; Kressner, D.; Tobler, C. A literature survey of low-rank tensor approximation techniques. *GAMM-Mitteilungen* **2013**, *36*, 53–78. [CrossRef]
19. Lu, C.; Feng, J.; Chen, Y.; Liu, W.; Lin, Z.; Yan, S. Tensor Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Tensors via Convex Optimization. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 5249–5257.
20. Lu, C.; Feng, J.; Chen, Y.; Liu, W.; Lin, Z.; Yan, S. Tensor Robust Principal Component Analysis with A New Tensor Nuclear Norm. *arXiv* **2018**, arXiv:1804.03728.
21. Semerci, O.; Hao, N.; Kilmer, M.E.; Miller, E.L. Tensor-based formulation and nuclear norm regularization for multienergy computed tomography. *IEEE Trans. Image Process.* **2014**, *23*, 1678–1693. [CrossRef]
22. Ciccone, V.; Ferrante, A.; Zorzi, M. Robust identification of "sparse plus low-rank" graphical models: An optimization approach. In Proceedings of the 2018 IEEE Conference on Decision and Control (CDC), Miami Beach, FL, USA, 17–19 December 2018; pp. 2241–2246.
23. Ciccone, V.; Ferrante, A.; Zorzi, M. Factor Models with Real Data: A Robust Estimation of the Number of Factors. *IEEE Trans. Autom. Control* **2018**. [CrossRef]
24. Fazel, S.M. Matrix Rank Minimization with Applications. Ph.D. Thesis, Stanford University, Stanford, CA, USA, 2003.
25. Recht, B.; Fazel, M.; Parrilo, P.A. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.* **2010**, *52*, 471–501. [CrossRef]
26. Kang, Z.; Peng, C.; Cheng, Q. Robust PCA Via Nonconvex Rank Approximation. In Proceedings of the 2015 IEEE International Conference on Data Mining (ICDM), Atlantic City, NJ, USA, 14–17 November 2015; pp. 211–220.
27. Lewis, A.S.; Sendov, H.S. Nonsmooth analysis of singular values. Part I: Theory. *Set-Valued Anal.* **2005**, *13*, 213–241. [CrossRef]
28. Liu, Y.; Chen, L.; Zhu, C. Improved Robust Tensor Principal Component Analysis via Low-Rank Core Matrix. *IEEE J. Sel. Top. Signal Process.* **2018**, *12*, 1378–1389. [CrossRef]
29. Chen, Y.; Wang, S.; Zhou, Y. Tensor Nuclear Norm-Based Low-Rank Approximation With Total Variation Regularization. *IEEE J. Sel. Top. Signal Process.* **2018**, *12*, 1364–1377. [CrossRef]
30. Kong, H.; Xie, X.; Lin, Z. t-Schatten-p Norm for Low-Rank Tensor Recovery. *IEEE J. Sel. Top. Signal Process.* **2018**, *12*, 1405–1419. [CrossRef]
31. Tarzanagh, D.A.; Michailidis, G. Fast Monte Carlo Algorithms for Tensor Operations. *arXiv* **2017**, arXiv:1704.04362.
32. Driggs, D.; Becker, S.; Boyd-Graber, J. Tensor Robust Principal Component Analysis: Better recovery with atomic norm regularization. *arXiv* **2019**, arXiv:1901.10991.
33. Sobral, A.; Baker, C.G.; Bouwmans, T.; Zahzah, E.H. Incremental and multi-feature tensor subspace learning applied for background modeling and subtraction. In Proceedings of the International Conference Image Analysis and Recognition, Vilamoura, Portugal, 22–24 October 2014; pp. 94–103.
34. Sobral, A.; Javed, S.; Ki Jung, S.; Bouwmans, T.; Zahzah, E.H. Online stochastic tensor decomposition for background subtraction in multispectral video sequences. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Santiago, Chile, 7–13 December 2015; pp. 106–113.
35. Bouwmans, T.; Sobral, A.; Javed, S.; Jung, S.K.; Zahzah, E.H. Decomposition into low-rank plus additive matrices for background/foreground separation: A review for a comparative evaluation with a large-scale dataset. *Comput. Sci. Rev.* **2017**, *23*, 1–71. [CrossRef]
36. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends® Mach. Learn.* **2011**, *3*, 1–122. [CrossRef]

37. Kilmer, M.E.; Braman, K.; Hao, N.; Hoover, R.C. Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. *SIAM J. Matrix Anal. Appl.* **2013**, *34*, 148–172. [CrossRef]
38. Tyrtyshnikov, E.E. *A Brief Introduction to Numerical Analysis*; Springer Science & Business Media: New York, NY, USA, 2012.
39. Marc Peter Deisenroth, A.A.F.; Ong, C.S. *Mathematics for Machine Learning*; Cambridge University Press: Cambridge, UK, 2019.
40. Moreau, J.J. Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France* **1965**, *93*, 273–299. [CrossRef]
41. Dong, W.; Shi, G.; Li, X.; Ma, Y.; Huang, F. Compressive Sensing via Nonlocal Low-Rank Regularization. *IEEE Trans. Image Process.* **2014**, *23*, 3618–3632. [CrossRef] [PubMed]
42. Zhang, G.; Lin, Y. *Lectures in Functional Analysis*; Peking University Press: Beijing, China, 1987.
43. Donoho, D.L. De-noising by soft-thresholding. *IEEE Trans. Inf. Theory* **1995**, *41*, 613–627. [CrossRef]
44. Lu, H.; Plataniotis, K.; Venetsanopoulos, A. *Multilinear Subspace Learning: Dimensionality Reduction of Multidimensional Data*; Chapman & Hall/CRC Machine Learning & Pattern Recognition Series; CRC Press: Boca Raton, FL, USA, 2013.
45. Zhang, Z.; Aeron, S. Exact Tensor Completion Using t-SVD. *IEEE Trans. Signal Process.* **2017**, *65*, 1511–1526. [CrossRef]
46. Luenberger, D.G. *Optimization by Vector Space Methods*; John Wiley & Sons: New York, NY, USA, 1997.
47. Liu, J.; Musialski, P.; Wonka, P.; Ye, J. Tensor Completion for Estimating Missing Values in Visual Data. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 208–220. [CrossRef] [PubMed]
48. Martin, D.; Fowlkes, C.; Tal, D.; Malik, J. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In Proceedings of the Eighth IEEE International Conference on Computer Vision (ICCV 2001), Vancouver, BC, Canada, 7–14 July 2001; Volume 2, pp. 416–423.
49. Huang, B.; Mu, C.; Goldfarb, D.; Wright, J. Provable models for robust low-rank tensor completion. *Pac. J. Optim.* **2015**, *11*, 339–364.
50. Goyette, N.; Jodoin, P.-M.; Porikli, F.; Konrad, J.; Ishwar, P. Changedetection.net: A new change detection benchmark dataset, 2012. In Proceedings of the IEEE Workshop on Change Detection (CDW-2012) at CVPR-2012, Providence, RI, USA, 16–21 June 2012.