Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 11, No. 2, pp. 403–418, April 2015

RANK-MIN-ONE AND SPARSE TENSOR DECOMPOSITION FOR SURVEILLANCE VIDEO*

Xianchao Xiu and Lingchen Kong

Abstract: Sparse tensor optimization has recently attracted much attention since it has many applications in areas such as biology, computer vision and information science. In this paper, we focus on the application of tensor optimization in surveillance video. Based on the static background of surveillance video, we introduce the new definition of rank-min-one tensor. Then we consider a rank-min-one and sparse tensor decomposition model for surveillance video. We establish the modified iterative reweighted ℓ_1 algorithm (MIRL1), and give its convergence analysis. For synthetic and real surveillance data, numerical experiments are also presented to illustrate the efficiency of our proposed MIRL1.

Key words: *surveillance video, rank-min-one and sparse tensor decomposition, modified iterative reweighted ℓ_1 algorithm*

Mathematics Subject Classification: *90C90, 65K10*

1 Introduction

In surveillance video, video signals are usually captured by fixed cameras and transferred to a data processing center. The number of cameras available worldwide has increased dramatically over the decade, which results in a huge amount of data. Therefore, how to detect the moving objects and anomalies in surveillance video quickly has become a significant problem.

Moving objects are detected by background subtraction, which is important in surveillance video, see, e.g., [17, 22, 23] and references therein. All methods for background subtraction to detect foreground are applied with the help of establishing the frame difference [16]. Subsequent approaches aimed to model the variations and uncertainties in background appearance, such as the mixture of Gaussian [11], the kernel density estimation [10]. These traditional methods cost plentiful time due to excessive management for all pixels of surveillance video. Note that Compressed Sensing states that if a signal has a sparse representation in some basis then it can be reconstructed from a small set of linear measurements, which was introduced by Donoho [9], Candès and Tao [5] and has been widely studied in various areas. Based on this idea, Hale et al. [14] and Cevher et al. [7] modeled the background subtraction in surveillance video as a sparse approximation problem and solved it via convex optimization. In the literature, each frame is stacked as a column of a matrix. The background component has relatively small changes over a short period, so the background

*The work was partly supported by the National Natural Science Foundation of China (11431002,11171018).

matrix can be denoted by a low-rank matrix. The foreground component consists of moving objects, so the foreground matrix can be seen as a sparse one with most of its entries being zero or nearly zero. Thus, the low-rank matrix can be used to capture the global structure of the surveillance video, while the sparse matrix can capture the local structure. Many authors began to take advantage of low-rank and sparse model to solve background subtraction problem, see, e.g., [4, 8, 26, 3].

With the development of modern computer, multi-dimensional data is becoming more and more prevalent in many areas, which results that large researchers have concentrated on tensor study. Tensors, which emerged as a generalization of vectors and matrices, make it possible to deal with data that has intrinsically many dimensions. However, traditional matrix-based data analysis is inherently two-dimensional, which limits its usefulness in extracting information from a multidimensional perspective. On the other hand, tensor-based multilinear data analysis has shown that tensor models are capable of taking full advantage of the multilinear structures to provide better understanding and more precision. Many authors have made great effort, including computer vision and information science, see [20, 12, 25, 13, 15] for example. Liu et al. [20] was the first to study on the tensor completion problem and laid the theoretical foundation of low-rank tensor completion. Gandy et al. [12] paid attention to the low n-rank tensor recovery problem. They introduced its convex relaxation of the n-rank, and proposed two algorithms to solve the low n-rank tensor recovery problem. Yang et al. [25] proposed a fixed point iterative method for low n-rank tensor pursuit and proved the convergence of the method under some assumptions. Recently, Goldfarb et al. [13] drew on recent advances in robust Principal Component Analysis, then proposed tailored optimization algorithms with global convergence guarantees. More recently, Huang et al. [15] proposed a class of convex recovery models that can be proved to guarantee exact recovery under certain conditions.

For surveillance video, the underlying low-dimensional structure and the irregular sparse patterns of the tensor data can be considered as the semantic features. How to detect the irregular sparse patterns of the tensor data? We utilize the low-rank and sparse tensor decomposition model to distinguish the irregular sparse patterns. In practice, the background part varies a little in a short time, so it can be approximately seen as a rank-min-one tensor (see Definition 2.1), which is not the usual rank-one tensor. Meanwhile, the foreground part records the dynamic contents, hence the foreground can be regarded as a sparse tensor. In this paper, we propose a rank-min-one and sparse tensor decomposition model to extract the moving part in surveillance video. We make use of favorable structure, and develop the modified iterative reweighted ℓ_1 algorithm. Instead of directly computing tensors, we suggest to transform tensors into matrices via unfolding operation. After calculation, we fold those matrices to acquire the desired tensors. To illustrate the efficiency of our proposed algorithm, we conduct numerical experiments including synthetic and real surveillance data.

This paper is organized as follows. In Section 2, we review some notations and present rank-min-one and sparse tensor decomposition model. We establish the modified iterative reweighted ℓ_1 algorithm and its convergence result in Section 3. Section 4 reports experimental results. Some final remarks are given in the last section.

2 Rank-min-one and Sparse Tensor Decomposition

In this section, we review some notations and tensor basic results, and propose the rank-min-one and sparse tensor decomposition. We begin with some basic facts about tensor which will be used in subsequent analysis. More details can be found in [18].

2.1 Notations and tensor basics

Throughout this paper we denote tensors by boldface Euler script letters, e.g., $\mathcal{X}$. Matrices are denoted by capital letters, e.g., X; vectors are denoted by boldface lowercase letters, e.g., $\mathbf{x}$; and scalars are denoted by lowercase letters, e.g., x.

Matricization and tensorization For the K-way tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K}$, its mode-i *fiber* is a n_i-dimensional column vector defined by fixing every index but the i-th of $\mathcal{X}$. The mode-i *unfolding* (matricization) of the tensor $\mathcal{X}$ is the matrix denoted by $\mathcal{T}_i(\mathcal{X}) := X_{(i)} \in \mathbb{R}^{n_i \times \prod_{j \neq i} n_j}$ that is obtained by concatenating all the mode-i fibers of $\mathcal{X}$ as column vectors. The opposite operation *folding* (tensorization) is defined as $fold_i(X_{(i)}) := \mathcal{X}$.

Inner product and norms The inner product of two tensors $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K}$ is defined as $\langle \mathcal{X}, \mathcal{Y} \rangle := vec(\mathcal{X})^T vec(\mathcal{Y})$, and the Frobenius norm of $\mathcal{X}$ is denoted by $\|\mathcal{X}\| := \sqrt{\langle \mathcal{X}, \mathcal{X} \rangle}$. $\|\cdot\|_0$ counts the number of nonzero entries, and $\|\cdot\|_1$ is defined as the sum of absolute of all entries. The nuclear norm $\|X\|_*$ of a matrix X is the sum of all its singular values, i.e. $\|X\|_* := \sum \sigma_i$, where the SVD of $X = U\text{Diag}(\sigma)V^T$. For vector $\mathbf{x}$, Diag($\mathbf{x}$) denotes a diagonal matrix with i-th diagonal element being x_i.

Tucker decomposition and rank-min-one tensor The Tucker decomposition approximates $\mathcal{X}$ as $\mathcal{X} := \mathcal{C} \times_1 A_1 \times_2 A_2 \times \cdots \times_K A_K$, where $\mathcal{C} \in \mathbb{R}^{r_1 \times r_2 \times \cdots \times r_K}$ is called the *core tensor*, and the factor matrices $A_i \in \mathbb{R}^{n_i \times r_i}$ are columnwise orthonormal. The Tucker rank of $\mathcal{X}$ is a K-dimensional vector whose i-th entry is the (matrix) rank of the mode-i unfolding $X_{(i)}$, i.e., $\text{rank}_{tc}(\mathcal{X}) := (\text{rank}(X_{(1)}), \text{rank}(X_{(2)}), \cdots, \text{rank}(X_{(K)}))$.

Definition 2.1. We call the tensor $\mathcal{X}$ is rank-min-one if

$$\text{rank}_{min}(\mathcal{X}) := \min\{\text{rank}(X_{(1)}), \text{rank}(X_{(2)}), \cdots, \text{rank}(X_{(K)})\} = 1.$$

For example, surveillance video can be described as a 3-way tensor, which consists of background part and foreground part. If we stack the background part frame by frame, then we get a rank-one matrix since the background is relatively static, which shows that the background part can be represented by a rank-min-one tensor.

At the end of this subsection, for each entry x_{ij} in X, we define

$$\text{sgn}(x_{ij}) = \begin{cases} 1, & x_{ij} \geq 0; \\ -1, & x_{ij} < 0. \end{cases}$$

For two matrices $X, Y \in \mathbb{R}^{m \times n}$, the Hadmard product is defined by $(X \circ Y)_{ij} = x_{ij} y_{ij}$.

2.2 Model analysis

We introduce the low n-rank and sparse tensor decomposition for surveillance video

$$\begin{aligned} \min \quad & \frac{1}{3} \sum_{i=1}^{3} \text{rank}(L_{(i)}) + \|\mathcal{S}\|_0 \\ \text{s.t.} \quad & \mathcal{L} + \mathcal{S} = \mathcal{M}, \end{aligned} \tag{2.1}$$

where $\mathcal{M}, \mathcal{L}, \mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ are 3-way tensors with identical size in each mode. $\mathcal{M}$, $\mathcal{L}$ and $\mathcal{S}$ represent the observed tensor data, the correspondent structured part and irregular sparse part, respectively. In the special case for matrix, model (2.1) reduces to the following low-rank and sparse matrix decomposition problem

$$\begin{aligned} \min \quad & \text{rank}(L) + \|S\|_0 \\ \text{s.t.} \quad & L + S = M, \end{aligned} \tag{2.2}$$

where $M \in \mathbb{R}^{m\times n}$ is the given measurement matrix of surveillance video, $L \in \mathbb{R}^{m\times n}$ is the background matrix, $S \in \mathbb{R}^{m\times n}$ is the foreground matrix. m denotes the number of pixels, n denotes the number of frames. However, (2.2) is NP-hard due to the discontinuity and nonconvexity of rank(L) and $\|S\|_0$. Chandrasekaran et al. [8] suggested the convex relaxation of (2.2) by replacing rank with nuclear norm and ℓ_0-norm with ℓ_1-norm, that is

$$\begin{aligned} \min \quad & \|L\|_* + \|S\|_1 \\ \text{s.t.} \quad & L + S = M. \end{aligned} \tag{2.3}$$

They also gave sufficient conditions for (2.3) to exactly recover (2.2) by introducing the notation of rank-sparsity incoherence, which expressed the uncertainty principle between the sparsity pattern and its row and column spaces. Candès et al. [4] put forward that under rather weak assumptions, they can recover the real L and S exactly via solving (2.3).

For low n-rank and sparse tensor decomposition problem (2.1), it is also NP-hard. It is natural to surrogate Tucker vector of ranks with a weighted sum of norms, thus a tractable convex optimization problem can be obtained

$$\begin{aligned} \min \quad & \frac{1}{3}\sum_{i=1}^{3} \beta_i \|L_{(i)}\|_* + \|\mathcal{S}\|_1 \\ \text{s.t.} \quad & \mathcal{L} + \mathcal{S} = \mathcal{M}, \end{aligned} \tag{2.4}$$

where β_i is a positive weighted to balance the weights of rank and sparsity. As we know, in surveillance video, the background component has relative small changes over a short period. If we stack the background tensor properly, then background part can be seen as a rank-min-one tensor, that is, $\text{rank}_{min}(\mathcal{L}) = 1$. While, the foreground component consists of moving objects, so the foreground part can be regarded as a sparse tensor. For matrix case, Li et al. [19] proposed to substitute low-rank matrix for rank-one matrix. Then, Ma et al. [21] and Xiu et al. [24] extended it to the application of medical imaging, and shown it works well. Therefore, the low n-rank and sparse tensor decomposition problem (2.1) can be reformulated to

$$\begin{aligned} \min \quad & \|\mathcal{S}\|_0 \\ \text{s.t.} \quad & \mathcal{L} + \mathcal{S} = \mathcal{M}, \\ & \text{rank}_{min}(\mathcal{L}) = 1. \end{aligned} \tag{2.5}$$

We denote the mode-3 unfolding of the tensors $\mathcal{L}, \mathcal{S}, \mathcal{M}$ as $\mathcal{T}_3(\mathcal{L}), \mathcal{T}_3(\mathcal{S}), \mathcal{T}_3(\mathcal{M}) \in \mathbb{R}^{n_3\times(n_1\times n_2)}$. Notice that $\text{rank}_{min}(\mathcal{L}) = 1$ and we can set $\mathcal{T}_3(\mathcal{L}) := \mathbf{1}\mathbf{u}^T$, where $\mathbf{1}$ is a column vector in $\mathbb{R}^{n_3}$, $\mathbf{u}$ is also a column vector in $\mathbb{R}^{n_1 n_2}$. Therefore, by replacing ℓ_0-norm with ℓ_1-norm, the unfolding relaxation of (2.5) can be rewritten as

$$\begin{aligned} \min \quad & \|\mathcal{T}_3(\mathcal{S})\|_1 \\ \text{s.t.} \quad & \mathbf{1}\mathbf{u}^T + \mathcal{T}_3(\mathcal{S}) = \mathcal{T}_3(\mathcal{M}). \end{aligned} \tag{2.6}$$

Much of the recent work focuses on the reweighted ℓ_1 minimization [1, 6, 27] for matrix case, and shows that it performs better than classical ℓ_1 minimization in recovering sparse signal. We define some necessary notations for ease of presentation

$$\begin{aligned} A &:= [\mathbf{1} \;\; I] \in \mathbb{R}^{n_3\times(n_3+1)}, \\ X &:= \begin{bmatrix} \mathbf{u}^T \\ \mathcal{T}_3(\mathcal{S}) \end{bmatrix} \in \mathbb{R}^{(n_3+1)\times n_1 n_2}, \end{aligned}$$

where I is the identity matrix. It then follows that,

$$AX = \mathbf{1}\mathbf{u}^T + \mathcal{T}_3(\mathcal{S}) = \mathcal{T}_3(\mathcal{M}).$$

Furthermore, combining with the above equality, problem (2.6) becomes

$$\begin{array}{ll} \min & \|W \circ X\|_1 \\ \text{s.t.} & AX = \mathcal{T}_3(\mathcal{M}), \end{array} \tag{2.7}$$

where $\circ$ denotes the Hadamard product, and W is the reweighted matrix defined as

$$W := \left[\begin{array}{c} 0 \\ \hline \overline{W} \end{array} \right] \in \mathbb{R}^{(n_3+1)\times n_1 n_2},$$

where $\overline{W} \in \mathbb{R}^{n_3 \times n_1 n_2}$. In order to solve (2.7), we get the following unconstrained optimization problem by putting the constraint into the objective function

$$\min \ \ \lambda\|W \circ X\|_1 + \frac{1}{2}\|AX - \mathcal{T}_3(\mathcal{M})\|^2 \tag{2.8}$$

with $\lambda > 0$ being a penalty parameter.

3 Modified Iterative Reweighted ℓ_1 Algorithm and Convergence

This section will consider the modified iterative reweighted ℓ_1 algorithm for rank-min-one and sparse tensor decomposition, then establish its global convergence.

Before presenting the algorithm, we first describe the idea of majorization,

$$\begin{array}{rcl} f(X) & := & \lambda\|W \circ X\|_1 + \dfrac{1}{2}\|AX - \mathcal{T}_3(\mathcal{M})\|^2 \\ & \leq & \lambda\|W \circ X\|_1 + \dfrac{1}{2}\|AY - \mathcal{T}_3(\mathcal{M})\|^2 + \langle A^T(AY - \mathcal{T}_3(\mathcal{M})), X - Y\rangle + \dfrac{L}{2}\|X - Y\|^2 \\ & := & F(X, Y), \end{array}$$

where $L > \lambda_{\max}(A^T A)$. Indeed, $F(X, Y)$ is a majorization function of $f(X)$, because of the fact that for any $X \neq Y$,

$$F(X, Y) > f(X), \qquad F(X, X) = f(X).$$

With the help of majorization, we can give the following lemma.

Lemma 3.1. *For*

$$\begin{array}{rcl} X^{k+1} & = & \text{argmin } F(X, X^k) \\ & = & \text{argmin } \lambda\|W \circ X\|_1 + \dfrac{L}{2}\|X - \widetilde{X}^k\|^2, \end{array} \tag{3.1}$$

where

$$\widetilde{X}^k = X^k - A^T(AX^k - \mathcal{T}_3(\mathcal{M}))/L.$$

The $(k+1)$-th iteration has a closed form solution

$$X^{k+1} = \max\left\{\widetilde{X}^k - (\lambda/L)W, 0\right\} \circ \text{sgn}(\widetilde{X}^k),$$

where sgn *is defined in subsection 2.1.*

Proof. X^{k+1} is an optimal solution if and only if

$$\begin{aligned} 0 &\in \lambda\partial\|W \circ X^{k+1}\|_1 + L(X^{k+1} - \widetilde{X}^k) \\ &= \lambda W \circ \partial\|X^{k+1}\|_1 + L(X^{k+1} - \widetilde{X}^k), \end{aligned} \tag{3.2}$$

where $\partial\|X^{k+1}\|_1$ is the subgradients of $\|\cdot\|_1$ at X^{k+1}. It is easy to see that (3.2) is equivalent to

$$0 \in X^{k+1} - \widetilde{X}^k + (\lambda/L)W \circ \partial\|X^{k+1}\|_1.$$

Then it demonstrates that the optimal solution can be described as

$$X^{k+1} = \max\left\{\widetilde{X}^k - (\lambda/L)W, 0\right\} \circ \operatorname{sgn}(\widetilde{X}^k).$$

The lemma holds immediately. □

Based on the above argument, we have derived the following majorized weighted ℓ_1 algorithm (MWL1) for (2.8).

Table 1: The framework of the MWL1.

Algorithm 1: Majorized Weighted ℓ_1 Algorithm (MWL1)
Input: the observed data $\mathcal{M}$, compute $\mathcal{T}_3(\mathcal{M})$;
Initialize: $\lambda,\ \varepsilon_0 > 0$, L, and $X^0,\ W$;
While $\|X^{k+1} - X^k\|^2 > \varepsilon_0$
Compute $\widetilde{X}^k = X^k - A^T(AX^k - \mathcal{T}_3(\mathcal{M}))/L$;
Compute $X^{k+1} = \max\left\{\widetilde{X}^k - (\lambda/L)W, 0\right\} \circ \operatorname{sgn}(\widetilde{X}^k)$;
End
Output: X, compute $\mathcal{L} = fold_3(\mathbf{1}\mathbf{u}^{kT})$, $\mathcal{S} = fold_3(\mathcal{T}_3(\mathcal{S}^k))$.

Lemma 3.2. *For given $\lambda > 0$ and $L > \lambda_{\max}(A^TA)$, if $\hat{X}$ is a global minimizer of $f(X)$, then $\hat{X}$ is also the global minimizer of $F(X, \hat{X})$.*

Proof. Since $F(X, \ \cdot\ )$ is a majorization of $f(X)$, we have

$$F(X, \hat{X}) \geq f(X) \geq f(\hat{X}) = F(\hat{X}, \hat{X}).$$

This completes the proof. □

Theorem 3.3. *For given $\lambda > 0$ and $L > \lambda_{\max}(A^TA)$, $\hat{X}$ is the global minimizer of* (2.8) *if and only if $\hat{X}$ satisfies the following fixed point equation*

$$\hat{X} = \max\left\{\widetilde{X} - (\lambda/L)W, 0\right\} \circ \operatorname{sgn}(\widetilde{X}), \tag{3.3}$$

where $\widetilde{X} = \hat{X} - A^T(A\hat{X} - \mathcal{T}_3(\mathcal{M}))/L$.

Proof. Necessity: Since $\hat{X}$ is a global minimizer of $f(X)$, by Lemma 3.2, $\hat{X}$ is also a global minimizer of $F(X, \hat{X})$. From Lemma 3.1, one can check that $\hat{X} = \max\{\widetilde{X} - (\lambda/L)W, 0\}\text{sgn}(\widetilde{X})$.

Sufficiency: (2.8) is a convex optimization and thus its global minimizer is its stationary point X^* which contents

$$\left\langle A^T(AX^* - \mathcal{T}_3(\mathcal{M})) + \lambda W, Y - X^* \right\rangle \geq 0.$$

If we have $\hat{X} = \max\{\widetilde{X} - (\lambda/L)W, 0\}\text{sgn}(\widetilde{X})$, one can check that

$$\begin{aligned} & \left\langle A^T(A\hat{X} - \mathcal{T}_3(\mathcal{M})) + \lambda W, Y - \hat{X} \right\rangle \\ = \; & L\left\langle \hat{X} - \left(\hat{X} - \left(A^T(A\hat{X} - \mathcal{T}_3(\mathcal{M})) + \lambda W\right)/L\right), Y - \hat{X} \right\rangle \\ = \; & L\left\langle \hat{X} - \left(\tilde{X} - (\lambda/L)W\right), Y - \hat{X} \right\rangle \\ \geq \; & 0, \end{aligned}$$

where the last inequality holds from the property of the projection. Therefore $\hat{X}$ is also a stationary point, namely $\hat{X}$ is also a global minimizer of (2.8). □

Theorem 3.4. *For given $\lambda > 0$ and $L > \lambda_{\max}(A^TA)$, let $\{X^k\}$ be the sequence generated by Algorithm* 1*, then*

A. *$\{f(X^k)\}$ is monotonically decreasing and converges to $f(\hat{X})$ where $\hat{X}$ is any accumulation point of $\{X^k\}$;*

B. *$\{X^k\}$ is asymptotically regular, that is, $\|X^{k+1} - X^k\|^2 \to 0$ when $k \to \infty$;*

C. *Any accumulation point of $\{X^k\}$ a global minimizer of* (2.8)*.*

Proof **A.** As $F(X, \ \cdot\ )$ is the majorization of $f(X)$, it holds

$$f(X^k) = F(X^k, X^k) \geq F(X^{k+1}, X^k) \geq f(X^{k+1}),$$

where the first inequality derives from Lemma 3.2. Therefore $\{f(X^k)\}$ is monotonically non-increasing. As $\{f(X^k)\}$ is bounded from below, $\{f(X^k)\}$ converges to a constant $\hat{f}$. From $\{X^k\} \subset \{X | f(X) \leq f(X^0)\}$ which is bounded, it yields that $\{X^k\}$ is also bounded, and thus $\{X^k\}$ has at least one accumulation point. Let $\hat{X}$ be an accumulation point of $\{X^k\}$. By the continuity of $f(X)$ and the convergence of $\{f(X^k)\}$, we get $\{f(X^k)\} \to \hat{f} = f(\hat{X})$ as $k \to \infty$.

B. Through simply computing, we have

$$\begin{aligned}
& f(X^k) - f(X^{k+1}) \\
\geq\ & F(X^{k+1}, X^k) - f(X^{k+1}) \\
=\ & \mu\|W \circ X^{k+1}\|_1 + \frac{1}{2}\|AX^k - \mathcal{T}_3(\mathcal{M})\|^2 + \langle A^T(AX^k - \mathcal{T}_3(\mathcal{M})), X^{k+1} - X^k\rangle \\
& + \frac{L}{2}\|X^{k+1} - X^k\|^2 - \lambda\|W \circ X^{k+1}\|_1 - \frac{1}{2}\|AX^k - \mathcal{T}_3(\mathcal{M})\|^2 \\
& - \langle A^T(AX^k - \mathcal{T}_3(\mathcal{M})), X^{k+1} - X^k\rangle - \frac{1}{2}(X^{k+1} - X^k)^T A^T A(X^{k+1} - X^k) \\
& - o(\|X^{k+1} - X^k\|^2) \\
\geq\ & \frac{1}{2}(X^{k+1} - X^k)^T(LI - A^T A)(X^{k+1} - X^k) - o(\|X^{k+1} - X^k\|^2) \\
\geq\ & \frac{\epsilon}{2}\|X^{k+1} - X^k\|^2 - o(\|X^{k+1} - X^k\|^2), \\
\geq\ & \frac{\epsilon}{3}\|X^{k+1} - X^k\|^2,
\end{aligned}$$

where $\epsilon \geq \lambda_{\min}(LI - A^T A) > 0$ and I is the identity matrix. The first inequality derives from the proof in **A**, the second equality is from the Taylor expansion of $\frac{1}{2}\|AX^{k+1} - \mathcal{T}_3(\mathcal{M})\|^2$ at point X^k , and the fourth inequality holds due to $L > \lambda_{\max}(A^T A)$.
Henceforth, for ϵ and anypositive integer N,

$$\sum_{k=0}^{N} \|X^{k+1} - X^k\|^2 \leq \frac{2}{\epsilon}\sum_{k=0}^{N}(f(X^k) - f(X^{k+1})) \leq \frac{2}{\epsilon}f(X^0),$$

which implies that $\sum_{k=0}^{\infty} \|X^{k+1} - X^k\|^2 < \infty$ and thus $\lim_{k\to\infty} \|X^{k+1} - X^k\|^2 = 0$, that is $\{X^k\}$ is asymptotically regular.
C. Let $\{X^{k_j}\}$ be a convergent subsequence of $\{X^k\}$ and $\hat{X}$ be its limit point, that is to say, $X^{k_j} \to X^*$ as $k \to \infty$.
Combining with the limitation above, one can immediately derive

$$\widetilde{X}^{k_j} = X^{k_j} - A^T(AX^{k_j} - \mathcal{T}_3(\mathcal{M}))/L \to \hat{X} - A^T(A\hat{X} - \mathcal{T}_3(\mathcal{M}))/L =: \widetilde{X},$$

as $k_j \to \infty$. From **B**, the following inequality holds

$$\|X^{k_j+1} - \hat{X}\|^2 \leq \|X^{k_j+1} - X^{k_j}\|^2 + \|X^{k_j} - \hat{X}\|^2 \to 0,$$

as $k_j \to \infty$, which guarantees that $\{X^{k_j+1}\}$ also converges to $\hat{X}$. On the other side, by utilizing Lemma 3.1, we have

$$X^{k_j+1} = \max\left\{\widetilde{X}^{k_j} - (\lambda/L)W, 0\right\} \circ \operatorname{sgn}(\widetilde{X}^{k_j}) \to \max\left\{\widetilde{X} - (\lambda/L)W, 0\right\} \circ \operatorname{sgn}(\widetilde{X}),$$

as $k_j \to \infty$, which manifests $\hat{X} = \max\left\{\widetilde{X} - (\lambda/L)W, 0\right\} \circ \operatorname{sgn}(\widetilde{X})$. Therefore, from Theorem 3.3, any accumulation point of $\{X^k\}$ is a global minimizer of problem (2.8). □

According to the above conclusions, one can easily get the following corollary.

Corollary 3.5. *The sequence $\{\mathcal{S}^k\}$ generated by Algorithm 1 from any initial point $\mathcal{S}^0$ converges to $\hat{\mathcal{S}}$.*

Since numerical experiments in literature [6, 28], particularly in compressed sensing, have indicated that the reweighted ℓ_1 minimization performs exceptionally well in recovering sparse structured solutions, we modify Algorithm 1 (MWL1) as MIRL1 (i.e., Modified Iterative Reweighted ℓ_1 Algorithm), whose framework is written as the following table.

Table 2: The framework of the MIRL1.

Algorithm 2: Modified Iterative Reweighted ℓ_1 Algorithm (MIRL1)

Input: the observed data $\mathcal{M}$, compute $\mathcal{T}_3(\mathcal{M})^T$;

Initialize: $M, \lambda,\ \varepsilon_0 \geq 0,\ L,\ X^0$, and W^0;

For τ=1: M

Initialize $X^{\tau,1} = X^{\tau-1}$;

While $\|X^{\tau,k+1} - X^{\tau,k}\|^2 > \varepsilon_0$

Compute $\widetilde{X}^{\tau,k} = X^{\tau,k} - A^T(AX^{\tau,k} - \mathcal{T}_3(\mathcal{M}))/L$;

Compute $X^{\tau,k+1} = \max\{\widetilde{X}^{\tau,k} - (\mu/L)W^\tau, 0\} \circ \mathrm{sgn}(\widetilde{X}^{\tau,k})$;

End

Update $X^\tau = X^{\tau,k+1}$;

Update $W^{\tau+1}$ from $X^{\tau-1}, X^\tau$;

End

Output: X, compute $\mathcal{L} = fold_3(\mathbf{1}\mathbf{u}^{kT})$, $\mathcal{S} = fold_3(\mathcal{T}_3(\mathcal{S}^k))$.

By exploiting the method from Literature [28] to update the weight, which is different from classic iterative reweighted ℓ_1 algorithm [6], we take W^τ as (for any $i \neq 1$)

$$W_{ij}^\tau = \begin{cases} \left[\frac{|(H_{ij})_i^\tau| + \varepsilon}{\max_{ij\in(\Omega^\tau)^C}|H_{ij}^\tau|}\right]^{q-1}, & (i,j) \in \Omega^\tau; \\ 1, & (i,j) \in (\Omega^\tau)^C, \end{cases}$$

and $W_{1,j}^\tau = 0$, where

$$H^\tau = (\mathcal{T}_3(\mathcal{S}))^\tau - (\mathcal{T}_3(\mathcal{S}))^{\tau-1}, \quad k^\tau = |\mathrm{supp}(\mathcal{T}_3(\mathcal{S}))^\tau)|,$$

Ω^τ denotes the indices of the k^τ largest entries (in absolute) of H^τ and $(\Omega^\tau)^C$ denotes its complementary set; $(\mathcal{T}_3(\mathcal{S}))^\tau, (\mathcal{T}_3(\mathcal{S}))^{\tau-1}$ are part of $X^{\tau-1}, X^\tau$ respectively; and $0 < p < 1$, $\varepsilon > 0$ is sufficiently small.

Proposition 3.6. *For given $\lambda > 0, L > \lambda_{\max}(A^TA)$, and W^τ, the inter loops of Algorithm 2 have the global convergence, namely, any accumulation point of $\{X^{\tau,k}\}$ is a global minimizer of $\{\min \mu\|W^\tau \circ X\|_1 + \frac{1}{2}\|AX - \mathcal{T}_3(\mathcal{M})\|^2\}$.*

4 Numerical Experiments

In this section, we conduct numerical experiments on gray and color data to demonstrate the efficiency of the proposed algorithm. The experiments are conducted in Matlab version 8.2 with Intel Core I5 2.60GHz CPU and 8GB of RAM, and the data is available at the web site http://perception.i2r.astar.edu.sg/bk_model/bk_index.html and

http://homepages.inf.ed.ac.uk/-rbf/CAVIAR/, respectively. There exists noise in surveillance video in reality, hence we add Gaussian white noise to test our algorithm. We compare our algorithm with ADM [17] and FISTA [2].

4.1 Synthetic data

In Figure 1, we show some frames of the test gray surveillance video. This database has 80 sequences, each of which has 256-by-256 pixels. The background is a relatively static benchmark image, then we add some significant variations as foregrounds to those backgrounds.

Figure 1: Some frames of the cameraman video.

We set the stopping criterion as $RelErr < 10^{-3}$, where

$$RelErr := \max\left\{\frac{\|\mathcal{L}^k - \mathcal{L}^*\|^2}{\|\mathcal{L}^*\|^2}, \ \frac{\|\mathcal{S}^k - \mathcal{S}^*\|^2}{\|\mathcal{S}^*\|^2}\right\},$$

where $\mathcal{L}^*$ and $\mathcal{S}^*$ are ground-truths of background and foreground respectively. We also choose $\mu = 0.01$, $\varepsilon_0 = 1e-3$ as suggested in [26].

Figure 2 – Figure 4 show the computational results for ADM, FISTA and MIRL1. The first row is some frames of the cameraman video, the second row is some frames of the extracted background, and the third row is some frames of the extracted foreground. From Figure 2 – Figure 4, we know that the extracted backgrounds and foregrounds by ADM, FISTA and MIRL1 for the cameraman video are about the same. In Table 3, we report the computational results for the cameraman video. "Iteration", "Time" and "RelErr" represent the number of iterations, computing time in seconds, and relative error, respectively. The "RelErr" of MIRL1 is lower than that of ADM and FISTA, even though it costs more time.

Table 3: Computational results for cameraman video.

Algorithm	Iteration	Time	RelErr
ADM	16	10.74	1.56E-02
FISTA	13	12.69	6.41E-03
MIRL1	13	13.25	3.81E-03

4.2 Real data

Some example frames of the test real color surveillance video can be found in Figure 5. We extract 160 sequences, each of which has 288-by-384 pixels. The video has a relatively static background, and a relatively moving parts.

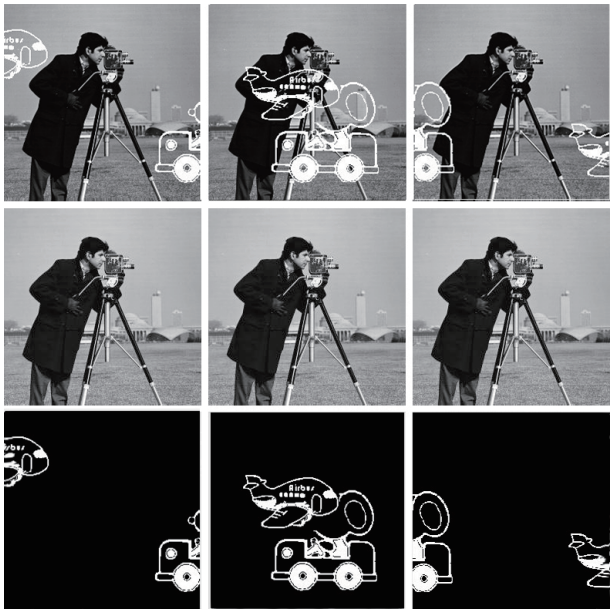

Figure 2: Computational results of ADM for cameraman video.

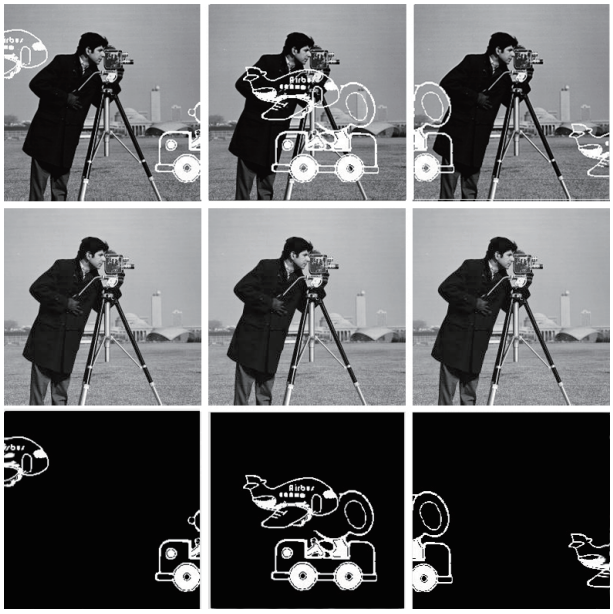

Figure 3: Computational results of FISTA for cameraman video.

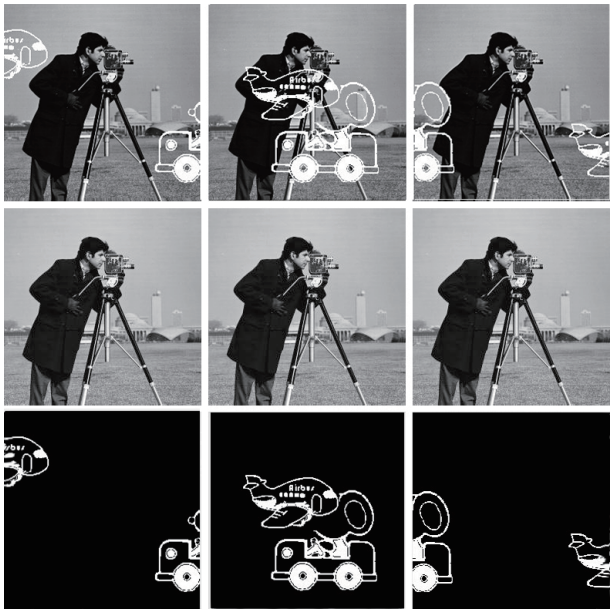

Figure 4: Computational results of MIRL1 for cameraman video.

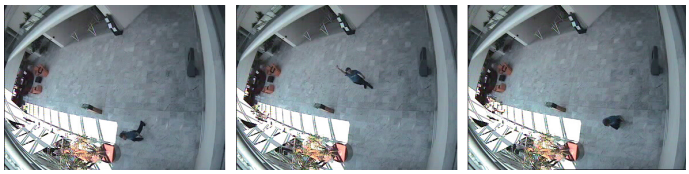

Figure 5: Some frames of the walk video.

For real color data, there don't exist ground-truths, so we choose the stopping criterion as

$$RelErr := \max\left\{\frac{\|\mathcal{L}^{k+1}-\mathcal{L}^{k}\|^2}{\|\mathcal{L}^{k}\|^2}, \frac{\|\mathcal{S}^{k+1}-\mathcal{S}^{k}\|^2}{\|\mathcal{S}^{k}\|^2}\right\},$$

where $\mathcal{L}^k, \mathcal{S}^k, \mathcal{L}^{k+1}, \mathcal{S}^{k+1}$ are k-th and $(k+1)$-th iteration value, respectively. Here, we also choose $\mu = 0.01$, and $\varepsilon_0 = 1e-3$.

The computational results are established in Figure 6–Figure 8. The first row is some frames of the walk video, the second row is some frames of the extracted background, and the third row is some frames of the extracted foreground. From Figure 6–Figure 8 and computational results, we find that MIRL1 gets a higher precision than ADM and FISTA. That is to say, MIRL1 extracts less artifacts, which is summarized in Table 4.

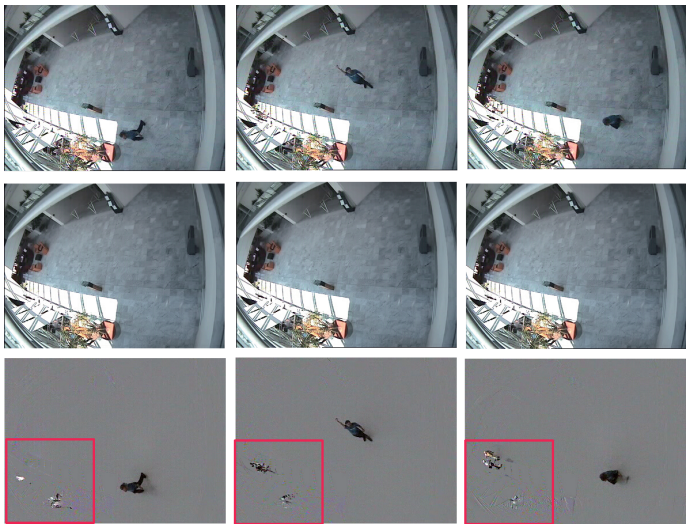

Figure 6: Computational results of ADM for walk video.

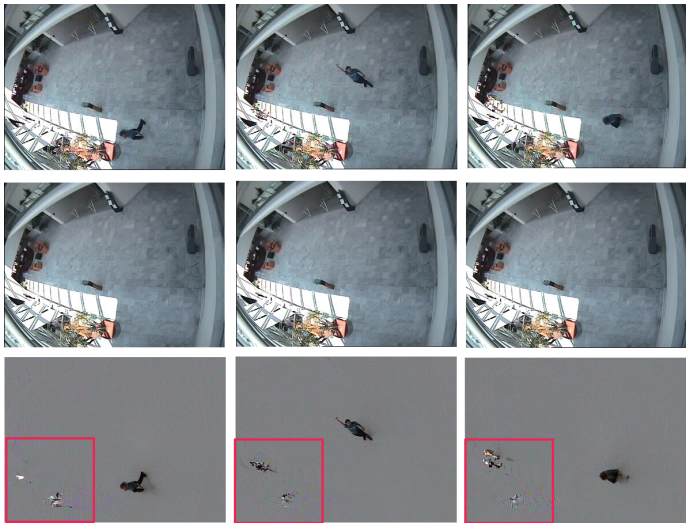

Figure 7: Computational results of FISTA for walk video.

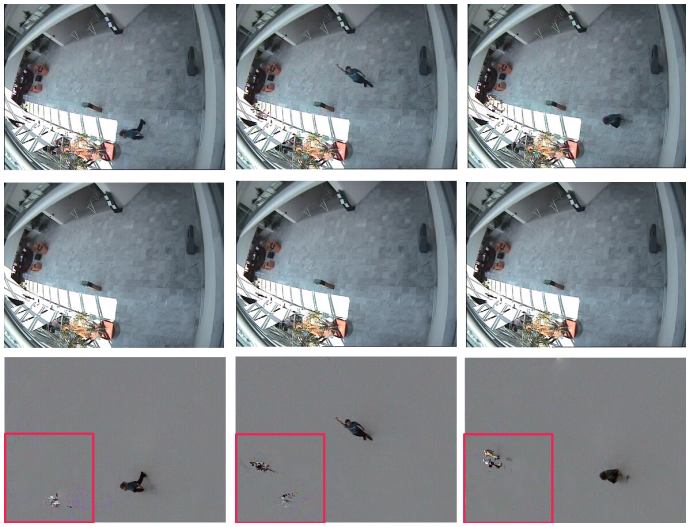

Figure 8: Computational results of MIRL1 for walk video.

Table 4: Computational results for walk video.

Algorithm	Iteration	Time	RelErr
ADM	25	24.83	2.54E-02
FISTA	19	30.29	8.35E-03
MIRL1	16	32.62	1.38E-03

5 Conclusions

In this paper, we focused on the application of tensor optimization in surveillance video. Since the background part was static, we introduced the definition of rank-min-one tensor, and proposed rank-min-one and sparse tensor decomposition for detecting the moving objects in surveillance video. Then, we established the modified iterative reweighted ℓ_1 algorithm for our model and gave its convergence analysis. Finally, numerical experiments, including synthetic and real data, illustrated the efficiency of our proposed algorithm.

Acknowledgement

The authors would like to thank the editors and two anonymous reviewers for numerous insightful comments and suggestions, which have greatly improved the paper. We also thank Professor Xiaoming Yuan from Hong Kong Baptist University for gracefully providing simulation data.

References

[1] K. Amin, W. Xu, A. Avestimehr and B. Hassibi, Weighted ℓ_1 minimization for sparse recovery with prior information, *IEEE Int. Sym. Inform. Theory* (2009) 483–487.

[2] A. Beck and T. Marc, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM J. Imaging Sci.* 2 (2009) 183–202.

[3] T. Bouwmans and E. Zahzah, Robust PCA via principal component pursuit: A review for a comparative evaluation in video surveillance, *Comput. Vision Imaging Underst.* 122 (2014) 22–34.

[4] E. Candès, X. Li, Y. Ma and J. Wright, Robust principal component analysis? *J. ACM* 58 (2011) 1–37.

[5] E. Candès and T. Tao, Decoding by linear programming, *IEEE Trans. Inform. Theory* 51 (2005) 4203–4215.

[6] E. Candès, M. Wakin and S. Boyd, Enhancing sparsity by reweighted ℓ_1 minimization, *J. Fourier Anal. Appl.* 14 (2008) 877–905.

[7] V. Cevher, A. Sankaranarayanan, M. Duarte, D. Reddy, R. Baraniuk and R. Chellappa, Compressive sensing for background subtraction, *Proc. ECCV* (2008) 155–168.

[8] V. Chandrasekaran, S. Sanghavi, P. Parrilo and A. Willsky, Rank-sparsity incoherence for matrix decomposition, *SIAM J. Optim.* 21 (2011) 572–596.

[9] D. Donoho, Compressed sensing, *IEEE Trans. Inform. Theory* 52 (2006) 1289–1306.

[10] A. Elgammal, R. Duraiswami, D. Harwood and L. Davis, Background and foreground modeling using nonparametric kernel density estimation for visual surveillance, *Proc. IEEE* 90 (2002) 1151–1163.

[11] N. Friedman and S. Russell, Image segmentation in video sequence: A probabilistic approach, *UAI* (1997) 175–181.

[12] S. Gandy, B. Recht and I. Yamada, Tensor completion and low-n-rank tensor recovery via convex optimization, *Inverse Problems* 27 (2011).

[13] D. Goldfarb and Z. Qin, Robust low-rank tensor recovery: models and algorithms, *SIAM J. Matrix Anal. Appl.* 35 (2014) 225–253.

[14] E. Hale, W. Yin and Y. Zhang, A Fixed-point continuation method for ℓ_1-regularized minimization with applications to compressed sensing, *CAAM* 07 (2007).

[15] B. Huang, C. Mu and D. Goldfarb et al., Provable low-rank tensor recovery, 2014.

[16] R. Jain and H. Nagel, On the analysis of accumulative differernce pictures from image sequences of real world scenes, *IEEE Trans. Pattern Anal. Mach. Intell.* 2 (1979) 206–214.

[17] H. Jiang, W. Deng and Z. Shen, Surveillance video processing using comprssive sensing, *Inverse Probl. Imaging* 6 (2012) 201–214.

[18] T. Kolda and B. Bader, Tensor decompositions and applications, *SIAM Rev.* 51 (2009) 457–464.

[19] X. Li, M. Ng and X. Yuan, Nuclear-norm-free variational models for background extraction from surveillance video, submitted to *IEEE Trans. Image Process.*.

[20] J. Liu, P. Musialski, P. Wonka and J. Ye, Tensor completion for estimating missing values in visual data, *ICCV* (2009) 2114–2121.

[21] S. Ma, D. Johnson, C. Ashby, et al., SPARCoC: a new framework for molecular pattern discovery and cancer gene identification, *PLoS ONE* to appear.

[22] A. Mcivor, Background subtraction techniques, *Image Vision Comput.* 1 (2000) 155–163.

[23] N. Obliver, B. Rosario and A. Pentland, A Bayesian computer vision system for modeling human interactions, *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (2000) 831–843.

[24] X. Xiu and L. Kong, Rank-one and sparse matrix decomposition for dynamic MRI, *Numer. Algebra Control Optim.* to appear.

[25] L. Yang, Z. Huang and X. Shi, A fixed point iterative method for low n-rank tensor pursuit, *IEEE Trans. Signal Process.* 61 (2013) 2952–2962.

[26] X. Yuan and J. Yang, Sparse and low-rank matrix decomposition via alternating direction methods, *Pac. J. Optim.* 9 (2013) 167–180.

[27] Y. Zhao and D. Li, Reweighted ℓ_1-minimization for sparse solutions to underdetermined linear systems. *SIAM J. Optim.* 22 (2012) 1065–1088.

[28] S. Zhou, N. Xiu, Y. Wang and L. Kong, Exact recovery for sparse signal via weighted ℓ_1 minimization, 2014.

Manuscript received 28 May 2014
revised 5 November 2014, 25 January 2014
accepted for publication 5 February 2015

Xianchao Xiu
Department of Applied Mathematics, School of Science
Beijing Jiaotong University, Beijing 100044, P. R. China
E-mail address: xianchaoxiu@163.com

Lingchen Kong
Department of Applied Mathematics, School of Science
Beijing Jiaotong University, Beijing 100044, P. R. China
E-mail address: konglchen@126.com