



Stochastic Decomposition into Low Rank and Sparse Tensor for Robust Background Subtraction

Sajid Javed, Thierry Bouwmans, Soon Ki Jung

► To cite this version:

Sajid Javed, Thierry Bouwmans, Soon Ki Jung. Stochastic Decomposition into Low Rank and Sparse Tensor for Robust Background Subtraction. 6th International Conference on Imaging for Crime Prevention and Detection (ICDP 2015), Jul 2015, Londres, United Kingdom. 10.1049/ic.2015.0105 . hal-01374218

HAL Id: hal-01374218

<https://hal.science/hal-01374218>

Submitted on 30 Sep 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Stochastic Decomposition into Low Rank and Sparse Tensor for Robust Background Subtraction

Sajid Javed¹, Thierry Bouwmans², and Soon Ki Jung¹

¹School of Computer Science and Engineering, Kyungpook National University
80 Daehakro, Buk-gu, Daegu, Republic of Korea
sajid@vr.knu.ac.kr, skjung@knu.ac.kr

²Laboratoire MIA (Mathématiques, Image et Applications)-
Université de La Rochelle, 1700, France
thierry.bouwmans@univ-lr.fr

Keywords: Tensor decomposition, Low-rank tensor, Stochastic optimization, Background/Foreground Separation.

Abstract

Background subtraction (BS) is a very important task for various computer vision applications. Higher-Order *Robust Principal Component Analysis* (HORPCA) based robust tensor recovery or decomposition provides a very nice potential for BS. The BG sequence is then modeled by underlying low-dimensional subspace called *low-rank* while the *sparse* tensor constitutes the foreground (FG) mask. However, traditional tensor based decomposition methods are sensitive to outliers and due to the batch optimization methods, high dimensional data should be processed. As a result, huge memory usage and computational issues arise in earlier approaches which are not desirable for real-time systems. In order to tackle these challenges, we apply the idea of stochastic optimization on tensor for robust *low-rank* and *sparse* error separation. Only one sample per time instance is processed from each unfolding matrices of tensor in our scheme to separate the *low-rank* and *sparse* component and update the low dimensional basis when a new sample is revealed. This iterative multi-dimensional tensor data optimization scheme for decomposition is independent of the number of samples and hence it reduces the memory and computational complexities. Experimental evaluations on both synthetic and real-world datasets demonstrate the robustness and comparative performance of our approach as compared to its batch counterpart without sacrificing the online processing.

1 Introduction

Background Subtraction (BS) is a very important task in video analysis. This low level task consists of accurate and efficient isolation of moving objects called “Foreground” (FG) from the static information called “Background” (BG) scene. The neat and clean BG model is then further used for high level computer vision and image processing tasks such as video compression, privacy, inpainting, augmented reality, computational photography, etc. [1] [10]. However, the task becomes more

complex when the data size grows since the real-world scenario requires larger data to be processed in a more efficient way.

Many algorithms have been proposed for BG/FG segmentation and several implementations are available in the BGS¹ library. Among them subspace learning model such as Higher-Order *Robust Principal Component Analysis* (HORPCA) based tensor decomposition provides a very nice framework for moving object detection. Tensor based decomposition methods [7] [19] also known as HORPCA have been recently proposed for exact *low-rank* recovery which exploits information using multidimensional N -arrays and provides an equivalent and more efficient framework than traditional RPCA-based approach [5]. LRS² library provides the implementation of many tensor based decomposition methods for FG detection. The BG sequence is considered using the underlying multi-dimensional subspace called *low-rank* tensor and *sparse* error constitutes the moving FG objects. Fig. 1 shows an example of BS using tensor decomposition, whereas Fig. 2 indicates only frontal slice of tensor of original sequences taken from CDnet [18] and iLIDS [3] dataset for better visualisation.

However, HORPCA based decomposition methods into *low-rank* and *sparse* component currently face some prominent challenges: First, the state of the art HORPCA [7] based approaches are based on batch optimization processing. In order to perform decomposition, a number of training frames are required to store in a memory before processing any data. Therefore, due to the high memory storage, they are not applicable to real-time processing. Second, HORPCA based algorithms are sensitive to outliers, i.e., *low-rank* tensor recovery is not always robust, especially for BG/FG separation where outliers always appear and hence it is not applicable for practical visual surveillance systems. For example, the last row in Fig. 2 depicts an example of outlier appearances in frontal slice of the tensor. Third, earlier HORPCA approaches need to access every frame for *Higher Order Singular Value Decomposition* (HOSVD) as a result the computational problems prevent them from efficient processing of big data.

¹<https://github.com/andrewssobral/bgslibrary>

²<https://github.com/andrewssobral/lrslibrary>

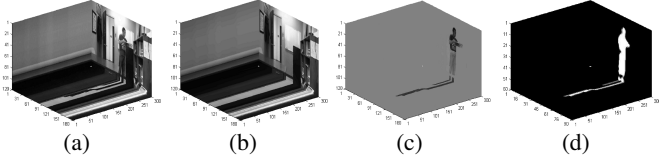


Figure 1. An example of input tensor decomposition: (a) input, (b) *low-rank*, (c) *sparse* tensor, and (d) foreground mask.

In order to address these major difficulties for BG/FG segmentation. This paper presents a robust recovery of *low-rank* tensor model for accurate FG segmentation. We briefly explain our methodology here. Traditional approaches are required to access each sample for HOSVD computation and therefore its nuclear norm tightly couples all the samples before processing. In this work, the stochastic (a.k.a online or iterative) optimization scheme where nuclear norm is re-formulated, recently proposed by Feng and Xu [6], is applied on each mode of the tensor. Our main technical contribution is an equivalent formulation of online RPCA [6] applied on tensor therefore we named it as, *Online Tensor Decomposition* (OTD) into *low-rank* and *sparse* component. OTD processes one sample per time instance from each vectorized tensor mode via online manners and hence reduces the memory usage and computational time. The *low-rank* tensor is then modeled by explicit product of basis and its coefficients, whereas the FG detection is obtained by thresholding the frontal slice of *sparse* tensor.

The rest of this paper is organized as follows. In Section 2, the related work is reviewed. Section 3 describes the proposed stochastic tensor decomposition for BS. Experimental results are discussed in Section 4, and finally Section 5 concludes our work.

2 Previous Work

In the literature, several algorithms have been proposed to cope with *low-rank* and *sparse* decomposition problem in computer vision. For example, Candes *et al.* [5] designed a very interesting framework called RPCA via Principal Component Pursuit (PCP) to decompose the given observation matrix into *low-rank* and *sparse* component. Excellent survey on BS using RPCA can be found in [2]. But these RPCA [5] matrix based decomposition methods used for BS [2] work only on single dimension and consider image as a vector and hence multidimensional data for efficient analysis can not be considered. In addition, the local spatial information sometimes lost and erroneous FG regions are obtained.

Therefore, tensor based algorithm called TSA (Tensor subspace learning) is developed by He *et al.* [8]. TSA is a very robust method for learning a subspace basis using multidimensional data but does not provide the convergence analysis. Wang *et al.* [17] proposed a convergent solution for tensor based *low-rank* learning. In addition, WTA (Window-based Tensor), STA (Streaming Tensor Analysis) and DTA (Dynamic Tensor Analysis) have been proposed by Sun *et al.* [15]. In [11], RSTD (Robust Subspace Tensor Decomposition) is developed for automatic robust subspace recovery using Block

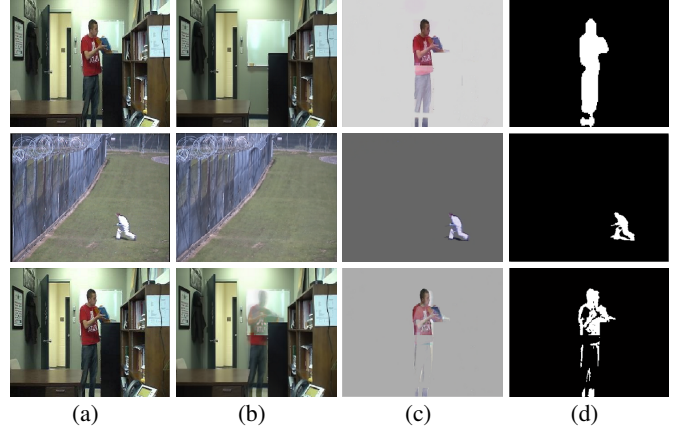


Figure 2. Frontal slice of tensor: (a) input, (b) *low-rank*, (c) *sparse* component, and (d) foreground mask.

Coordinate Descent (BCD) approach on unconstrained problem via variable splitting strategy, a number of computer vision applications such as image restoration, BS and face recognition are addressed in [11]. But parameters tuning and complexity of optimization method are the main drawbacks in RSTD. Donald and Qin [7] developed an extended *Alternating Direction Augmented Lagrangian* (ADAL) and HORPCA methods for robust tensor recovery. Convergence guarantee and proofs of each method are presented in [7]. Recently, Zhao *et al.* [19] proposed a Robust Bayesian Tensor Factorization (BRTF) scheme for incomplete tensor completion data. BRTF provides a very fast multi-way data convergence but tuning of annoying parameters and batch processing are the major difficulties in this approach.

All these matrix or tensor based decomposition methods discussed above are based on batch optimization mode. Therefore, a real time processing is loss due to the major challenges presented above. In order to tackle these problems, this paper presents a robust BS algorithm based on stochastic decomposition of *low-rank* and *sparse* component. We apply the idea of online optimization of RPCA [6] on tensor and update the individual basis iteratively followed by the processing of one frame per time instance using each unfolding matrices of tensor.

3 Proposed Methodology

In this section, stochastic tensor decomposition for robust BG/FG separation is presented. First, we describe some tensor basics and notations, then the proposed algorithm is presented in detail.

3.1 Tensor Basics and Notation

Tensor is considered as a multi-dimensional numerical array. The notations used in this paper are similar as used in [9]. The order of a tensor is the number of its dimensions (a.k.a *modes* or *ways*). The bold face calligraphic letters represent higher order tensors, (order ≥ 2) e.g., $\mathcal{X}$. Scalars are represented by lowercase letter e.g., x , whereas vectors and matrices are denoted by boldface lowercase and capital letters e.g., $\mathbf{x}$, and $\mathbf{X}$,

respectively.

Let say an N th order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, then its elements are denoted by symbolic name with indexes. For example, the element (i, j, k) of third-order tensor $\mathcal{X}$ is represented by $x_{i,j,k}$. The analogue of matrix rows and columns for the tensors are called *fibers*, which is obtained by fixing all indexes but one (a.k.a vector-values subtensor). A tensor $\mathcal{X}$ has column, row and tube fibers represented by $x_{:jk}$, $x_{i:k}$, and $x_{ij:}$. Similarly, *slices* of a tensor are two dimensional sub-array that can be obtained by fixing all indexes but two. A tensor $\mathcal{X}$ has *horizontal*, *lateral* and *frontal* slices indicated by $x_{i:}$, $x_{j:}$, and $x_{k:}$. Reshaping is always required for manipulation of tensors, a common operation is called matrix *unfolding* or *matricization* for reformatting tensors into matrix. For N th order tensor, its unfolding matrices are denoted by $\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_N$. Vectorization of tensor into vector is denoted by $\text{vec}(\mathcal{X})$. The squared Frobenius norm of a tensor is $\|\mathcal{X}\|_F^2 = \langle \mathcal{X}, \mathcal{X} \rangle$. More details on tensor operations can be found in [9].

3.2 Stochastic Tensor Decomposition

Let say that $\mathcal{Y}$ is an input N th order observation tensor, which is corrupted by *outliers*, say $\mathcal{E}$, then $\mathcal{Y}$ can be reconstructed by separating it into *low-rank* tensor $\mathcal{X}$ (corresponds to BG), and *sparse* error $\mathcal{E}$ (corresponds to FG objects), i.e., $\mathcal{Y} = \mathcal{X} + \mathcal{E}$, under the convex optimization framework developed in [7] as

$$\min_{\mathcal{X}, \mathcal{E}} \frac{1}{2} \sum_{i=1}^N \|\mathcal{Y}_i - \mathcal{X}_i - \mathcal{E}_i\|_F^2 + \lambda_1 \|\mathcal{X}_i\|_* + \lambda_2 \|\mathcal{E}_i\|_1, \quad (1)$$

where $\|\mathcal{X}_i\|_*$ and $\|\mathcal{E}_i\|_1$ denote the nuclear and l_1 norm of each mode- i unfolding matrices of $\mathcal{X}$ and $\mathcal{E}$, respectively. Efficient methods such as CANDECOMP/PARAFAC (CP)-decomposition and Tucker decomposition [9] (a.k.a HOSVD) are used for *low-rank* decomposition of tensor. In addition, APG, HORPCA-s based on ADAL and HORPCA-M based on I-ADAL are also developed in [7] to solve Eq.1. However, as mentioned above, these methods are based on batch optimization and not suitable for scalable data.

In this work, online optimization is considered to solve Eq.1. The major challenge is the computation of HOSVD because nuclear norm keeps all the samples tightly and therefore all samples are accessed during optimization at each iteration. Therefore, it suffers from high computational complexities. In contrast, an equivalent nuclear norm is used in this work for each mode- i unfolding matrices of $\mathcal{X}$, whose rank is upper bounded as shown in [13] as

$$\|\mathcal{X}_i\|_* = \inf_{\mathbf{L} \in \mathbb{R}^{p \times r}, \mathbf{R} \in \mathbb{R}^{n \times r}} \left\{ \frac{1}{2} (\|\mathbf{L}_i\|_F^2 + \|\mathbf{R}_i\|_F^2) \right. \\ \left. s.t. \mathcal{X}_i = \mathbf{L}_i \mathbf{R}_i^T \right\}, \quad (2)$$

where p denotes the dimension of each sample e.g., *width* $\times$ *height*, n is the number of samples and r is a rank. Eq. 2 shows that mode- i unfolding matrices of *low-rank* tensor $\mathcal{X}$ can be an explicit product of each low-dimensional subspace basis $\mathbf{L} \in$

$\mathbb{R}^{p \times r}$ and its coefficients $\mathbf{R} \in \mathbb{R}^{n \times r}$ and this re-formulated nuclear norm is shown in recent works [4], [13] [14]. Hence, Eq. 1 is re-formulated by substituting Eq.2 as

$$\min_{\mathcal{X}_1, \dots, \mathcal{X}_N, \mathbf{L} \in \mathbb{R}^{p \times r}, \mathbf{R} \in \mathbb{R}^{n \times r}, \mathcal{E}} \frac{1}{2} \sum_{i=1}^N \|\mathcal{Y}_i - \mathcal{X}_i - \mathcal{E}_i\|_F^2 + \frac{\lambda_1}{2} (\|\mathbf{L}_i\|_F^2 + \|\mathbf{R}_i\|_F^2) + \lambda_2 \|\mathcal{E}_i\|_1, \quad s.t. \mathcal{X}_i = \mathbf{L}_i \mathbf{R}_i^T. \quad (3)$$

For objective function minimization, avoiding the constraints in Eq.3 and put $\mathcal{X}_i = \mathbf{L}_i \mathbf{R}_i^T$ as

$$\min_{\mathcal{X}_1, \dots, \mathcal{X}_N, \mathbf{L} \in \mathbb{R}^{p \times r}, \mathbf{R} \in \mathbb{R}^{n \times r}, \mathcal{E}} \frac{1}{2} \sum_{i=1}^N \|\mathcal{Y}_i - \mathbf{L}_i \mathbf{R}_i^T - \mathcal{E}_i\|_F^2 + \quad (4)$$

$$\frac{\lambda_1}{2} (\|\mathbf{L}_i\|_F^2 + \|\mathbf{R}_i\|_F^2) + \lambda_2 \|\mathcal{E}_i\|_1, \quad (5)$$

where λ_1 and λ_2 are regularization parameters for *low-rank* and *sparsity* patterns. Eq 5 is the main equation for stochastic tensor decomposition which is not completely convex with respect to $\mathbf{L}$ and $\mathbf{R}$. However, Eq. 3 is the global optimal solutions to the original optimization problem in Eq. 2 as proved in [6]. The following cost function is required to optimize for solving Eq 3 as

$$f_n(\mathbf{L}) = \frac{1}{n} \sum_{i=1}^N \sum_{t=1}^n l(\mathcal{Y}_i^t, \mathbf{L}_i) + \frac{\lambda_1}{2n} \|\mathbf{L}_i\|_F^2, \quad (6)$$

where $\mathcal{Y}_i^t$ denotes i^{th} mode of a tensor $\mathcal{Y}$ at a time t instance given by

$$l(\mathcal{Y}_i^t, \mathbf{L}) = \min_{\mathbf{r}, \mathbf{e}} \|\text{vec}(\mathcal{Y}_i^t) - \mathbf{L} \mathbf{r} - \mathbf{e}\|_2^2 + \frac{\lambda_1}{2} \|\mathbf{r}\|_2^2 + \lambda_2 \|\mathbf{e}\|_1. \quad (7)$$

Finally, the objective function $l_t(\mathbf{L})$ for updating the mode- i basis $\mathbf{L}_i$ matrix of multidimensional subspace tensor $\mathcal{X}$ at a time t instance is given by

$$l_t(\mathbf{L}_i) = \frac{1}{n} \sum_{t=1}^n \left\{ \frac{1}{2} \|\text{vec}(\mathcal{Y}_i^t) - \mathbf{L}_i^t \mathbf{r}^t - \mathbf{e}^t\|_2^2 + \frac{\lambda_1}{2} \|\mathbf{r}^t\|_2^2 \right. \\ \left. + \lambda_2 \|\mathbf{e}^t\|_1 \right\} + \frac{\lambda_1}{2t} \|\mathbf{L}_i^t\|_F^2, \quad (8)$$

where $\mathbf{r}^t$ and $\mathbf{e}^t$ are vectors of coefficient and noise at a time t for matrix R_i , respectively, and mode- i matrix of *sparse* tensor $\mathcal{E}$. The main goal is to minimize the cost function Eq. 6 through stochastic optimization method as shown in Algorithm 1.

In case of BG modeling, one video frame at a time t is processed in an online manners. The coefficient $\mathbf{r}$, *sparse* matrix $\mathbf{e}$ and basis $\mathbf{L}$ are optimized in an iterative way. First, the coefficient $\mathbf{r}$ and noise $\mathbf{e}$ matrices are estimated with fixed random basis $\mathbf{L}$ by projecting one sample using Eq. 11. This subproblem in step. 6 requires to solve a following small-scale convex optimization problem at a time instance t as

$$\mathbf{r}^t = (\mathbf{L}^T \mathbf{L} + \lambda_1 \mathbf{I})^{-1} \mathbf{L}^T \{ \text{vec}(\mathcal{Y}_i^t) - \mathbf{e}^{t-1} \}, \quad (9)$$

$$\mathbf{e}^t = \begin{cases} \mathbf{M}^t(k) - \lambda_2, & \text{if } \mathbf{M}^t(k) > \lambda_2, \\ \mathbf{M}^t(k) + \lambda_2, & \text{if } \mathbf{M}^t(k) < \lambda_2, \\ 0, & \text{otherwise,} \end{cases} \quad (10)$$

where $\mathbf{M} = \text{vec}(\mathcal{Y}_i^t) - \mathbf{L}\mathbf{r}^t$ and $\mathbf{M}_t(k)$ is the k^{th} element in $\mathbf{M}$ at a time t . Second, the basis $\mathbf{L}^t$ is estimated using the E.q. 14 through minimizing the previously computed coefficients $\mathbf{r}$ and $\mathbf{e}$. These basis $\mathbf{L}^t$ for low-dimensional subspace learning is then updated using Algorithm 2 by the results of $\mathbf{r}$ and $\mathbf{e}$. If the rank r is given and basis L are estimated as above which is a fully rank, then L converges to the optimal solution asymptotically as compared to its batch counterpart as shown in [6].

Finally each i^{th} mode low-dimensional subspace tensor $\mathcal{X}$ is estimated by a multiple of basis $\mathbf{L}$ and coefficients $\mathbf{R}$. The BG sequence is then modeled by *low-rank* tensor $\mathcal{X}$ which changes at a time instance t , whereas the resulting *sparse* tensor $\mathcal{E}$ is obtained by the matricization of $\mathbf{e}$ entries. Finally, a hard thresholding scheme is applied on a *sparse* component to get the binary FG mask for BG/FG separation.

4 Experimental Evaluations

In this section, we present our experimental results in detail. We first evaluate the proposed algorithm performance on synthetic data then the qualitative and quantitative analysis using Change Detection (CDnet) [18] and Background Models Challenge (BMC) [16] dataset for real time BG/FG segmentation is presented in detail.

4.1 Evaluation on Synthetic Data

The proposed method is first quantitatively tested on synthetic data. For data evaluation, a true *low-rank* tensor $\mathcal{X}$ of size $30 \times 30 \times 30$ is generated by rank-3 factor matrices e.g., $\mathbf{Z}^n \in \mathbb{R}^{30 \times 3}$ where $n = 1, 2, 3$. Each factor matrix $\mathbf{Z}^n$ consists of three components such as $[\sin(\frac{2\pi n i_n}{30}), \cos(\frac{2\pi n i_n}{30}), \text{sgn}(\sin(0.5\pi i_n))]$. The first two components are different and third one is common in all modes. A random entries of this *low-rank* tensor $\mathcal{X}$ is corrupted by outliers from uniform distribution $U(-|H|, |H|)$ and small noise $N(0, 0.01)$ is also considered. We use a well known measure for evaluation called “Root Relative Square Error” (RRSE) given by $\frac{\|\hat{\mathcal{X}} - \mathcal{X}\|_2}{\|\mathcal{X}\|_2}$, where $\hat{\mathcal{X}}$ is a reconstructed *low-rank* tensor. We compare our RRSE performance with earlier methods based on batch optimization processing such as BRTF [19], CP-ARD [12], CP-ALS [9], HORPCA [7], and HOSVD [7], respectively. Fig. 3 shows the value of RRSE with results of reconstructed tensor $\hat{\mathcal{Y}}$. We consider two cases for robust tensor recovery for true data generation in Fig. 3. First, the magnitude is considered within a range of true data as shown in Fig. 3 (a). However, Fig. 3 (b) shows that the magnitude is taken larger

Algorithm 1 Stochastic Optimization Method

Input: $\mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$.

Initialize: $\mathcal{X} = \mathcal{E} = 0$, $\mathbf{L} \in \mathbb{R}^{p \times r}$ (initial basis), r , $\mathbf{A} \in \mathbb{R}^{r \times r}$, $\mathbf{B} \in \mathbb{R}^{p \times r}$, $\mathbf{r} \in \mathbb{R}^r$, $\mathbf{R} \in \mathbb{R}^{n \times r}$, $\mathbf{e} \in \mathbb{R}^p$, Unitary Matrix $\mathbf{I}$, $\lambda_1 = 0.25\lambda_2$, and $\lambda_2 = \frac{1}{\sqrt{\max(\text{size}(\mathcal{Y}))}}$.

- 1: **for** $t = 1$ to n **do** {Access each sample}
- 2: **for** $i = 1$ to N **do** {each tensor mode}
- 3: Access each frame from i^{th} mode of tensor $\mathcal{Y}$ by $\mathcal{Y}_i^t \leftarrow \text{unfold}(\mathcal{Y})$
- 4: Compute the coefficients $\mathbf{r}$ and noise $\mathbf{e}$ by projecting the new sample as

$$\{\mathbf{r}^t, \mathbf{s}^t\} = \underset{\mathbf{r}, \mathbf{s}}{\text{argmin}} \frac{1}{2} \|\mathcal{Y}_i^t - \mathbf{L}^{t-1} \mathbf{r} - \mathbf{s}\|_2^2 + \frac{\lambda_1}{2} \|\mathbf{r}\|_2^2 + \lambda_2 \|\mathbf{s}\|_1. \quad (11)$$

- 5: $\mathbf{R}(:, t) \leftarrow \mathbf{r}^t$. Compute the accumulation matrices $\mathbf{A}^t$ and $\mathbf{B}^t$ for each i^{th} mode

$$\mathbf{A}^t \leftarrow \mathbf{A}^{t-1} + \mathbf{r} \mathbf{r}^T, \quad (12)$$

$$\mathbf{B}^t \leftarrow \mathbf{B}^{t-1} + (\mathcal{Y}_i^t - \mathbf{s}^t) \mathbf{r}^T. \quad (13)$$

- 6: Compute $\mathbf{L}^t$ with previous iteration $\mathbf{L}^{t-1}$ and update the basis using Algorithm. 2

$$\mathbf{L}^t = \underset{\mathbf{L}}{\text{argmin}} \frac{1}{2} \text{Tr}[\mathbf{L}^T (\mathbf{A}^t + \lambda_1 \mathbf{I}) \mathbf{L}] - \text{Tr}(\mathbf{L}^T \mathbf{B}^t). \quad (14)$$

- 7: $\mathcal{L}_i^t \leftarrow \mathbf{L} \mathbf{R}^T$ (low-dimensional subspace for each unfold i^{th} mode)
- 8: $\text{vec}(\mathcal{E}_i^t) \leftarrow \mathbf{e}^t$ (*sparse* error)
- 9: **end for**
- 10: **end for**

Output: $\mathcal{X} = \frac{1}{N} \sum_{i=1}^N \mathcal{X}_i$, $\mathcal{E} = \sum_{i=1}^N \mathcal{E}_i$.

for corrupting some entries in true *low-rank*. In each case, the proposed method outperforms with previous HORPCA methods. However, it shows a comparative performance to its batch counter-part such as BRTF. The proposed scheme processes one frame per time instance and achieve almost a real-time processing, whereas BRTF is based on batch optimization strategy and hence it is not applicable for processing high dimensional data.

4.2 Evaluation on CDnet and BMC Dataset for BS

We have also evaluated our method for robust BS using synthetic and real videos of BMC and CDnet dataset. Due to the space limitations, we only present 5 selected synthetic videos of BMC dataset for qualitative analysis. The image size of each video sequence is 640×480 and half of the resolution is used. Both grayscale and color videos are evaluated. The size of grayscale input tensor $\mathcal{Y}$ is $320 \times 240 \times 200$ with 200 frames

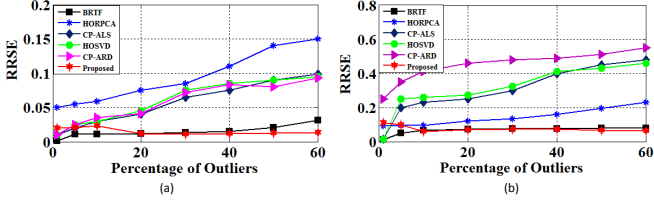


Figure 3. Performance of reconstructed *low-rank* tensor, (a) $O=\max(\text{vec}(\mathcal{X}))$, and (b) $O=10.\text{std}(\text{vec}(\mathcal{X}))$.

Algorithm 2 Basis Update

Input: $\mathbf{L} = [\mathbf{l}_1, \dots, \mathbf{l}_r] \in \mathbb{R}^{p \times r}$, $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_r] \in \mathbb{R}^{r \times r}$,
 $\mathbf{B} = [\mathbf{m}_1, \dots, \mathbf{m}_r] \in \mathbb{R}^{p \times r}$, $\tilde{\mathbf{A}} \leftarrow \mathbf{A} + \lambda_1 \mathbf{I}$.
1: **for** $j = 1$ to r **do** {access each column of $\mathbf{L}$ }
2: Update each column of basis matrix $\mathbf{L}$

$$\mathbf{l}_j \leftarrow \frac{1}{\tilde{\mathbf{A}}_{j,j}} (\mathbf{b}_j - \mathbf{L} \tilde{\mathbf{a}}_j) + \mathbf{l}_j \quad (15)$$

3: **end for**
4: **return** $\mathbf{L}$ (Updated basis for t^{th} mode)

and color video frame tensor as $320 \times 240 \times 3$. Stochastic tensor decomposition is applied on each video frame to separate $\mathcal{X}$ and $\mathcal{E}$ component. The value of rank $r = 10$ in Eq. 5 above. Additionally, median filtering is applied as a post-processing step to filter binary mask for morphological smoothing. Fig. 4 (i)-(m) shows the visual results of synthetic sequences from *Street* and *Rotary* category of BMC dataset [16].

From CDnet dataset [18], 8 real videos from category *Baseline* and *Thermal* are tested to validate the proposed scheme. The resolution of each sequence is $320 \times 240 \times 3$ and only color videos are tested. So the size of input tensor $\mathcal{Y}$ is $320 \times 240 \times 3 \times n$, where n is the number of input video frames and it varies from 500 to 4,000 according to the sequences presented in CDnet dataset. The *Thermal* category contains real videos under increasing percentage of practical camera noise. Fig. 4 (a)-(h) depicts the visual results on 8 real videos selected from *Baseline* and *Thermal* category.

The proposed scheme is also evaluated for quantitative analysis. A well known *F-measure* metric is computed as $F = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$ by comparing our results with available corresponding ground truth images. Table 1 shows that our proposed method is the top performer as compared to earlier batch optimization based algorithms. The “-” line in table 1 means that these methods are not applicable to large scale videos ($n > 500$) due to the batch optimization methods. Computational complexity is also observed during our experiments. Time is recorded in CPU time as $[hh : mm : ss]$ and table 2 shows the computational time of each method for first 100 frames with varying image resolution. All the experiments are conducted in Matlab R2013a with 3.40 GHz Intel core i5 processor with 4 GB RAM. These all robust experimental evaluations are the consequences of our proposed scheme.

Method	HORPCA	CP-ALS	BRTF	OTD
160×120	00:01:35	00:00:40	00:00:22	00:00:04
320×240	00:04:56	00:02:09	00:03:50	00:00:12

Table 2. Computational time according to different image resolutions.

5 Conclusion

In this paper, we proposed stochastic tensor decomposition for robust BS application. Experimental results show that we have achieved almost a real time processing since one frame is processed according to online optimization scheme. The coefficients and noise matrix are obtained using randomized basis then the basis is updated which is very robust against gross outliers. However, we have not integrated any robust features neither we applied this idea on highly dynamic BG scenes. Therefore, we will extend this work by fusing robust features such as disparity and texture information for more robust FG detection. Moreover, this idea can be used for other computer vision tasks such as online tracking using *low-rank sparse* tensor as a tracker.

6 Acknowledgement

This work is supported by the World Class 300 project, Development of HD video/network-based video surveillance system(10040370), funded by the Ministry of Trade, Industry, and Energy (MOTIE, Korea).

References

- [1] Thierry Bouwmans, Fida El Baf, Bertrand Vachon, et al. Statistical Background Modeling for foreground detection: A survey. *Handbook of Pattern Recognition and Computer Vision*, pages 181–199, 2010.
- [2] Thierry Bouwmans and El Hadi Zahzah. Robust PCA via Principal Component Pursuit: A review for a comparative evaluation in video surveillance. *Computer Vision and Image Understanding*, pages 22–34, 2014.
- [3] Home Office Scientific Development Branch. Imagery library for intelligent detection systems I-LIDS. In *Crime and Security, 2006. The Institution of Engineering and Technology Conference on*, pages 445–448, June 2006.
- [4] Samuel Burer and Renato DC Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95(2):329–357, 2003.
- [5] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust Principal Component Analysis? *Journal of the ACM (JACM)*, 58(3):11–37, 2011.
- [6] Jiashi Feng, Huan Xu, and Shuicheng Yan. Online Robust PCA via Stochastic Optimization. In *Advances in Neural Information Processing Systems*, pages 404–412, 2013.

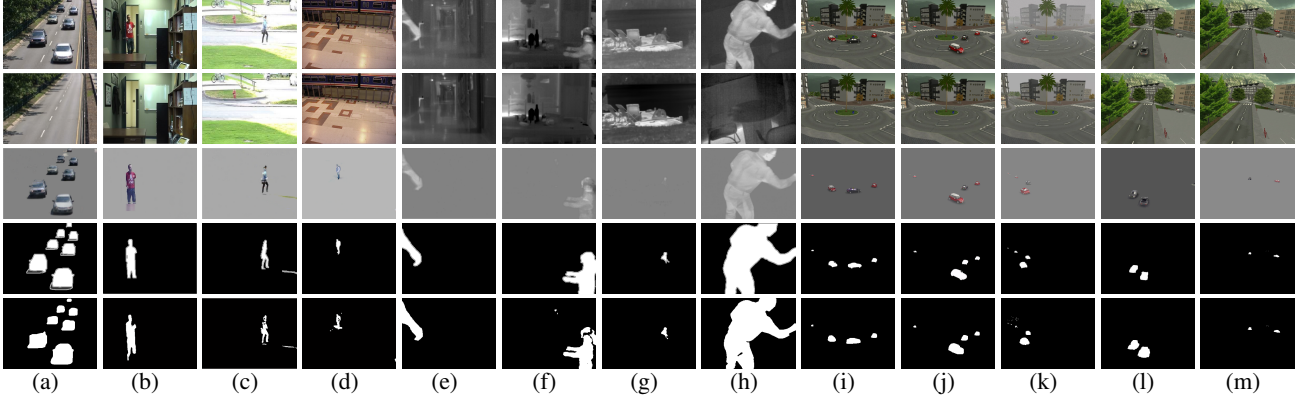


Figure 4. Qualitative results of CDnet [18] and BMC [16] dataset. From (a) to (d) *Baseline*, and (e) to (h) *Thermal* category of CDnet, and (i)–(m) *Synthetic* videos of BMC. From top to bottom, input, *low-rank*, *sparse*, ground truth, and binary mask of proposed methodology.

Table 1. Qualitative results of BMC [16] and CDnet [18] dataset: Average *F-measure* score in (%) on each video.

Method	BMC [16]										Average	CDnet [18]								Average
	Street					Rotary						Baseline				Thermal				
	112	212	312	412	512	122	222	322	422	522		highway	office	pedestrians	PETS2006	corridor	diningRoom	lakeSide	library	
HORPCA [7]	81.30	84.33	69.94	69.46	75.63	82.34	86.10	72.45	67.20	72.56	76.13	-	-	-	-	-	-	-	-	
CP-ALS [9]	84.70	84.21	78.65	76.32	79.12	85.36	81.23	71.60	78.05	80.52	79.97	-	-	-	-	-	-	-	-	
BRTF [19]	87.83	85.63	81.10	82.29	84.13	88.10	90.36	85.55	80.69	86.61	85.22	-	-	-	-	-	-	-	-	
Ours	91.65	87.10	84.80	81.61	85.93	93.35	90.21	87.65	81.20	84.63	86.50	88.36	85.92	90.63	74.41	86.30	80.68	76.58	94.66	

- [7] Donald Goldfarb and Zhiwei Qin. Robust low-rank tensor recovery: Models and algorithms. *SIAM Journal on Matrix Analysis and Applications*, 35(1):225–253, 2014.
- [8] Xiaofei He, Deng Cai, and Partha Niyogi. Tensor subspace analysis. In *Advances in neural information processing systems*, pages 499–506, 2005.
- [9] Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.
- [10] Wang Hee Lee, Kuntal Sengupta, and Rajeev Sharma. Augmented reality with occlusion rendering using background-foreground segmentation and trifocal tensors. In *Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on*, volume 2, pages II–93. IEEE, 2003.
- [11] Yin Li, Junchi Yan, Yue Zhou, and Jie Yang. Optimum subspace learning and error correction for tensors. In *Computer Vision–ECCV 2010*, pages 790–803. Springer, 2010.
- [12] Morten Mørup and Lars Kai Hansen. Automatic relevance determination for multi-way models. *Journal of Chemometrics*, 23(7-8):352–363, 2009.
- [13] Benjamin Recht, Maryam Fazel, and Pablo A Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010.
- [14] Jasson DM Rennie and Nathan Srebro. Fast maximum margin matrix factorization for collaborative prediction. In *Proceedings of the 22nd international conference on Machine learning*, pages 713–719. ACM, 2005.
- [15] Jimeng Sun, Dacheng Tao, Spiros Papadimitriou, Philip S Yu, and Christos Faloutsos. Incremental tensor analysis: Theory and applications. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 2(3):11, 2008.
- [16] Antoine Vacavant, Thierry Chateau, Alexis Wilhelm, and Laurent Lequière. A Benchmark Dataset for Outdoor Foreground/Background Extraction. In *Computer Vision–ACCV 2012 Workshops*, pages 291–300. Springer, 2013.
- [17] Huan Wang, Shuicheng Yan, Thomas S Huang, and Xiaou Tang. A convergent solution to tensor subspace learning. In *IJCAI*, pages 629–634, 2007.
- [18] Yi Wang, Pierre-Marc Jodoin, Fatih Porikli, Janusz Konrad, Yannick Benezeth, and Prakash Ishwar. Cdnet 2014: An expanded change detection benchmark dataset. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on*, pages 393–400. IEEE, 2014.
- [19] Qibin Zhao, Guoxu Zhou, Liqing Zhang, Andrzej Cichocki, and Shun-ichi Amari. Robust Bayesian Tensor Factorization for Incomplete Multiway data. *arXiv preprint arXiv:1410.2386*, 2014.