Online Stochastic Tensor Decomposition for Background Subtraction in Multispectral Video Sequences

Andrews Sobral
University of La Rochelle, France
Lab. L3I/MIA
andrews.sobral@univ-lr.fr

Sajid Javed, Soon Ki Jung
Kyungpook National University, Korea
School of Computer Science and Engineering
sajid@vr.knu.ac.kr, skjung@knu.ac.kr

Thierry Bouwmans, El-hadi Zahzah
University of La Rochelle, France
Lab. MIA, Lab. L3I
thierry.bouwmans@univ-lr.fr, elhadi.zahzah@univ-lr.fr

Abstract

Background subtraction is an important task for visual surveillance systems. However, this 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, and in some cases, in a continuous manner. Until now, most of background subtraction algorithms were designed for mono or trichromatic cameras within the visible spectrum or near infrared part. Recent advances in multispectral imaging technologies give the possibility to record multispectral videos for video surveillance applications. Due to the specific nature of these data, many of the bands within multispectral images are often strongly correlated. In addition, processing multispectral images with hundreds of bands can be computationally burdensome. In order to address these major difficulties of multispectral imaging for video surveillance, this paper propose an online stochastic framework for tensor decomposition of multispectral video sequences (OSTD). First, the experimental evaluations on synthetic generated data show the robustness of the OSTD with other state of the art approaches then, we apply the same idea on seven multispectral video bands to show that only RGB features are not sufficient to tackle color saturation, illumination variations and shadows problem, but the addition of six visible spectral bands together with one near infra-red spectra provides a better background/foreground separation.

1. Introduction

Background subtraction (BS) is usually the first step to detect moving objects in many fields of computer vision applications such as video surveillance (to detect persons, vehicles, animals, etc.), human-computer interface, motion detection and multimedia applications [24]. This basic operation consists of separating the moving objects called "foreground" (FG) from the static information called "background" (BG). However, in most cases, the background model is not always static due to the complexity of natural scenes: wind in the trees, moving water, waves, etc. In recent decades, a large number of algorithms have been proposed for background subtraction [24] and several implementations can be found in the BGS[1] [22] library.

However, this task becomes more complex when the data size grows (i.e. massive multidimensional data) since the real-world scenario requires larger data to be processed in a more efficient way, and in some cases, in a continuous manner (streaming data). Until now, most of background subtraction algorithms were designed for mono (i.e. graylevel) or trichromatic cameras (i.e. RGB) within the visible spectrum or near infrared part (NIR). Recent advances in multispectral imaging technologies give the possibility to record multispectral videos for video surveillance applications [1].

The primary advantage of multispectral cameras for video surveillance is the possibility to take into account the spatial (or spatiotemporal) relationships among the different spectra in a neighbourhood, allowing more elaborate spectral-spatial (and -temporal) models for a more accurate segmentation. However, the primary disadvantages are cost and complexity due its massive and multidimensional characteristics.

Usually a multispectral video consist of a sequence of multispectral images sensed from contiguous spectral bands. Each multispectral image can be represented as a

[1] https://github.com/andrewssobral/bgslibrary

three-dimensional data cube or *tensor*, and we call *frame* the measurements corresponding to a single spectral band (frontal slice of the tensor). Due to the specific nature of these data, many of the bands within multispectral images are often strongly correlated. In addition, processing multispectral images with hundreds of bands can be computationally burdensome.

In order to address these major difficulties of multispectral imaging for video surveillance (in particular, the detection of moving objects), this paper propose an online stochastic framework for tensor decomposition of multispectral video sequences (OSTD). In short, the main contributions of this paper are:

- An online stochastic framework for tensor decomposition to deal with multi-dimensional and streaming data.

- And, the use of multispectral video sequences instead of standard mono/trichromatic images, enabling a better background subtraction.

First, we start with the related works (Section 2), and next we describe the notation conventions used in this paper (Section 3). A brief introduction to tensors are detailed in Section 4, and the proposed method is shown in Section 5. Finally, in Sections 6 and 7, the experimental results are shown as well as conclusions.

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.* [4] designed an interesting framework called RPCA via Principal Component Pursuit (PCP) to decompose an observation matrix into *low-rank* and *sparse* components. The observation matrix is represented as: $\mathbf{M} = \mathbf{L} + \mathbf{S}$ where $\mathbf{L}$ is a low-rank matrix and $\mathbf{S}$ is a matrix that can be sparse or not. The low-rank minimization concerning $\mathbf{L}$ offers a suitable framework for background modeling due to the correlation between frames. Minimizing $\mathbf{L}$ and $\mathbf{S}$ implies that the background is approximated by a low-rank subspace that can gradually change over time, while the moving foreground objects constitute the correlated sparse outliers which are contained in $\mathbf{S}$. A survey on RPCA applied for BS can be found in Bouwmans and Zahzah [2].

However, these RPCA matrix based decomposition methods used for BS works 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 foreground regions are obtained. Some authors use a tensor representation to solve this problem. Wang and Ahuja

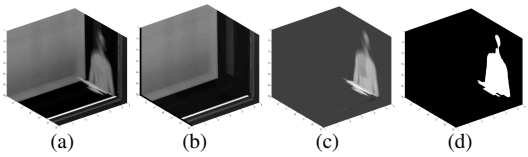

Figure 1: An example of tensor decomposition: (a) input, (b) *low-rank*, (c) *sparse* tensor, and (d) foreground mask.

[27] propose a rank-R tensor approximation which can capture spatiotemporal redundancies in the tensor entries. He *et al.*[11] present a tensor subspace analysis algorithm which detects the intrinsic local geometrical structure of the tensor space by learning a lower dimensional tensor subspace. The algorithm learn a subspace basis using multidimensional data but does not provide the convergence analysis. Sun *et al.* [25] introduce a general framework, incremental tensor analysis (ITA), which efficiently computes a compact summary for high-order and high-dimensional data, and also reveals the hidden correlations. However, Li *et al.* [12] explains that the previous work cannot be applied to background modeling and object tracking directly. To solve this problem, Li *et al.* [12] propose a high-order tensor learning algorithm called incremental rank-(R1,R2,R3) tensor-based subspace learning. This online algorithm builds a low-order tensor eigenspace model where the mean and the eigenbasis are updated adaptively. However, the previous method uses only the gray-scale and color information. In some situations, these features are insufficient to perform a robust foreground detection. To deal with this situation, Sobral *et al.* [21] propose an incremental tensor subspace learning approach that uses a multi-feature model and updates the *low-rank* component incrementally. In [16], RSTD (Robust Subspace Tensor Decomposition) is developed for automatic robust subspace recovery using Block Coordinate Descent (BCD) approach on unconstrained problem via variable splitting strategy, and some computer vision applications such as image restoration, BS and face recognition are addressed in [16]. But the parameters tuning and the complexity of the optimization method are the main drawbacks in the RSTD. Otherwise, Tran *et al.* [26] proposed a tensor-based method for video anomaly detection applying the Stable PCP [4] decomposition in each tensor mode. The proposed method is a one-shot framework to determine which frames are anomalous in a video. Moreover, Donald and Qin [7] developed some Higher-order RPCA (HoRPCA) methods for robust tensor recovery. Convergence guarantee and proofs of each method are presented in [7]. Recently, Zhao *et al.* [29] proposed a Robust Bayesian Tensor Factorization (BRTF) scheme for incomplete tensor completion data. BRTF provides a fast multi-way data convergence but tunning of annoying parameters and batch processing are the major difficulties of this approach.

All these tensor-based decomposition methods discussed above are based on generalization of matrix based decomposition problems, and the majority of them work as a batch optimization processing. In addition, most of incremental tensor subspace learning approaches apply matrix SVD in the unfolded matrices. However, the matrix factorization step in SVD is computationally very expensive, especially for large matrices. Therefore, a real time processing is sacrificed due to the major challenges discussed above. In order to address these problems, this paper proposes a robust and fast online tensor-based algorithm for RGB videos as well for MSVS (multispectral video sequences). The proposed algorithm is based on stochastic decomposition of *low-rank* and *sparse* components. The idea of online stochastic RPCA optimization was previously proposed by Feng. *et al.* [5] and Goes *et al.* [6], and it was successfully applied to background subtraction in [13, 14]. In this work we extend this approach for tensor analysis. The stochastic optimization is applied on each mode of the tensor and the individual basis are updated iteratively followed by the processing of one video frame per time instance. In addition, a comparison of RGB and MSVS is provided which shows that visible together with NIR spectral bands provide an improved foreground estimation as compare to RGB features.

3. Basic Notations

This paper follows the notation conventions in multilinear and tensor algebra as in [15, 8]. Scalars are denoted by lowercase letters, e.g., x; vectors are denoted by lowercase boldface letters, e.g., $\mathbf{x}$; matrices by uppercase boldface, e.g., $\mathbf{X}$; and tensors by calligraphic letters, e.g., $\mathcal{X}$. In this paper, only real-valued data are considered.

4. Tensor Introduction

A $tensor$ can be considered as a multidimensional or N-way array. As in [15, 8], an Nth-order tensor is denoted as: $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, where $I_n (n = 1, \ldots, N)$. Each element in this tensor is addressed by $\mathcal{X}_{i_1,\ldots,i_n}$, where $1 \leq i_n \leq I_N$. The $order$ of a tensor is the number of dimensions, also know as ways or modes [15]. A tensor $\mathcal{X}$ has column, row and tube fibers represented by $\mathcal{X}_{:jk}$, $\mathcal{X}_{i:k}$, and $\mathcal{X}_{ij:}$. Similarly, *slices* of a tensor are two dimensional subarray that can be obtained by fixing all indexes but two. A tensor $\mathcal{X}$ has *horizontal*, *lateral* and *frontal* slices indicated by $\mathcal{X}_{i::}$, $\mathcal{X}_{j::}$, and $\mathcal{X}_{k::}$. By unfolding a tensor along a mode, a tensor's unfolding matrix corresponding to this mode is obtained. This operation is also known as mode-n matricization[2]. For a Nth-order tensor $\mathcal{X}$, its unfolding matrices are denoted by $\mathcal{X}^{(1)}, \mathcal{X}^{(2)}, \ldots, \mathcal{X}^{(N)}$. A more general review of tensor operations can be found in [15].

[2]Can be regarded as a generalization of the mathematical concept of vectorization.

5. Online 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 decompostion [15] (a.k.a HOSVD) are used for *low-rank* decomposition of tensors. 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 or streaming data.

In this work, an 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 [18] as:

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

where p denotes the dimension of each sample e.g., $width \times height$, n is the number of samples and r is the 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 [3], [18], [19]. Hence, Eq. 1 is re-formulated by substituting Eq.2 by:

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

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

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

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

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 [5].The following cost function is required to optimize for solving Eq. 3 as:

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

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

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

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:

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

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 (i.e. RGB image) at a time t is processed in an online manner. 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:

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

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

where $\mathbf{M}=vec(\mathcal{Y}_i^t)-\mathbf{L}\mathbf{r}^t$ 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 Eq. 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 [5].

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.

6. Experimental Evaluations

In this section, we present our experimental results in detail. We first evaluate the proposed method performance on synthetic generated data then the qualitative and quantitative analysis on MSVS are presented for BS application.

6.1. Evaluation on Synthetic Data

The proposed method is first quantitatively tested on synthetic data. For data evaluation, a true *low-rank* tensor $\mathcal{L}$ of size $30\times30\times30$ is generated by rank-3 factor matrices e.g., $\mathbf{Y}^k\in\mathbb{R}^{30\times3}$ where $k=1,2,3$. Each factor matrix $\mathbf{Y}^k$ consists of three components such as $[\sin(\frac{2\pi n i_n}{30}),\cos(\frac{2\pi n i_n}{30}),sgn(\sin(0.5\pi i_n))]$. The first two components are different and third one is common in all modes. A random entries of $\mathcal{L}$ 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 $\frac{||\widehat{\mathcal{L}}-\mathcal{L}||_2}{||\mathcal{L}||_2}$, where $\widehat{\mathcal{L}}$ is the estimated *low-rank* tensor. We compare our RRSE performance with other state of the art methods such as BRTF [29], CP-ARD [17], CP-ALS [15], HORPCA [7] and HOSVD [7] respectively. Fig. 2 shows the value of RRSE with a results of recovered tensor $\widehat{\mathcal{L}}$. We consider two cases for robust tensor recovery for true data generation in Fig. 2. First, the magnitude is considered within a range of true data as shown in Fig. 2 (a). However, Fig. 2 (b) shows that the magnitude is taken larger for corrupting some entries in true *low-rank*. In each case, the proposed method shows a very significant improvements as compare to its batch counter-part such as BRTF.

Algorithm 1 Online Stochastic Tensor Decomposition

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

Initialize: $\mathcal{X} = \mathcal{E} = 0$ (*low-rank* and *sparse* components), $\mathbf{L} \in \mathbb{R}^{p \times r}$ (initial basis), r, $\mathbf{A} \in \mathbb{R}^{r \times r}$, $\mathbf{B} \in \mathbb{R}^{p \times r}$, $\mathbf{r} \in \mathbb{R}^r$, $\mathbf{R} \in \mathbb{R}^{n \times r}$, $\mathbf{e} \in \mathbb{R}^p$, $\mathbf{I} \in \mathbb{R}^{r \times r}$ (unitary matrix), $\lambda_1 = \frac{1}{\sqrt{\max(size(\mathcal{Y})}}$, and $\lambda_2 = 10\lambda_1$.

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 $\mathcal{Y}_i^t \leftarrow unfold(\mathcal{Y})$
4: Compute the coefficients $\mathbf{r}$ and noise $\mathbf{e}$ by projecting the new sample as:

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

5: $\mathbf{R}(:, \mathbf{t}) \leftarrow \mathbf{r}^t$. 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

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

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

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

Algorithm 2 Basis Update

Input: $\mathbf{L} = [\mathbf{l}_1, ..., \mathbf{l}_r] \in \mathbb{R}^{p \times r}$, $\mathbf{A} = [\mathbf{a}_1, ..., \mathbf{a}_r] \in \mathbb{R}^{r \times r}$, $\mathbf{B} = [\mathbf{m}_1, ..., \mathbf{m}_r] \in \mathbb{R}^{p \times r}$, $\widetilde{\mathbf{A}} \leftarrow \mathbf{A} + \lambda_1 \mathbf{I}$.

1: **for** $j = 1$ to r **do** {access each column of $\mathbf{L}$}
2: Update each column of basis matrix $\mathbf{L}$

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

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

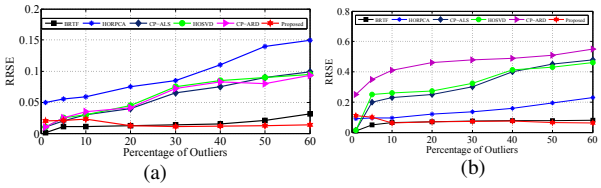

Figure 2: Performance of reconstructed *low-rank* tensor. (a) O=max(vec($\mathcal{L}$)), and (b) O=10.std(vec($\mathcal{L}$)).

6.2. BS of Multispectral Video Sequences

We evaluate the proposed method on a set of MSVS data set [1]. This is a first data set on MSVS[3] available for research community in background subtraction. The main purpose of this data set is not to show the robustness of the BS methods but the integration of multispectral information shows efficient FG/BG separation when color saturation and illumination variation issues occur. A set of both qualitative and quantitative results are presented.

The MSVS data set contains a set of 5 video sequences with 7 multispectral bands (6 visible spectrum and 1 NIR spectra). Each sequence presents a well known BS challenges such as color saturation and dynamic background. Fig. 3 shows the result from RGB image, 6 visible spectrum and 1 NIR spectral band together with visible spectra. Fig. 5 shows the visual comparison of the proposed approach for BS task over three scenes of MSVS data set. Fig. 6 (a) and (b) shows the visual results of these sequences using individual band with RGB features. This qualitative evaluation shows that BS using stochastic tensor decomposition on 7 multispectral bands together with visible spectra provides the best FG segmentation.

The proposed method is also tested for quantitative analysis. MSVS data set contains an image size of 658×492 for each band and $658 \times 492 \times 3$ for RGB image. So, the size of the input tensor $\mathcal{A}$ with 7 multispectral bands is $658 \times 492 \times 7$ for each video frame. A well-known *F-measure* metric is computed for each video sequence with its available ground truth images. Table. 2 shows a fair comparison of RGB and 7 multispectral bands (MSB). The average *F-measure* score is computed for each video by comparing our results with 3 other methods such as CP-ALS [15], HORPCA [7], and BRTF [29]. The experimental evaluations show that the proposed methodology outperforms the other approaches.

The proposed scheme processes each multispectral or RGB image (third-order tensor) per time instance reaching almost a real-time processing, whereas CP-ALS, HORPCA, and BRTF are based on batch optimization strategy. Due to this limitation, the CP-ALS, HORPCA, and BRTF were

[3] http://ilt.u-bourgogne.fr/benezeth/projects/ICRA2014/

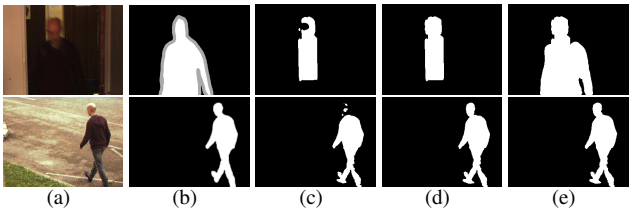

Figure 3: FG results on 1^{st} and 2^{nd} videos of the MSVS data set [1]. (a) input image, (b) ground truth, (c) results from RGB, (d) 6 visible bands, and (e) 1 NIR spectral band.

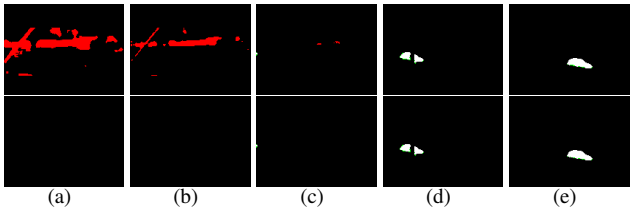

Figure 4: FG results on the 3^{rd} video of the MSVS data set [1]. From top to bottom: basis initialization with UDRN and BRP. From left to right, the FG mask at: (a) frame 1, (b) frame 5, (c) frame 10, (d) frame 15, and (e) frame 20.

applied for each 100 frames at time of the whole video sequence (fourth-order tensor). In this paper, the parameter r in the Algorithm 1 was defined experimentally as 10. For CP-ALS, the rank was defined as 50 for better visual results. For HORPCA and BRTF, we used its default parameters. To obtain the foreground mask, the sparse component $\mathcal{E}$ needs to be thresholded. First, we calculate the mean of $\mathcal{E}$ along the third dimension, generating a matrix $\mathbf{E}$, then a hard threshold function is applied by:

$$\mathbf{FG} = \begin{cases} 1 & \text{if } (0.5\mathbf{E})^2 > \beta \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

where $\beta = 0.5(std(vec(\mathbf{E})))^2$, and $std(.)$ denotes the standard deviation of a data vector.

6.3. Basis Initialization with BRP

Bilateral Random Projections (BRP) was first proposed by Zhou and Tao [31] as a fast *low-rank* approximation method for dense matrices. The effectiveness and the efficiency of BRP was verified previously in the GoDec [30] algorithm for *low-rank* and *sparse* decomposition. Given r bilateral random projections of an $m \times n$ dense matrix $\mathbf{X}$, the low-rank approximation $\mathbf{L}$ can be rapidly built by:

$$\mathbf{L} = \mathbf{Y}_1(\mathbf{A}_2^T\mathbf{Y}_1)^{-1}\mathbf{Y}_2^T \tag{17}$$

where $\mathbf{Y}_1 = \mathbf{X}\mathbf{A}_1$, $\mathbf{Y}_2 = \mathbf{X}^T\mathbf{A}_2$, and $\mathbf{A}_1 \in \mathbb{R}^{n\times r}$ and $\mathbf{A}_2 \in \mathbb{R}^{m\times r}$ are random matrices.

In this section, we evaluate the robustness of BRP for the basis initialization instead of the traditional uniformly distributed random numbers (UDRN). As expected, Fig. 4 shows a fast background modeling convergence for the first 20 video frames on the 3^{rd} video of the MSVS data set [1]. As can be seen, BRP enable a fast and effective *low-rank* approximation, reducing the amount of false positive pixels in the background model initialization task. Finally, the power scheme modification proposed by Zhou and Tao [31] can accelerate the *low-rank* recovery when the singular values of $\mathbf{X}$ decay slowly.

Size	HORPCA	CP-ALS	BRTF	Proposed
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 1: Computational time according to different image resolutions.

6.4. Computational Time

Computational complexity is also observed during our experiments. The time is recorded in CPU time as $[hh : mm : ss]$ and Table 1 shows the computational time of each method for the first 100 frames varying the image resolution. The algorithms were implemented in MATLAB (R2014a) running on a laptop computer with Windows 7 Professional 64 bits, 2.7 GHz Core i7-3740QM processor and 32Gb of RAM. The MATLAB implementation of the proposed approach is available in `https://github.com/andrewssobral/ostd`, and the implementation of the selected algorithms are available in the LRS[4] [23] library.

7. Conclusion

In this paper, we proposed an online stochastic tensor decomposition algorithm for robust BS application. Experimental results show that the proposed methodology outperforms the other approaches, and we have achieved almost a real time processing since one video frame is processed according to online optimization scheme. As previously discussed, the basis initialization with BRP can accelerate the *low-rank* approximation reducing the amount of false positive pixels in the background model initialization step. In addition, the basis is updated incrementally making it more robust against gross outliers. Moreover, the stochastic optimization applied on each mode of the tensor can be replaced by other incremental subspace learning approaches such as GRASTA [10], GOSUS [28], ReProCS [9] and incPCP [20]. Finally, this idea can be used as an online tracking using the *low-rank* and *sparse* components as a tracker.

[4] `http://github.com/andrewssobral/lrslibrary`

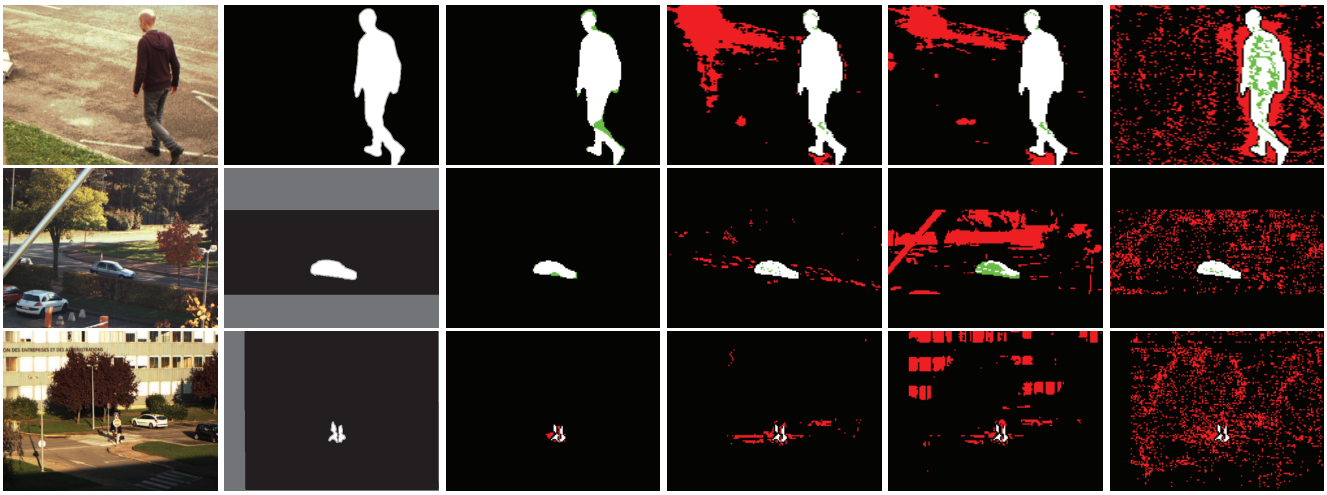

Figure 5: Visual comparison of background subtraction results over three scenes of the MSVS data set [1]. From left to right: (a) input image, (b) ground truth, (c) proposed approach, (d) BRTF [29], (e) HORPCA [7], and (f) CP-ALS [15]. The true positives (TP) pixels are in white, true negatives (TN) pixels in black, false positives (FP) pixels in red and false negatives (FN) pixels in green.

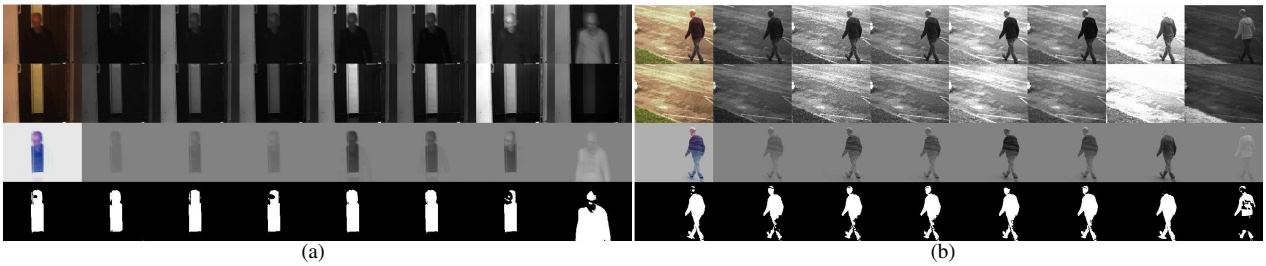

Figure 6: Visual results of the proposed method on each RGB and multispectral bands. From top to bottom: input image, *low-rank* component, *sparse* component, and the foreground mask. From left to right: RGB image, set of 7 visible, and 1 NIR spectrum are shown in each of columns respectively.

Table 2: MSVS data set [1]: *Comparison of average F-measure score in (%) with other approaches.*

Methods	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	**Avg**
CP-ALS [15]	RGB 58.69 MSB 71.61	RGB 71.25 MSB 83.50	RGB 51.32 MSB 68.54	RGB 60.21 MSB 78.63	RGB 49.35 MSB 66.97	RGB 58.16 MSB 73.85
HORPCA [7]	RGB 63.23 MSB 80.65	RGB 78.52 MSB 84.79	RGB 55.69 MSB 68.12	RGB 67.56 MSB 77.56	RGB 58.80 MSB 74.47	RGB 64.76 MSB 77.11
BRTF [29]	RGB 68.56 MSB 85.30	RGB 79.21 MSB 89.63	RGB 63.56 MSB 68.11	RGB 73.22 MSB 84.65	RGB 62.51 MSB 77.91	RGB 70.32 MSB 82.76
Proposed	RGB **78.63** MSB **93.65**	RGB **85.96** MSB **95.17**	RGB **79.56** MSB **90.64**	RGB **76.32** MSB **89.29**	RGB **71.23** MSB **92.66**	RGB **76.69** MSB **92.28**

8. Acknowledgments

The authors gratefully acknowledge the financial support of CAPES (Brazil) for granting a PhD scholarship to the first author.

References

[1] Y. Benezeth, D. Sidibe, and J. B. Thomas. Background subtraction with multispectral video sequences. In *ICRA*, 2014.

[2] T. Bouwmans and E. Zahzah. Robust PCA via Principal Component Pursuit: A review for a comparative evaluation in video surveillance. *CVIU*, pages 22–34, 2014.

[3] S. Burer and R. D. Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95(2):329–357, 2003.

[4] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust Principal Component Analysis? *Journal of the ACM*, 58(3):11–37, 2011.

[5] J. Feng, H. Xu, and S. Yan. Online Robust PCA via Stochastic Optimization. In *NIPS*, 2013.

[6] J. Goes, T. Zhang, R. Arora, and G. Lerman. Robust Stochastic Principal Component Analysis. In *AISTATS*, 2014.

[7] D. Goldfarb and Z. Qin. Robust low-rank tensor recovery: Models and algorithms. *SIAM Journal on Matrix Analysis and Applications*, 35(1):225–253, 2014.

[8] L. Grasedyck, D. Kressner, and C. Tobler. A literature survey of low-rank tensor approximation techniques. 2013.

[9] H. Guo, C. Qiu, and N. Vaswani. An online algorithm for separating sparse and low-dimensional signal sequences from their sum. *IEEE Trans. on Signal Processing*, 62(16):4284–4297, 2014.

[10] J. He, L. Balzano, and A. Szlam. Incremental gradient on the grassmannian for online foreground and background separation in subsampled video. In *CVPR*, 2012.

[11] X. He, D. Cai, and P. Niyogi. Tensor subspace analysis. In *NIPS*, 2005.

[12] W. Hu, X. Li, X. Zhang, X. Shi, S. Maybank, and Z. Zhang. Incremental tensor subspace learning and its applications toforeground segmentation and tracking. *IJCV*, 91(3):303–327, 2011.

[13] S. Javed, S. Ho Oh, A. Sobral, T. Bouwmans, and S. Ki Jung. OR-PCA with MRF for robust foreground detection in highly dynamic backgrounds. In *ACCV*, pages 284–299, 2014.

[14] S. Javed, A. Sobral, T. Bouwmans, and S. Ki Jung. OR-PCA with dynamic feature selection for robust background subtraction. In *30th Annual ACM Symposium on Applied Computing*, pages 86–91, 2015.

[15] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Review*, 2008.

[16] Y. Li, J. Yan, Y. Zhou, and J. Yang. Optimum subspace learning and error correction for tensors. In *ECCV*, pages 790–803. Springer, 2010.

[17] M. Mørup and L. K. Hansen. Automatic relevance determination for multi-way models. *Journal of Chemometrics*, 23(7-8):352–363, 2009.

[18] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010.

[19] J. D. Rennie and N. Srebro. Fast maximum margin matrix factorization for collaborative prediction. In *ICML*, pages 713–719, 2005.

[20] P. Rodriguez and B. Wohlberg. A matlab implementation of a fast incremental principal component pursuit algorithm for video background modeling. In *ICIP*, pages 3414–3416, 2014.

[21] A. Sobral, C. Baker, T. Bouwmans, and E. Zahzah. Incremental and multi-feature tensor subspace learning applied for background modeling and subtraction. In *ICIAR*. 2014.

[22] A. Sobral and T. Bouwmans. BGS Library: A library framework for algorithms evaluation in foreground/background segmentation. In *Background Modeling and Foreground Detection for Video Surveillance*. CRC Press, Taylor and Francis Group.

[23] A. Sobral, T. Bouwmans, and E. Zahzah. LRS Library: low-rank and sparse tools for background modeling and subtraction in videos. In *Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing*. CRC Press, Taylor and Francis Group.

[24] A. Sobral and A. Vacavant. A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. *CVIU*, 122(0):4–21, 2014.

[25] J. Sun, D. Tao, S. Papadimitriou, P. S. Yu, and C. Faloutsos. Incremental tensor analysis: Theory and applications. *ACM Trans. on KDD*, 2(3):11, 2008.

[26] L. Tran, C. Navasca, and J. Luo. Video detection anomaly via low-rank and sparse decompositions. In *Image Processing Workshop (WNYIPW), 2012 Western New York*, pages 17–20, Nov 2012.

[27] H. Wang and N. Ahuja. Rank-r approximation of tensors using image-as-matrix representation. In *CVPR*, volume 2, pages 346–353 vol. 2, June 2005.

[28] J. Xu, V. Ithapu, L. Mukherjee, J. M. Rehg, and V. Singh. GOSUS: Grassmannian Online Subspace Updates with Structured-sparsity. In *ICCV*, 2013.

[29] Q. Zhao, Z. L. Zhou, G. and, A. Cichocki, and S. Amari. Robust bayesian tensor factorization for incomplete multiway data. *CoRR*, abs/1410.2386, 2014.

[30] T. Zhou and D. Tao. Godec: Randomized low-rank & sparse matrix decomposition in noisy case. In *ICML*, pages 33–40, June 2011.

[31] T. Zhou and D. Tao. Bilateral random projections. In *IEEE Int. Symposium on Information Theory*, pages 1286–1290, 2012.