

RESEARCH

Open Access

Object detection oriented video reconstruction using compressed sensing

Bin Kang^{1*}, Wei-Ping Zhu^{1,2} and Jun Yan¹

Abstract

Moving object detection plays a key role in video surveillance. A number of object detection methods have been proposed in the spatial domain. In this paper, we propose a compressed sensing (CS)-based algorithm for the detection of moving object in video sequences. First, we propose an object detection model to simultaneously reconstruct the foreground, background, and video sequence using the sampled measurement. Then, we use the reconstructed video sequence to estimate a confidence map to improve the foreground reconstruction result. Experimental results show that the proposed moving object detection algorithm outperforms the state-of-the-art approaches and is robust to the movement turbulence and sudden illumination changes.

Keywords: Compressed sensing; Low-rank optimization; Moving object detection; Moving turbulence mitigation

1 Introduction

With the strong market demand on sensor networks for video surveillance purpose, the design of multimedia sensors equipped with high-resolution video acquisition systems to adapt to particular environment and the limited bandwidth is of crucial importance. In the multimedia sensor networks, the video sequences captured are first encoded and then transmitted to the processing center for video analysis. Moving object detection, aiming to locate and segment interesting objects in a video sequence, is a key to video surveillance.

A common approach for detecting moving objects, called background subtraction (BS) [1], is to estimate a background model first, and then compare video frames with the background model to detect the moving objects. When processing real video surveillance sequences, BS algorithms face with several challenges such as sudden illumination changes, movement turbulence, etc. [2]. A sudden illumination change strongly affects the appearance of the background and thus causes false foreground subtraction. Movement turbulence may contain (1) periodical or irregular turbulences such as waving trees and water ripples; (2) the objects being suddenly introduced or removed from the scene. It is still an open problem to

eliminate the movement turbulence due to its complex structure. Recently, Tsai et al. [3] proposed a fast background subtraction scheme using independent component analysis (ICA) for object detection. This scheme is tolerant of sudden illumination changes in indoor surveillance videos. Zhang et al. [4] proposed a kernel similarity modeling method for motion detection in complex and dynamic environments. This approach is robust to simple movement turbulence. Kim et al. [5] proposed a fuzzy color histogram (FHC)-based background subtraction algorithm to detect the moving object in the dynamic background. This algorithm can minimize the color variations generated by background motion. Chen et al. [6] suggested a hierarchical background model based on the fact that the background images consist of different objects whose conditions may change frequently. In the same year, Han et al. [7] proposed a piecewise background model which integrates color, gradient, and Haar-like features to handle spatiotemporal variations. This model is robust to the illumination change and shadow effect. All the aforementioned BS algorithms operate in the spatial domain and require a large amount of training sequences to estimate a background model. The training process always imposes high computational complexity, so it actually limits the application of BS algorithms in the multimedia sensor networks.

Compressed sensing (CS) [8-10] is a recently proposed sampling method which states that if a signal is sparse, it

* Correspondence: greatisgood@126.com

¹College of Communication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

Full list of author information is available at the end of the article

can be faithfully reconstructed from a small number of random measurements. The number of measurements required by CS is much smaller than that required by Nyquist sampling rate. CS can perform image sensing and compression simultaneously with low computational complexity. It has the superiority in reducing the computational cost of video encoder [11]. Due to its advantages, CS has become an attractive solution in object detection. One early attempt of using CS algorithm for object detection is to utilize the sampled measurements of the image background to train an object silhouette firstly, and then use the trained object silhouette to detect the moving object [12]. This algorithm needs a large amount of storage and computation for training the object silhouette, which is not suitable for real-time multimedia sensor networks. In 2012, Jiang et al. [13] proposed an object detection model to perform low-rank and sparse decomposition by using the compressed measurements. Although this model adapts to the limited bandwidth of multimedia sensor networks, it is not robust to the movement turbulence and sudden illumination change because the wavelet transform coefficients of the video sequence are not sparse when encountering with the movement turbulence. In 2013, Yang et al. [14] proposed a CS-based algorithm for object detection. This algorithm can exactly and simultaneously reconstruct the video foreground and background by using only 10% of sampled measurements. However, it still uses the wavelet transform as [13] does to achieve sparse decomposition. This causes false foreground reconstruction in the movement turbulence and sudden illumination change. Write et al. [15] proposed an algorithm called compressive principal component pursuit for analyzing the performance of the natural convex heuristic of solving the problem as to how to recover the low-rank matrix and the sparse component from a small set of linear measurement. This algorithm can be used to achieve object detection in the compressed domain. In this paper, we propose a new CS-based algorithm for detecting moving object. We firstly use three-dimensional circulant sampling method to obtain sampled measurement, based on which we reconstruct simultaneously the video foreground and background by solving an optimization problem. The main contributions of this paper are as follows:

1. There is a key problem as to how to obtain a robust video foreground reconstruction result using the compressed measurement. In order to solve this problem, we first propose a new object detection model to simultaneously reconstruct the video foreground, background, and video sequence using a small number of compressed measurements. Then, we use the reconstructed video sequence to estimate

a confidence map, which is used to further refine the foreground reconstruction result.

2. An efficient alternating algorithm is proposed for solving the minimization problem of the new object detection model. We prove that the alternating algorithm is guaranteed to yield a feasible background, foreground, and video reconstruction result.

The paper is organized as follows: Section 2 discusses how to solve the key problem in the CS-based object detection algorithm. Section 3 develops an alternating algorithm for solving the new object detection model. The experimental results of the proposed approach are then given in Section 4. Finally, conclusion is provided in Section 5.

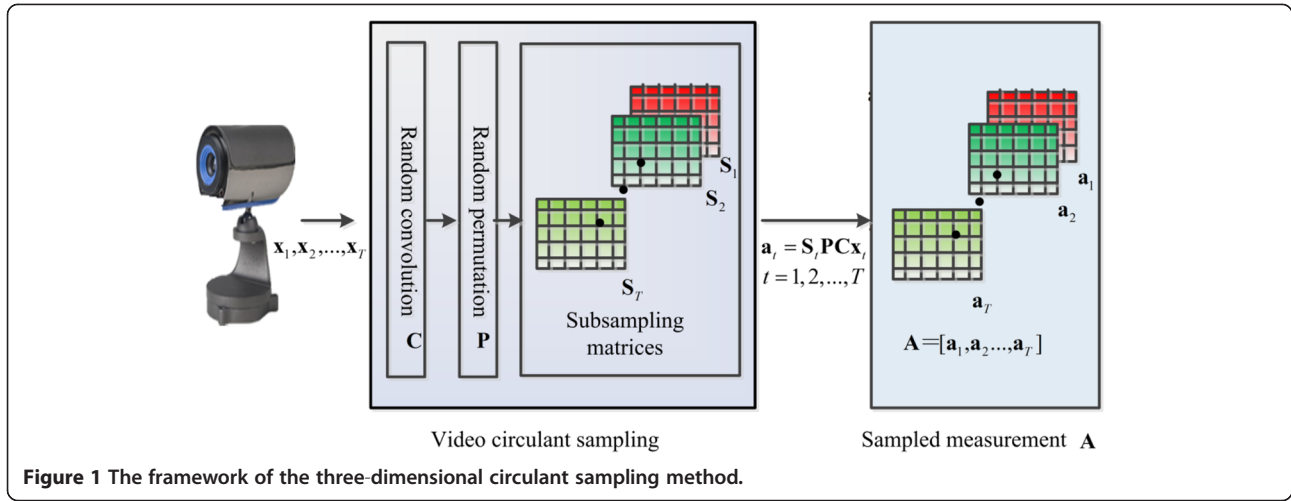
2 Problem formulation

The authors of [16] have proposed a three-dimensional circulant sampling method, as shown in Figure 1, which can perform video sensing and compression simultaneously with low computational complexity and easy hardware implementation. This method achieves video compression in two steps: the first step is random convolution, which yields circulant measurements $\mathbf{C}\mathbf{x}_t$ through convolving the original vectorized video frames \mathbf{x}_t ($t = 1, 2, \dots, T$) with a circulant matrix \mathbf{C} . The second step is random subsampling, which aims to reduce the number of circulant measurements $\mathbf{C}\mathbf{x}_t$. In this step, a random permutation is first applied to each vector $\mathbf{C}\mathbf{x}_t$ by a permutation matrix \mathbf{P} . Then the permuted vectors (measurements) $\mathbf{P}\mathbf{C}\mathbf{x}_t$ are each subsampled utilizing a subsampling matrix \mathbf{S}_t to generate the compressed (dimension-reduced) measurements $\mathbf{a}_t = \mathbf{S}_t\mathbf{P}\mathbf{C}\mathbf{x}_t$. In the figure, the whole compressed measurements are denoted by matrix $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_T]$.

Given the sampled measurement matrix \mathbf{A} , how to reconstruct the foreground and background becomes a key problem in the CS-based object detection. In 2009, Candes et al. proposed a robust principal component analysis (RPCA) model to simultaneously reconstruct the video foreground and background by solving the following minimization problem:

$$\min_{\mathbf{B}, \mathbf{F}} \|\mathbf{B}\|_* + \lambda \|\mathbf{F}\|_1 \quad s. t. \quad \mathbf{X} = \mathbf{B} + \mathbf{F} \quad (1)$$

where $\mathbf{X} \in R^{(MN) \times T}$ is the original video sequence, and \mathbf{B} and \mathbf{F} represent the background and foreground of the video, respectively. There are two drawbacks with the RPCA model. (1) RPCA cannot reconstruct \mathbf{B} and \mathbf{F} using the sampled measurement \mathbf{A} directly because the original video sequence \mathbf{X} is required in object detection. Obviously, the requirement for the original video reconstruction imposes a high computational complexity.



(2) In RPCA, the foreground reconstruction result is robust only to the corruption that has a sparse distribution [17,18]. In the real-world video sequence, however, there rarely exists the movement turbulence that is sparse in nature.

The so-called three-dimensional total variation (*TV3D*) has recently been proposed for CS-based video reconstruction [16], which can exploit both intra-frame and inter-frame correlations of the video sequence. The advantage of *TV3D* is that it can guarantee the performance of video reconstruction result with a low computational complexity ($O(3 \times MN \times T)$). The *TV3D* model is formulated as:

$$TV3D(\mathbf{X}) = \|D_1 \mathbf{X}\|_1 + \|D_2 \mathbf{X}\|_1 + \rho \|D_3 \mathbf{X}\|_1 \quad (2)$$

where D_1 and D_2 are, respectively, the horizontal and vertical difference operators within a frame, and D_3 is the time-varying difference operator.

In order to detect the moving object from the sampled measurement directly, we propose a new object detection model, by combining *TV3D* and RPCA, that can simultaneously reconstruct foreground, background, and video sequence. The proposed object detection model is described as:

$$\begin{aligned} \min_{\mathbf{B}, \mathbf{F}, \mathbf{X}} TV3D(\mathbf{X}) + \gamma * rank(\mathbf{B}) + \eta \|\mathbf{F}\|_1 \\ s. t. \Phi \mathbf{X} = \mathbf{A}, \mathbf{X} = \mathbf{B} + \mathbf{F} \end{aligned} \quad (3)$$

where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T]$ represents the original video sequence to be reconstructed $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_T]$ is the background, $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_T]$ is the foreground (moving object), and Φ is the measurement matrix. Since the accuracy of the reconstructed background and foreground relies on the performance of the video reconstruction result, the *TV3D* is used to enhance the quality of the

reconstructed video. As mentioned earlier, *TV3D* has a low computational complexity (see (2)), while (3) gives a similar computational complexity as RPCA does. Therefore, problem (3) is less insensitive to the variable initialization, and we can initialize \mathbf{X} , \mathbf{B} and \mathbf{F} as zero matrices. Note that solving the minimization problem of $rank(\mathbf{B})$ in (3) is NP-hard due to its nonconvexity and discontinuous nature [17]. We would like to relax the $rank(\mathbf{B})$ function through a nuclear norm, leading (3) to:

$$\begin{aligned} \min_{\mathbf{B}, \mathbf{F}, \mathbf{X}} \sum_{i=1}^3 a_i \|D_i \mathbf{X}\|_1 + \gamma \|\mathbf{B}\|_* + \eta \|\mathbf{F}\|_1 \\ s. t. \Phi \mathbf{X} = \mathbf{A}, \mathbf{X} = \mathbf{B} + \mathbf{F} \end{aligned} \quad (4)$$

The difference between problem (4) and the 3DCS model in [16] is that: the 3DCS model is aimed to give a high video reconstruction result, where not only *TV3D* is used for video reconstruction but also the nuclear norm is adopted to make use of the low-rank property of the video sequence in the wavelet domain. Problem (4) in this paper is, however, aimed to exactly reconstruct the video foreground and background using a small number of sampled measurements. To achieve this goal, we employ *TV3D* to guarantee the exact low-rank and sparse decomposition.

By solving problem (4), we can obtain the reconstructed foreground $\hat{\mathbf{F}}$, background $\hat{\mathbf{B}}$, and the video sequence $\hat{\mathbf{X}}$. Since the reconstructed $\hat{\mathbf{F}}$ is not robust to strong movement turbulence, Borenstein et al. have proposed in [19] an algorithm to achieve an excellent image segmentation performance by using a confidence map to identify the image region. Inspired by this idea, we use the reconstructed video sequence $\hat{\mathbf{X}}$ to construct a confidence map denoted as $\mathbf{O} = [\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_T]$, where the

element of O is 0 or 1. We then use O to further improve the reconstructed foreground $\hat{\mathbf{F}}$ through $\odot \hat{\mathbf{F}}$, where \odot denotes the Hadamard (point-wise) product. Note that the confidence map is a binary matrix, in which the location of the movement turbulence is set to 0 and the location of the moving object is set to 1.

In real-world video surveillance, movement turbulence is repetitive and locally centered [20,21], which can be modeled by Gaussian distribution [22,23]. In this paper, we utilize the following mixed Gaussian model to estimate the intensity distribution of a pixel undergoing movement turbulence [22].

$$f(x_{ij}) = \omega f_1(x_{ij}, \mu_x, \sigma_x) + (1-\omega) f_2(x_{ij}, \mu_p, \Sigma_p) \quad (5)$$

where $f(x_{ij})$ represents the probability density of a pixel x_{ij} at j th element in the i th column of $\hat{\mathbf{X}}$, ω is the weight of the two Gaussian models, μ_x and σ_x are the mean and the standard deviation, which are estimated by the EM algorithm, and μ_p and Σ_p are the mean and the covariance matrix, which are estimated from the particle trajectory of x_{ij} [22]. Particle trajectory aims to capture the deformation caused by movement turbulence, which can be obtained by using Lagrangian particle trajectory advection approach [24,25].

The confidence map is obtained as follows: we first estimate each pixel's probability density $f(x_{ij})$ using (5), then we decide which pixels belong to the movement turbulence and which ones belong to the moving object using an threshold θ . If $f(x_{ij}) > \theta$, we set it as 1. Otherwise, we set it as 0. The obtained binary matrix is the final confidence map.

3 Reconstruction algorithm

In problem (4), we generalize the process of video compression as $\mathbf{a}_t = \Phi \mathbf{x}_t$. Since we use \mathbf{P} , \mathbf{C} , and \mathbf{S}_t ($t = 1, 2, \dots, T$) to generate the compressed measurement \mathbf{A} (see Figure 1), we should use the specific form $\mathbf{r}_t = \mathbf{C} \mathbf{x}_t$ and $\mathbf{S}_t \mathbf{P} \mathbf{r}_t = \mathbf{a}_t$ ($t = 1, 2, \dots, T$) to replace the $\Phi \mathbf{X} = \mathbf{A}$ in (4) and rewrite it as:

$$\begin{aligned} \min_{\mathbf{B}, \mathbf{F}, \mathbf{G}_i} \sum_{i=1}^3 \alpha_i \|\mathbf{G}_i\|_1 + \gamma \|\mathbf{B}\|_* + \eta \|\mathbf{F}\|_1 \\ \text{s.t. } \mathbf{G}_i = \mathbf{D}_i \mathbf{X}, \mathbf{X} = \mathbf{B} + \mathbf{F}, \mathbf{R} = \mathbf{C} \mathbf{X}, \mathbf{S}_t \mathbf{P} \mathbf{r}_t = \mathbf{a}_t \quad (t = 1, 2, \dots, T) \end{aligned} \quad (6)$$

where $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_T]$ is the circulant measurement.

Next, we propose an alternating algorithm for the reconstruction of \mathbf{X} , \mathbf{B} , and \mathbf{F} in (6). Each iteration of the alternating algorithm contains two steps: R-step, which aims at reconstructing the original video \mathbf{X} ; and S-step, which is to segment background and foreground.

In R-step, we reconstruct \mathbf{X} by solving the following problem:

$$\begin{aligned} \min_{\mathbf{G}_i} \sum_{i=1}^3 \alpha_i \|\mathbf{G}_i\|_1 \text{ s.t. } \mathbf{G}_i = \mathbf{D}_i \mathbf{X}, \mathbf{R} = \mathbf{C} \mathbf{X}, \mathbf{S}_t \mathbf{P} \mathbf{r}_t \\ = \mathbf{a}_t \quad (t = 1, 2, \dots, T) \end{aligned} \quad (7)$$

We adopt the augmented Lagrange multiplier (ALM) algorithm [26] to solve problem (7). The augmented Lagrange function of (7) is given by:

$$\begin{aligned} \mathcal{L}(\mathbf{X}, \mathbf{G}_i, \mathbf{R}, \boldsymbol{\lambda}_i, \mathbf{v}) = \sum_{i=1}^3 \left(\|\mathbf{G}_i\|_1 + \frac{\beta_i}{2} \|\mathbf{G}_i - \mathbf{D}_i \mathbf{X} - \boldsymbol{\lambda}_i\|_F^2 \right) \\ + \frac{\beta_4}{2} \|\mathbf{R} - \mathbf{C} \mathbf{X} - \mathbf{v}\|_F^2 \text{ s.t. } \mathbf{S}_t \mathbf{P} \mathbf{r}_t = \mathbf{a}_t \quad (t = 1, 2, \dots, T) \end{aligned} \quad (8)$$

where $\boldsymbol{\lambda}_i$ and \mathbf{v} are Lagrange multiplier matrices. The constrained optimization problem in (7) has been replaced by problem (8). The ALM algorithm is to solve the minimization problem of (8) by iteratively minimizing the Lagrange function and updating the Lagrange multiplier,

$$(\mathbf{G}_i^{k+1}, \mathbf{R}^{k+1}, \mathbf{X}^{k+1}) = \arg \min_{\mathbf{X}, \mathbf{G}_i, \mathbf{R}} \mathcal{L}(\mathbf{X}, \mathbf{G}_i, \mathbf{R}, \boldsymbol{\lambda}_i, \mathbf{v}) \quad (9)$$

$$\boldsymbol{\lambda}_i^{k+1} = \boldsymbol{\lambda}_i^k - \tau (\mathbf{G}_i^{k+1} - \mathbf{D}_i \mathbf{X}^{k+1}) \quad \forall i = 1, 2, 3 \quad (10)$$

$$\mathbf{v}^{k+1} = \mathbf{v}^k - \tau (\mathbf{R}^{k+1} - \mathbf{C} \mathbf{X}^{k+1}) \quad (11)$$

Note that it is difficult to solve (9) directly. One can use an alternating strategy to minimize the augmented Lagrange function with respect to each component separately, namely,

$$\mathbf{G}_i^{k+1} = \arg \min_{\mathbf{G}_i} \mathcal{L}(\mathbf{G}_i, \mathbf{X}^k, \mathbf{R}^k, \boldsymbol{\lambda}_i^k, \mathbf{v}^k) \quad (12)$$

$$\mathbf{R}^{k+1} = \arg \min_{\mathbf{R}} \mathcal{L}(\mathbf{R}, \mathbf{G}_i^{k+1}, \mathbf{X}^k, \boldsymbol{\lambda}_i^k, \mathbf{v}^k) \quad (13)$$

$$\mathbf{X}^{k+1} = \arg \min_{\mathbf{X}} \mathcal{L}(\mathbf{X}, \mathbf{G}_i^{k+1}, \mathbf{R}^{k+1}, \boldsymbol{\lambda}_i^k, \mathbf{v}^k) \quad (14)$$

The sub-problem in (12) is solved as follows:

$$\mathbf{G}_i^{k+1} = S_{1/\beta_i}(\mathbf{D}_i \mathbf{X}^k + \boldsymbol{\lambda}_i^k) \quad (15)$$

where $S_\alpha(\cdot)$ is a soft-thresholding operator, which is defined, for a scalar x , as:

$$S_\alpha(x) = \text{sign}(x) \cdot \max\{|x| - \alpha, 0\} \quad (16)$$

where α is represented as a soft-thresholding. Suppose there is a matrix $Z = (z_{ij})$. Then, $S_\alpha(Z)$ outputs a matrix which defines an operator for matrix Z with respect to scalar α , i.e., the elements of $S_\alpha(Z)$ follow the definition in (16).

Next, we solve the sub-problem (13) through the following two steps [16].

$$\mathbf{r}_t^{k+1} = \mathbf{C}\mathbf{x}_t^k \quad (t = 1, 2, \dots, T) \quad (17)$$

$$\mathbf{r}_t^{k+1}(\text{Pic}S_t) = \mathbf{a}_t \quad (t = 1, 2, \dots, T) \quad (18)$$

where $\text{Pic}S_t$ is the index of measurements which is selected by \mathbf{S}_t , and \mathbf{r}_t is the t th column in \mathbf{R} .

In sub-problem (14), \mathbf{X} is updated through solving a quadratic problem.

By fixing \mathbf{X}^{k+1} , we reconstruct \mathbf{B} and \mathbf{F} in S-step by solving the following problem:

$$\min_{\mathbf{B}, \mathbf{F}} \gamma \|\mathbf{B}\|_* + \eta \|\mathbf{F}\|_1 \text{ s.t. } \mathbf{X}^{k+1} = \mathbf{B} + \mathbf{F} \quad (19)$$

The augmented Lagrange function of (19) can be expressed as:

$$\begin{aligned} \mathcal{L}(\mathbf{F}, \mathbf{B}, \mathbf{Y}) &= \|\mathbf{B}\|_* + \eta \|\mathbf{F}\|_1 + \langle \mathbf{X}^{k+1} - \mathbf{B} - \mathbf{F}, \mathbf{Y} \rangle \\ &> + \frac{\beta_5}{2} \|\mathbf{X}^{k+1} - \mathbf{B} - \mathbf{F}\|_F^2 \end{aligned} \quad (20)$$

where \mathbf{Y} is the Lagrange multiplier matrix, and $\langle \cdot, \cdot \rangle$ denotes the matrix inner product. We use ALM algorithm to solve the minimization problem in (20) by the following two steps:

$$(\mathbf{F}^{k+1}, \mathbf{B}^{k+1}) = \arg \min_{\mathbf{F}, \mathbf{B}} \mathcal{L}(\mathbf{F}, \mathbf{B}, \mathbf{Y}) \quad (21)$$

$$\mathbf{Y}^{k+1} = \mathbf{Y}^k + \mu(\mathbf{X}^{k+1} - \mathbf{B}^{k+1} - \mathbf{F}^{k+1}) \quad (22)$$

Similarly, we use an alternating strategy to minimize problem (21) with respect to each component separately:

$$\mathbf{F}^{k+1} = \arg \min_{\mathbf{F}} \mathcal{L}(\mathbf{F}, \mathbf{B}^k, \mathbf{Y}^k) \quad (23)$$

$$\mathbf{B}^{k+1} = \arg \min_{\mathbf{B}} \mathcal{L}(\mathbf{B}, \mathbf{F}^{k+1}, \mathbf{Y}^k) \quad (24)$$

The complete algorithm proposed to solve problem (6) is summarized in Algorithm 1 below.

Algorithm 1 Proposed alternating algorithm for solving problem (6)

Input: $\mathbf{C}, \mathbf{P}, \mathbf{S}_t, \mathbf{A}$

Output: $\mathbf{X}, \mathbf{B}, \mathbf{F}$

Set $\mathbf{X}^0 = \boldsymbol{\lambda}_i^0 = \text{zeros}(MN, T)$ for $i = 1, 2, 3$

$$\mathbf{B}^0 = \mathbf{F}^0 = \mathbf{R}^0 = \mathbf{Y}^0 = \mathbf{v}^0 = \text{zeros}(MN, T)$$

While $\frac{\|\mathbf{X}^{k+1} - \mathbf{X}^k\|_F}{\|\mathbf{X}^k\|_F} > 10^{-6}$ **do**

1. **R-step**

Updating \mathbf{G}, \mathbf{R}

$$\mathbf{G}_i^{k+1} = S_{1/\beta_i}(\mathbf{D}_i \mathbf{X}^k + \boldsymbol{\lambda}_i^k) \quad (i = 1, 2, 3)$$

$$\mathbf{r}_t^{k+1} = \mathbf{C}\mathbf{x}_t^k \quad (t = 1, 2, \dots, T)$$

$$\mathbf{r}_t^{k+1}(\text{Pic}S_t) = \mathbf{a}_t \quad (t = 1, 2, \dots, T)$$

Updating \mathbf{X}

$$\mathbf{X}^{k+1} = \mathbf{M}^{-1} \left(\sum_{i=1}^3 \alpha_i \beta_i \mathbf{D}_i^T (\mathbf{G}_i^{k+1} - \boldsymbol{\lambda}_i^k) + \beta_4 \mathbf{C}^T (\mathbf{R}^{k+1} - \mathbf{v}^k) \right)$$

2. **S-step**

$$\mathbf{B}^{k+1} = \mathbf{D}_{1/\beta_5}(\mathbf{X}^{k+1} - \mathbf{F}^k - \mathbf{B}^k + \beta_5^{-1} \mathbf{Y}^k)$$

$$\mathbf{F}^{k+1} = S_{1/\beta_5}(\mathbf{X}^{k+1} - \mathbf{B}^{k+1} - \mathbf{F}^k + \beta_5^{-1} \mathbf{Y}^k)$$

3. **Updating Lagrange multipliers**

$$\boldsymbol{\lambda}_i^{k+1} = \boldsymbol{\lambda}_i^k - \tau(\mathbf{G}_i^{k+1} - \mathbf{D}_i \mathbf{X}^{k+1}) \quad i = 1, 2, 3$$

$$\mathbf{v}^{k+1} = \mathbf{v}^k - \tau(\mathbf{R}^{k+1} - \mathbf{C}\mathbf{X}^{k+1})$$

$$\mathbf{Y}^{k+1} = \mathbf{Y}^k + \mu(\mathbf{X}^{k+1} - \mathbf{B}^{k+1} - \mathbf{F}^{k+1})$$

4. $k = k+1$

end while

In the above algorithm, $\mathbf{M} = \sum_{i=1}^3 \alpha_i \beta_i \mathbf{D}_i^T \mathbf{D}_i + \beta_4 \mathbf{C}^T \mathbf{C}$, $\mathcal{D}_\alpha(\cdot)$ is the singular value shrinkage operator [27], which is defined as follows: suppose the SVD of a matrix Z is given by $Z = \mathbf{U}\Sigma\mathbf{V}^T$, where Σ is a rectangular diagonal matrix in which each diagonal entries Σ_{ii} is the singular value of Z , \mathbf{U} and \mathbf{V} are real unitary matrix. The singular value shrinkage operator for matrix Z is defined as $\mathcal{D}_\alpha(Z) = \mathbf{U}S_\alpha(\Sigma)\mathbf{V}^T$, where $S_\alpha(\cdot)$ is soft-thresholding operator for matrix Σ with respect to α . In Algorithm 1, the

termination criterion is set as $\frac{\|x^{k+1}-x^k\|_F}{\|x^k\|_F} = 10^{-6}$ considering that the reconstruction of \mathbf{B} and \mathbf{F} rely on the reconstruction of \mathbf{X} .

The solution to problem (7) does not guarantee a global minimum solution for problem (6). Moreover, it is difficult to rigorously prove the convergence of the proposed alternating algorithm for problem (7). But we can prove that there exists a feasible solution for \mathbf{X} , \mathbf{B} , and \mathbf{F} that can minimize the cost function in (6). This feasible solution is stated in the following theorem.

Theorem 1: The sequence $\{\mathbf{X}^k\}$, $\{\mathbf{B}^k\}$, and $\{\mathbf{F}^k\}$ generated in Algorithm 1 are bounded, and there exists a feasible point $(\mathbf{X}^*, \mathbf{B}^*, \mathbf{F}^*)$ for the solution of problem (6).

The proof of Theorem 1 is given in Appendix.

4 Experimented results

In this section, we perform numerical experiments to show the performance of the proposed object detection algorithm. We focus on the illustration of the moving object reconstruction result and show that the new object detection algorithm is robust to the movement turbulence.

For quantitative evaluation, we utilize F -measure to evaluate the accuracy of the moving object detection result. The F -measure is defined as:

$$F\text{-measure} = \frac{2 \times (\text{precision} \times \text{recall})}{\text{precision} + \text{recall}} \quad (25)$$

where ‘precision’ and ‘recall’ are given by:

$$\text{precision} = \frac{TP}{TP + FP}, \text{ recall} = \frac{TP}{TP + FN} \quad (26)$$

‘Precision’ and ‘recall’ are two kinds of classification accuracy parameters which are widely used to measure the accuracy of the background subtraction result [28]. In the ‘precision’ and ‘recall’, TP, FP, and FN are the number of true positives, the number of false positives, and the number of false negatives, respectively. The higher the F -measure, the better the accuracy of the moving object detection is. The major parameters used in Algorithm 1 are shown in Table 1. In our experiments, we compare the proposed object detection algorithm with the RPCA method as well as a widely used background subtraction algorithm called improved Gaussian mixture model (GMM) [29]. Both RPCA and

GMM are operated in the spatial domain. All the experiments are performed on an Acer PC (CPU is Inter(R) Core(TM) i3-2310 M 2.10 GHz).

The testing video sequence for all the experiments are chosen from the database which are detailed in Table 2.

4.1 The new object detection model

Here, we choose Fountain sequence as an example to show first the video reconstruction result of the new object detection model. In this experiment, we compare the video reconstruction result of the proposed object detection model with three known video reconstruction sparsity measures: 2DTV, DWT, and 2DTV + DWT. The simulation results in terms of peak signal-to-noise ratio (PSNR) of the four methods are shown in Figure 2.

It is seen that the PSNR of the reconstructed video using the proposed object detection model is significantly higher than that of 2DTV, DWT, and 2DTV + DWT. Figure 3 shows the twentieth frame of the original video sequence and the corresponding reconstruction results of the four methods. Evidently, the reconstructed video frame using the proposed object detection model is clearer than that from 2DTV, DWT, and 2DTV + DWT. We can conclude from this experiment that the proposed reconstruction model is able to yield superior video reconstruction performance.

Next, we illustrate the video reconstruction and object detection results of our proposed model versus the sampling rate as shown in Figure 4. The chosen video sequence is from an airport video which contains a large amount of edge information and thus can highlight the difference of video foreground reconstruction results at different sampling rates. In addition, we compare our object detection results with compressive principal component pursuit (PCP) in Figure 4, which clearly shows the advantage of using $TV3D$ norm in our object detection model.

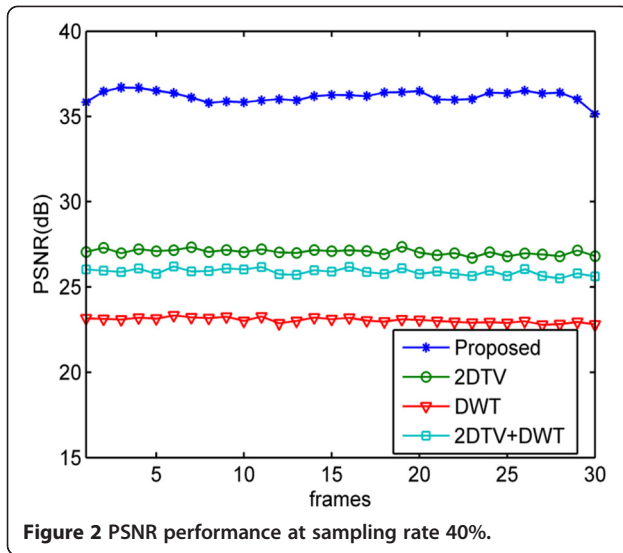
Clearly, Figure 4b,c,d give an exact foreground reconstruction result, where in order to see the difference among three images, we have given the local magnified images of the foreground reconstruction result. It is seen that Figure 4d gives the best foreground reconstruction

Table 1 Parameters used in solving the proposed reconstruction model

η_1	β_1	β_2	β_3	β_4	β_5	τ	μ
$\frac{1}{\sqrt{M \times N}}$	100	100	100	100	100	1.6	1

Table 2 Sequence information used in experiments

Name	Size ($m \times n \times T$)	Ref.
Airport	176 × 144 × 30	
Lobby	160 × 128 × 30	
Canteen	160 × 120 × 30	
Shopping mall	320 × 256 × 30	[30]
Campus	160 × 128 × 40	
Fountain	160 × 128 × 40	
Pedestrian	360 × 240 × 30	[31]



result. Figure 4c gives a slightly better performance than Figure 4b does. This is because the sampling rate used in Figure 4c is higher than that in Figure 4b. Figure 4a does not give a clear foreground reconstruction result due to the poor performance of the video reconstruction result. Comparing with Figure 4a,b,c,d, Figure 4e,f,g,h give poor video foreground and background reconstruction results. This is because that Figure 4e, f, g, h are reconstructed by compressive PCP, which is a special case of problem (6) when $\alpha_i = 0$ ($i = 1, 2, 3$). In this special case, the poor video reconstruction performance has become the bottleneck that precludes good video background and foreground reconstruction at low sampling rate. We can conclude from this experiment that using $TV3D$ norm in our model can guarantee a high object detection performance at low sampling rate. In addition to the above subjective measure of the object detection performances at different sampling rates, we choose PSNR and root mean square error (RMSE) as objective evaluation parameters to further illustrate the performance of the proposed object detection model and compressive PCP at different sampling rates.

In Table 3, PSNR is used to measure the video reconstruction result and RMSE_B is utilized to evaluate the

RMSE of the background reconstruction result. From Figure 4 and Table 3, we could see that at 20% sampling rate, the PSNR of our video reconstruction result is already above 30 dB, this means that we have obtained enough information for the exact foreground reconstruction result.

4.2 The moving object detection result

Here, we illustrate the performance of the proposed object detection algorithm with an emphasis on the reconstruction of foreground and background. In order to compare with GMM algorithm, we give the binary form of our foreground reconstruction result in the following experiments. We choose four indoor video sequences (airport, lobby, canteen, and shopping mall) to illustrate that the proposed object detection algorithm is able to give a similar performance as the popular spatial-domain moving object detection methods do. The reconstruction results of our proposed algorithm for four indoor video sequences are shown in Figure 5, where columns 1 to 4 are the moving object detection results of airport, lobby, canteen, and shopping mall video sequences, respectively. It is seen that our proposed algorithm using only 20% sampled measurement can give a similar moving object detection performance as the RPCA and GMM methods do. In the lobby and canteen video sequences, the proposed moving object detection algorithm is able to reduce the shadow turbulence. Table 4 gives objective evaluation results in terms of the F -measure of the proposed algorithm along with the two known methods for the four video sequences. We can see that the F -measure of the proposed object detection algorithm is obviously higher than that of the GMM method. From this experiment, we can conclude that the proposed moving object detection algorithm is able to exactly detect the moving object using only 20% sampled measurements in the indoor video sequences.

Figure 6 shows object detection results of the lobby video sequence with a sudden illumination change from the 10th frame to 11th frame. It is clearly seen that the proposed algorithm is robust to the sudden illumination changes of the indoor video sequence.

We now illustrate the performance of the proposed algorithm in outdoor video sequence. The outdoor video

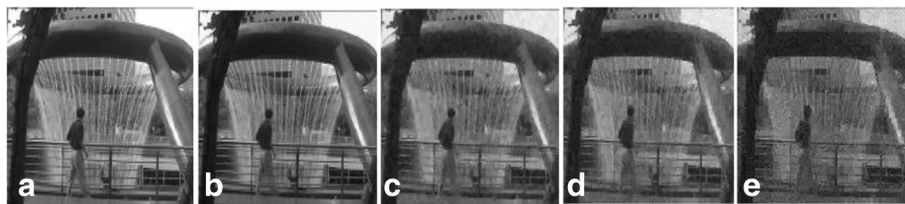


Figure 3 The 20th frame of original video sequence and reconstruction results. (a) Original video. (b) The proposed object detection model, PSNR: 36.25 dB. (c) 2DTV, PSNR: 27.08 dB. (d) 2DTV + DWT, PSNR: 26.11 dB. (e) DWT, PSNR: 23.06 dB.

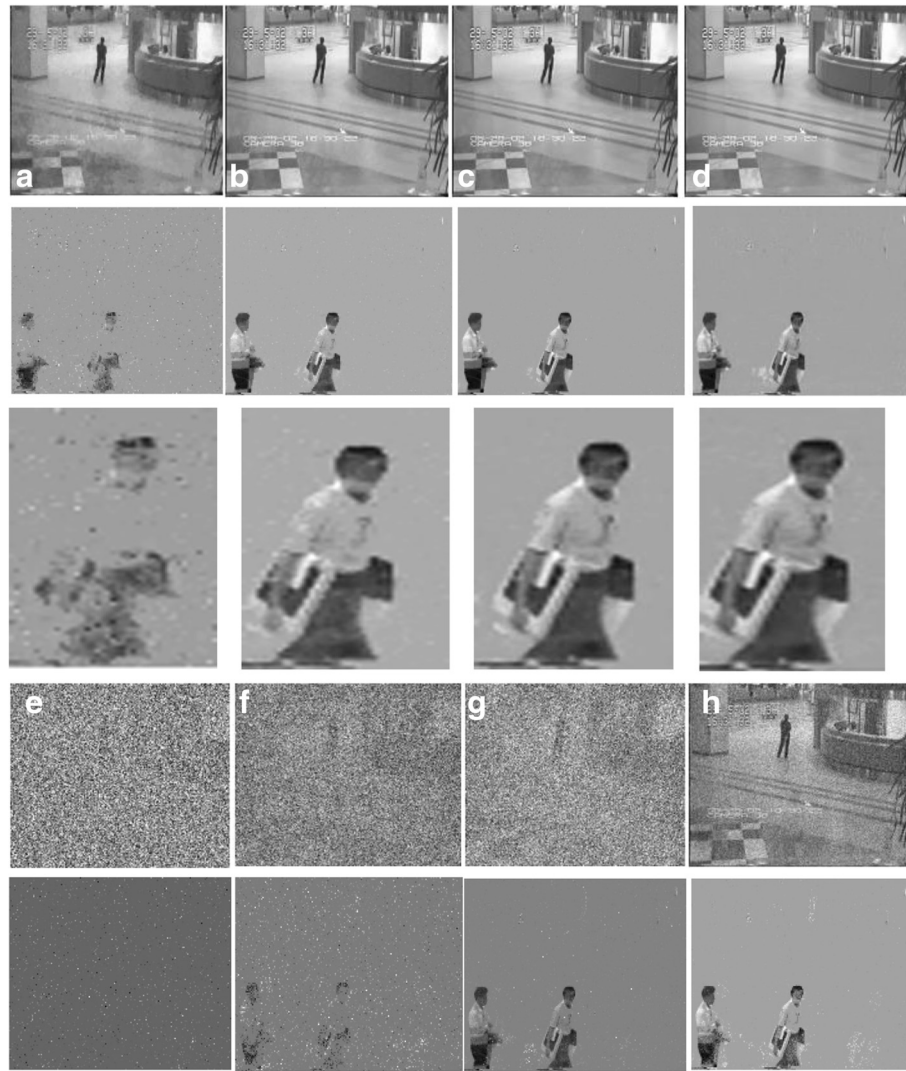


Figure 4 The object detection performances at different sampling rate. **(a)** The background first row and foreground second row reconstruction results using proposed model at 10% sampling rate. **(b)** The background and foreground reconstruction results using proposed model at 30% sampling rate. **(c)** The background and foreground reconstruction results using proposed model at 50% sampling rate. **(d)** The background and foreground reconstruction results using RPCA. The third row of **a, b, c, d** are the local magnification images. **(e, f, g, h)** are the background and foreground reconstruction results using compressive PCP at 10%, 30%, 50%, and 70% sampling rate, respectively.

sequence usually contains strong movement turbulence. We choose campus, fountain, and pedestrian video sequences for this experiment. The pedestrian video sequence is captured by a COTS camera (the SONY DCW-TRV 740).

This test case is very challenging because the whole video sequence is strongly disturbed by the swaying tree and flag. From Figure 7, it is obvious that the proposed algorithm is able to effectively eliminate the turbulence of the swaying trees (Figure 7b), while RPCA is not robust to this kind of strong movement turbulence (Figure 7c). The post-processing result of RPCA (Figure 7e) can give a slightly better performance than the proposed algorithm does. Although the GMM method can reduce the

movement turbulence (Figure 7d), its foreground reconstruction result is not better than that of the proposed object detection algorithm. We can conclude from this experiment that the proposed object detection algorithm is able to give a robust foreground reconstruction result using only 40% sampled measurement.

In this experiment, the background involves a huge fountain, which would strongly disturb the moving object. It is seen from Figure 8 that the new object detection algorithm is able to efficiently eliminate the fountain turbulence, and it gives a better foreground reconstruction result than the GMM method does (Figure 8b,d). RPCA is still not robust to this kind of movement turbulence (Figure 8c). The post-processing result of RPCA (Figure 8e)

Table 3 Evaluation of the proposed model at different sampling rate

Sampling rate	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	
Proposed object detection model									
PSNR	24.06	31.81	34.20	36.77	40.37	43.61	46.35	49.45	RMSE_B in RPCA
	dB	dB	dB	dB	dB	dB	dB	dB	
RMSE_B	0.080	0.061	0.059	0.057	0.052	0.049	0.047	0.046	
Compressive PCP									
PSNR	4.61	5.16	6.83	7.19	7.32	12.51	22.43	30.56	Model is 0.045
	dB	dB	dB	dB	dB	dB	dB	dB	
RMSE_B	0.924	0.828	0.632	0.577	0.522	0.292	0.130	0.079	

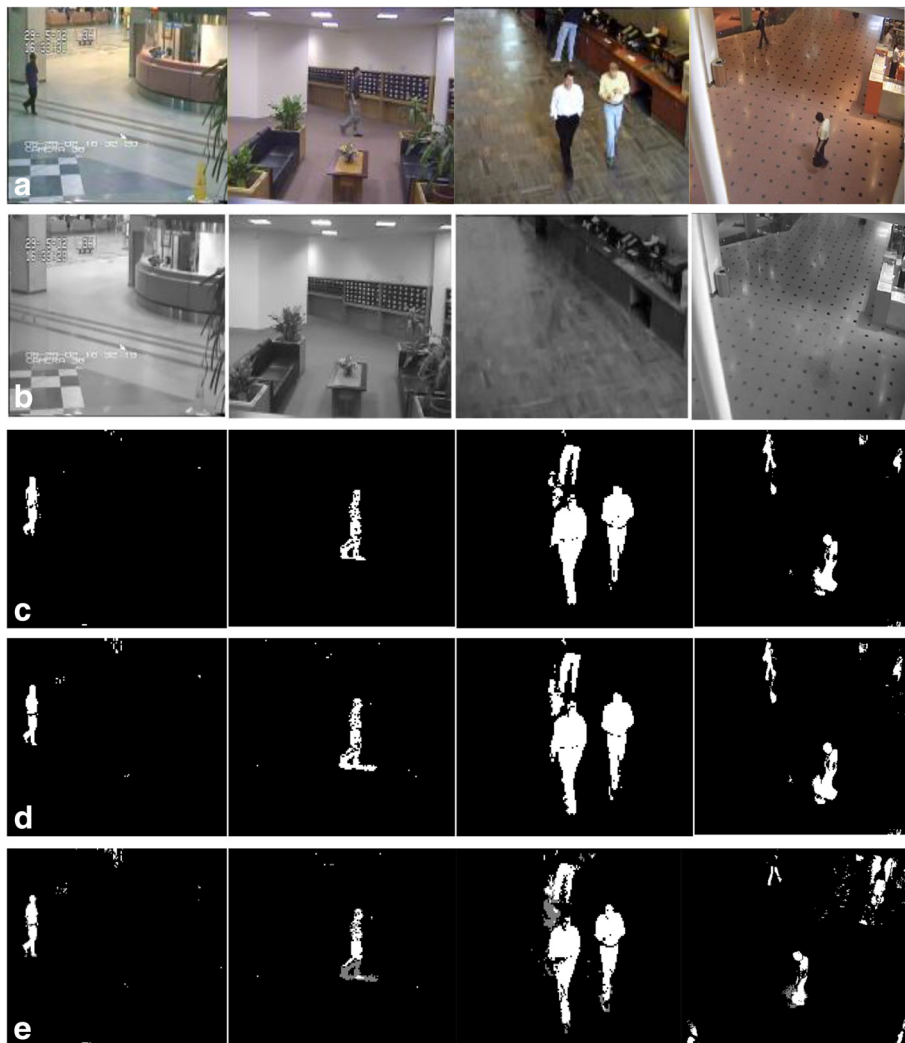


Figure 5 Object detection results of four indoor video sequences. (a) The four kinds of original video sequences. (b, c) The reconstructed background and foreground using the proposed object detection algorithm. (d) The reconstructed foreground using RPCA. (e) The reconstructed foreground using GMM. The sampling rate of our object detection algorithm is 20%. The 24th frame of the airport sequence, 11th frame of the lobby sequence, 19th frame of the canteen sequence and 6th frame of the shopping mall sequence are randomly selected.

Table 4 Quality evaluation (*F*-measure) of the detection results in Figure 5

Sequence	Proposed	RPCA	GMM
Airport	0.55	0.56	0.50
Lobby	0.56	0.45	0.43
Canteen	0.63	0.61	0.59
Shopping mall	0.49	0.50	0.39

is better than the proposed algorithm due to the fact that RPCA is operated in the spatial domain. The original video sequence can give RPCA a large amount of detailed information.

We choose a real-world outdoor video sequence to conduct this experiment. The chosen video sequence contains ordinary turbulence such as shadow and camera noise. We randomly select four frames to show the moving object detection performance of different methods. It is clearly shown in Figure 9b that the proposed object detection algorithm is able to exactly distinguish the contour outlines of the moving person. It can completely eliminate the camera noise. Both of RPCA and GMM (see Figure 9c,d) could not give a clear moving object detection result. The averaged *F*-measure of Figures 7, 8, and 9 are given in Table 5, which show that the proposed algorithm gives an obviously higher *F*-measure than the RPCA and GMM methods do.

5 Conclusion

In this paper, we have proposed a CS-based algorithm for detecting the moving object in video sequences. In order to achieve robust foreground reconstruction result using only a small number of sampled measurements, we have first proposed an object detection model to simultaneously reconstruct the foreground, background, and

the original video sequence using the sampled measurements. Then, the reconstructed video sequence is used to estimate a confidence map to refine the foreground reconstruction result. It has been shown through experiment that the proposed moving object detection algorithm can give a good performance for both indoor and outdoor video sequences. Especially for outdoor video sequence, the proposed reconstruction model is able to effectively eliminate the movement turbulence such as waving trees, water fountain, and video noise. In conclusion, the proposed moving object detection algorithm can achieve an accuracy comparable to some known spatial-domain methods with a significantly reduced number of sampled measurements. The limitation of the proposed method includes: (1) In Algorithm 1, solving nuclear norm imposes high computational complexity. (2) There is a lack of theoretical analysis of the impact of the sampling rate on the object detection result. To solve those problems in future work, (1) we will use an online version of object detection model to achieve background reconstruction, and (2) we will refer to [15] for possible theoretical analysis of the performance of the proposed model.

6 Appendix: Proof of Theorem 1

The proof of Theorem 1 is based on following two Lemmas.

Lemma 1: Let $\mathbf{a}_i^k = \lambda_i^k - \tau(\mathbf{G}_i^{k+1} - \mathbf{D}_i \mathbf{X}^k)$ ($i = 1, 2, 3$), $\mathbf{b}^k = \mathbf{Y}^k + \mu(\mathbf{X}^{k+1} - \mathbf{B}^{k+1} - \mathbf{F}^k)$ $\mathbf{c}^k = \mathbf{Y}^k + \mu(\mathbf{X}^{k+1} - \mathbf{B}^{k+1} - \mathbf{F}^{k+1})$. The sequence $\{\mathbf{a}_i^k\}$ ($i = 1, 2, 3$), $\{\mathbf{b}^k\}$ and $\{\mathbf{c}^k\}$ are bounded.

Proof of Lemma 1

(i) We prove first $\{\mathbf{a}_i^k\}$ ($i = 1, 2, 3$) is bounded.

In each iteration of Algorithm 1, \mathbf{G}_i ($i = 1, 2, 3$) is updated through solving problem (7), and \mathbf{B} and \mathbf{F} are

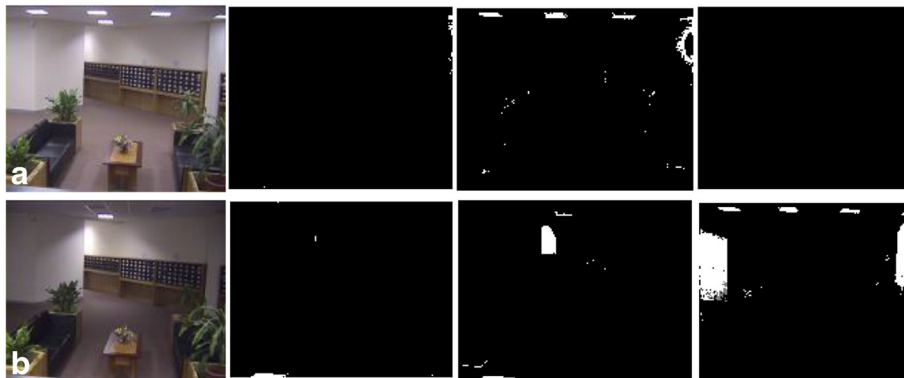


Figure 6 Object detection results of lobby video sequence with sudden illumination change. (a) The 10th video foreground reconstruction result. **(b)** The 11th video foreground reconstruction result. The first column is the 10th and 11th frames of the original video sequence. The second column is the foreground reconstruction results of the proposed object detection algorithm. The third column is RPCA's foreground reconstruction results. Finally, the fourth one is the foreground reconstruction results of GMM.

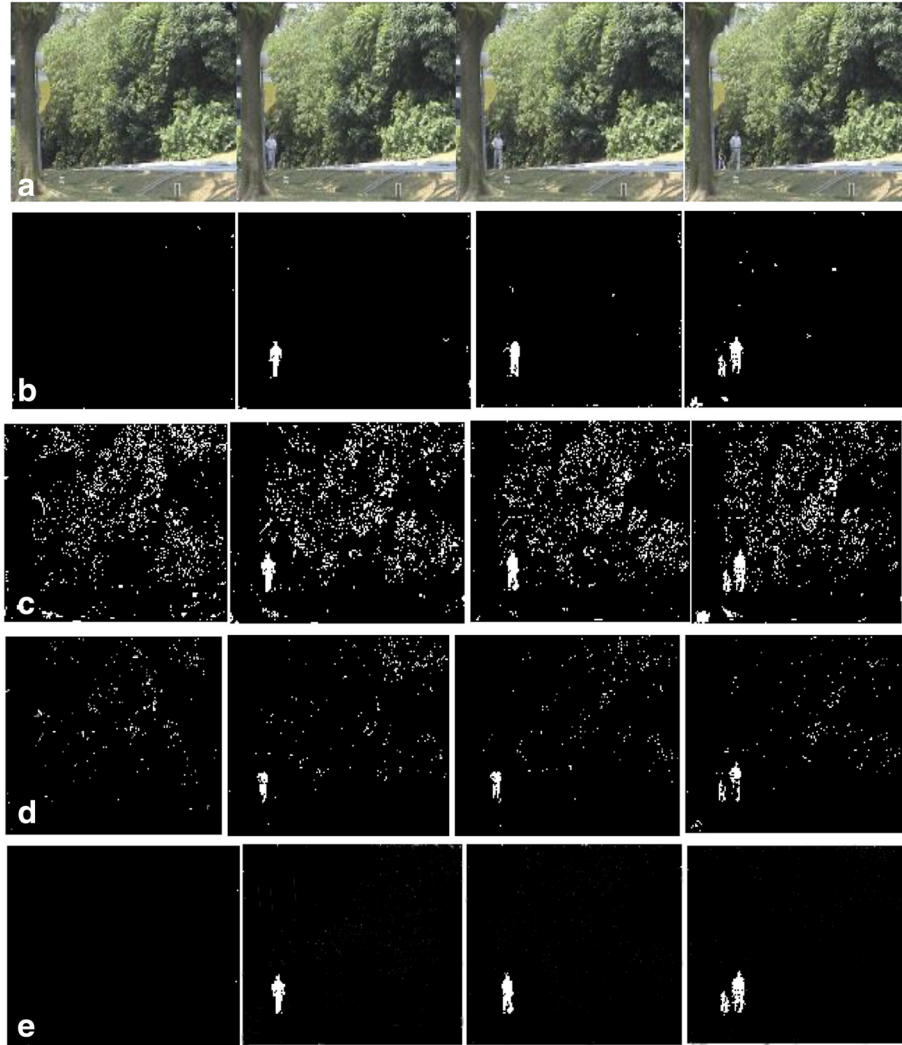


Figure 7 Object detection results of campus video sequence. **(a)** Original video sequence. **(b)** Reconstructed foreground using the proposed algorithm. **(c)** Reconstructed foreground using RPCA. **(d)** Reconstructed foreground using GMM. **(e)** Reconstructed foreground using modified RPCA (the manual postprocess result of RPCA). The sampling rate of our algorithm is 40% and four frames, i.e., 1st, 18th, 25th, and 35th frames are randomly selected.

reconstructed through solving problem (19). When minimizing the Lagrange function in (8), we can obtain:

$$\begin{aligned} \mathbf{G}_i^{k+1} &= \min_{\mathbf{G}_i} \mathcal{L}(\mathbf{G}_i, \mathbf{R}^k, \mathbf{X}^k, \boldsymbol{\lambda}_i^k, \mathbf{v}^k) \\ &\Rightarrow 0 \in \partial_{\mathbf{G}_i} \mathcal{L}(\mathbf{G}_i^{k+1}, \mathbf{R}^k, \mathbf{X}^k, \boldsymbol{\lambda}_i^k, \mathbf{v}^k) \\ &\Rightarrow \mathbf{a}_i^k \in \partial_{\mathbf{G}_i} \|\mathbf{G}_i^{k+1}\|_1 \quad (i = 1, 2, 3) \end{aligned}$$

$\therefore \mathbf{a}_i^k \in \partial_{\mathbf{G}_i} \|\mathbf{G}_i^{k+1}\|_1$, hence, $\{\mathbf{a}_i^k\} (i = 1, 2, 3)$ is bounded [22].

(ii) Now we prove $\{\mathbf{b}^k\}$ and $\{\mathbf{c}^k\}$ are bounded.

When minimizing the Lagrange function in (20), we can obtain that:

$$\begin{aligned} \mathbf{B}^{k+1} &= \min_{\mathbf{B}} \mathcal{L}(\mathbf{B}, \mathbf{F}^k, \mathbf{Y}^k) \\ &\Rightarrow 0 \in \partial_{\mathbf{B}} \mathcal{L}(\mathbf{B}^{k+1}, \mathbf{F}^k, \mathbf{Y}^k) \\ &\Rightarrow \mathbf{b}^k \in \partial_{\mathbf{B}} \|\mathbf{B}^{k+1}\|_* \end{aligned}$$

$\therefore \mathbf{b}^k \in \partial_{\mathbf{B}} \|\mathbf{B}^{k+1}\|_*$, so the sequence $\{\mathbf{b}^k\}$ is bounded [22].

$$\begin{aligned} \mathbf{F}^{k+1} &= \min_{\mathbf{F}} \mathcal{L}(\mathbf{B}^{k+1}, \mathbf{F}, \mathbf{Y}^k) \\ &\Rightarrow 0 \in \partial_{\mathbf{F}} \mathcal{L}(\mathbf{B}^{k+1}, \mathbf{F}^{k+1}, \mathbf{Y}^k) \\ &\Rightarrow \mathbf{c}^k \in \partial_{\mathbf{F}} \|\mathbf{F}^{k+1}\|_1 \end{aligned}$$

$\therefore \mathbf{c}^k \in \partial_{\mathbf{F}} \|\mathbf{F}^{k+1}\|_1$, hence the sequence $\{\mathbf{c}^k\}$ is bounded [22].

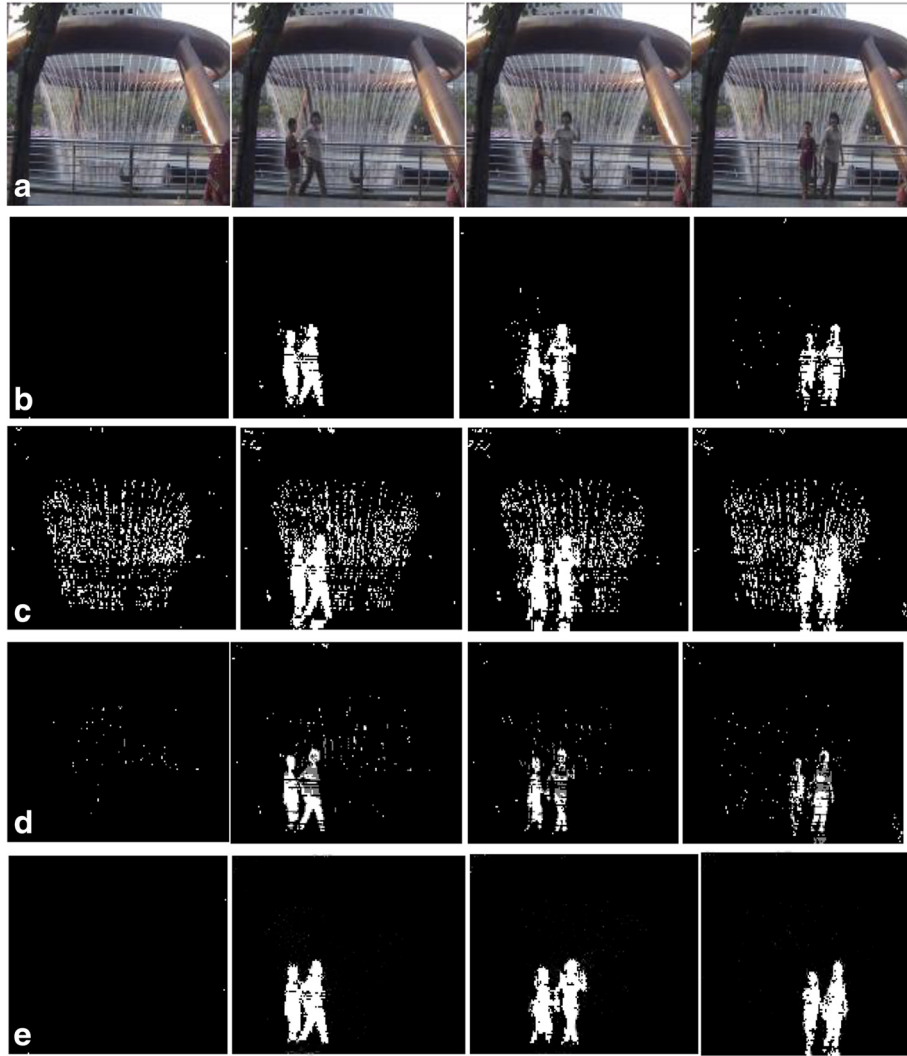


Figure 8 Object detection results of fountain video sequence. (a) Original video sequence. (b) Reconstructed foreground using the proposed algorithm. (c) Reconstructed foreground using RPCA. (d) Reconstructed foreground using GMM. (e) Reconstructed foreground using modified RPCA (the manual postprocess result of RPCA). The sampling rate of our algorithm is 40% and four frames, i.e., 1st, 12th, 14th, and 34th frames are randomly selected.

Lemma 2: Let $\mathcal{L}(\mathbf{B}, \mathbf{F}, \mathbf{Y})$ be the Lagrange function of problem (19). Then we have $\mathcal{L}^{k+1} = \mathcal{L}(\mathbf{B}^{k+1}, \mathbf{F}^{k+1}, \mathbf{Y}^k)$ and $\mathcal{L}^{k+1} - \mathcal{L}^k \leq \frac{1}{\mu} e^k$ with $e^k = \|\mathbf{Y}^k - \mathbf{Y}^{k-1}\|_F$, $k = 1, 2, \dots$

Proof of Lemma 2

Let $\mathcal{L}^{k+1} = \mathcal{L}(\mathbf{B}^{k+1}, \mathbf{F}^{k+1}, \mathbf{Y}^k)$, then:

$$\begin{aligned} \mathcal{L}^{k+1} &\leq \mathcal{L}(\mathbf{B}^k, \mathbf{F}^k, \mathbf{Y}^k) = \|\mathbf{B}^k\|_* + \eta_1 \|\mathbf{F}^k\|_1 + \langle \mathbf{X}^k - \mathbf{B}^k - \mathbf{F}^k, \mathbf{Y}^k \rangle \\ &\quad + \frac{\beta_5}{2} \|\mathbf{X}^k - \mathbf{B}^k - \mathbf{F}^k\|_F^2 = \mathcal{L}^k + \langle \mathbf{X}^{k+1} - \mathbf{B}^k - \mathbf{F}^k, \mathbf{Y}^k - \mathbf{Y}^{k-1} \rangle \end{aligned} \quad (\text{A1})$$

$$\begin{aligned} \mathbf{Y}^k &= \mathbf{Y}^{k-1} + \mu (\mathbf{X}^{k+1} - \mathbf{B}^k - \mathbf{F}^k) \Rightarrow \frac{\mathbf{Y}^k - \mathbf{Y}^{k-1}}{\mu} \\ &= \mathbf{X}^{k+1} - \mathbf{B}^k - \mathbf{F}^k \end{aligned} \quad (\text{A2})$$

Substituting (A2) into (A1), we can obtain that $\mathcal{L}^{k+1} - \mathcal{L}^k \leq \frac{1}{\mu} \|\mathbf{Y}^k - \mathbf{Y}^{k-1}\|_F$. End of proof of Lemma 2.

In Algorithm. 1, we need to solve two Lagrange functions: $\mathcal{L}(\mathbf{X}, \mathbf{G}_i, \mathbf{R}, \lambda_i, \mathbf{v})$ and $\mathcal{L}(\mathbf{B}, \mathbf{F}, \mathbf{Y}^k)$ for the reconstruction of \mathbf{X} , \mathbf{B} and \mathbf{F} . If there exists a feasible point $(\mathbf{X}^*, \mathbf{B}^*, \mathbf{F}^*)$ for the solution of problem (6), we must make sure that $\mathcal{L}(\mathbf{X}, \mathbf{G}_i, \mathbf{R}, \lambda_i, \mathbf{v})$ and $\mathcal{L}(\mathbf{B}, \mathbf{F}, \mathbf{Y}^k)$ are all bounded.

In Theorem 1 of [16], it has been proved that $\mathcal{L}(\mathbf{X}, \mathbf{G}_i, \mathbf{R}, \lambda_i, \mathbf{v})$ is bounded, so we only need to prove that $\mathcal{L}(\mathbf{B}, \mathbf{F}, \mathbf{Y}^k)$ is bounded. We have proved that: $\{\mathbf{c}^k\}$ is bounded in Lemma 1.

Noting that $\mathbf{Y}^{k+1} = \mathbf{c}^k$, we have bounded $\{\mathbf{Y}^k\}$.

Since $\mathcal{L}^{k+1} - \mathcal{L}^k \leq \frac{1}{\mu} \|\mathbf{Y}^k - \mathbf{Y}^{k-1}\|_F$ and $\{\mathbf{Y}^k\}$ is bounded, we conclude that $\mathcal{L}(\mathbf{B}, \mathbf{F}, \mathbf{Y}^k)$ is bounded.

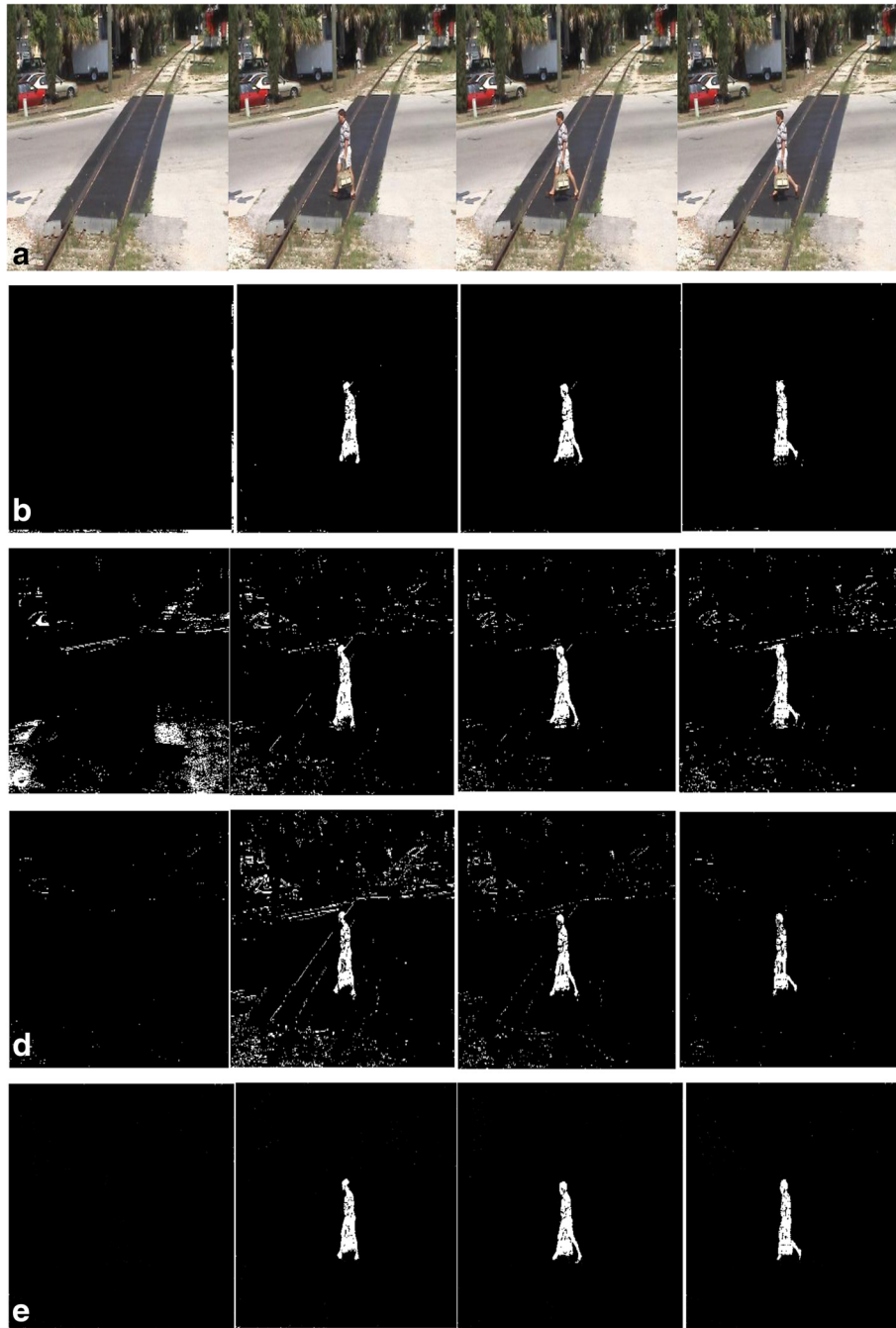


Figure 9 Object detection results of pedestrian video sequence. (a) Original video sequence. (b) Reconstructed foreground using the proposed algorithm. (c) Reconstructed foreground using RPCA. (d) Reconstructed foreground using GMM. (e) Reconstructed foreground using modified RPCA (the manual postprocess result of RPCA). The sampling rate of our algorithm is 40% and four frames, i.e., 1st, 8th, 12th, and 15th frame are randomly selected.

Table 5 Quality evaluation (*F*-measure) on Figures 7, 8, and 9

Experiments	Proposed	RPCA	GMM	Modified RPCA
Figure 6	0.36	0.07	0.13	0.38
Figure 7	0.49	0.18	0.47	0.55
Figure 8	0.61	0.43	0.42	0.62

Next we prove that $\{\mathbf{X}^k\}$, $\{\mathbf{B}^k\}$ and $\{\mathbf{F}^k\}$ are bounded.

As it has been proved that $\mathcal{L}(\mathbf{X}, \mathbf{G}_b, \mathbf{R}, \boldsymbol{\lambda}_b, \mathbf{v})$ is bounded, we can obtain that the Lagrange multiplier $\{\boldsymbol{\lambda}_i^k\}$ is bounded.

From (10) in paper, we obtain:

$$\boldsymbol{\lambda}_i^{k+1} = \boldsymbol{\lambda}_i^k - \tau(\mathbf{G}_i^{k+1} - \mathbf{D}_i \mathbf{X}^{k+1}) \quad (\text{A3})$$

From Lemma 1, we have:

$$\lambda_i^k = \mathbf{a}_i^k + \tau(\mathbf{G}_i^{k+1} - \mathbf{D}_i \mathbf{X}^k) \quad (\text{A4})$$

Subtracting (A4) into (A3), we get:

$$\lambda_i^{k+1} = \mathbf{a}_i^k + \tau \mathbf{D}_i (\mathbf{X}^{k+1} - \mathbf{X}^k)$$

$\{\lambda_i^k\}$ and $\{\mathbf{a}_i^k\}$ are bounded, so $\{\mathbf{X}^k\}$ is bounded.

Noting that:

$$\begin{aligned} \|\mathbf{Y}^{k+1}\|_F^2 &= \|\mathbf{Y}^k + \mu(\mathbf{X}^{k+1} - \mathbf{B}^{k+1} - \mathbf{F}^{k+1})\|_F^2 = \|\mathbf{Y}^k\|_F^2 \\ &\quad + 2\mu \langle \mathbf{Y}^k, \mathbf{X}^{k+1} - \mathbf{B}^{k+1} - \mathbf{F}^{k+1} \rangle \\ &\quad + \mu^2 \|\mathbf{X}^{k+1} - \mathbf{B}^{k+1} - \mathbf{F}^{k+1}\|_F^2 \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{\|\mathbf{Y}^{k+1}\|_F^2 - \|\mathbf{Y}^k\|_F^2}{2\mu} &= \langle \mathbf{Y}^k, \mathbf{X}^{k+1} - \mathbf{B}^{k+1} - \mathbf{F}^{k+1} \rangle \\ &\quad + \frac{\mu}{2} \|\mathbf{X}^{k+1} - \mathbf{B}^{k+1} - \mathbf{F}^{k+1}\|_F^2 \end{aligned}$$

Since \mathbf{Y}^k is bounded, we get $\langle \mathbf{Y}^k, \mathbf{X}^{k+1} - \mathbf{B}^{k+1} - \mathbf{F}^{k+1} \rangle + \frac{\mu}{2} \|\mathbf{X}^{k+1} - \mathbf{B}^{k+1} - \mathbf{F}^{k+1}\|_F^2$ converges to 0.

Thus, we have:

$$\begin{aligned} \|\mathbf{B}^{k+1}\|_* + \eta_1 \|\mathbf{F}^{k+1}\|_1 &= \mathcal{L}^{k+1} - \langle \mathbf{Y}^k, \mathbf{X}^{k+1} - \mathbf{B}^{k+1} - \mathbf{F}^{k+1} \rangle, \\ \mathbf{Y}^k &> -\frac{\beta_5}{2} \|\mathbf{X}^{k+1} - \mathbf{B}^{k+1} - \mathbf{F}^{k+1}\|_F^2 \end{aligned}$$

As \mathcal{L}^{k+1} is bounded and $\langle \mathbf{Y}^k, \mathbf{X}^{k+1} - \mathbf{B}^{k+1} - \mathbf{F}^{k+1} \rangle + \frac{\mu}{2} \|\mathbf{X}^{k+1} - \mathbf{B}^{k+1} - \mathbf{F}^{k+1}\|_F^2$ converge to 0, we can prove that $\{\mathbf{B}^k\}$ and $\{\mathbf{F}^k\}$ are bounded.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No.61372122 and 61302103); the Innovation Program for Postgraduate in Jiangsu Province under Grant (No. CXZZ13_0491).

Author details

¹College of Communication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China. ²Department of Electrical and Computer Engineering, Concordia University, Montreal, QC H3G 1 M8, Canada.

Received: 17 August 2014 Accepted: 5 January 2015

Published online: 26 February 2015

References

1. O Barnich, M Van Droogenbroeck, ViBe: a universal background subtraction algorithm for video sequences. *IEEE Trans. Image Process.* **20**(6), 1709–1724 (2011)
2. Brutzer, B Hoferlin, G Heidemann, Evaluation of background subtraction techniques for video surveillance, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 1937–1944
3. T Du-Ming, L Shia-Chih, Independent component analysis-based background subtraction for indoor surveillance. *IEEE Trans. Image Process.* **18**(1), 158–167 (2009)
4. Z Baochang, G Yongsheng, Z Sanqiang, Z Bineng, Kernel similarity modeling of texture pattern flow for motion detection in complex background. *IEEE Trans. Circuits Syst. Video Technol.* **21**(1), 29–38 (2011)
5. K Wonjun, K Changick, Background subtraction for dynamic texture scenes using fuzzy color histograms. *IEEE Signal Process. Lett.* **19**(3), 127–130 (2012)
6. S Chen, J Zhang, Y Li, J Zhang, A hierarchical model incorporating segmented regions and pixel descriptors for video background subtraction. *IEEE Trans. Inf. Theory* **8**(1), 118–127 (2012)
7. H Bohyung, LS Davis, Density-based multifeature background subtraction with support vector machine. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(5), 1017–1023 (2012)
8. R Baraniuk, Compressive sensing. *IEEE Signal Process. Mag.* **24**(4), 118–121 (2007)
9. DL Donoho, Compressed sensing. *IEEE Trans. Inf. Theory* **52**(4), 1289–1306 (2006)
10. EJ Candes, MB Wakin, An introduction to compressive sampling. *IEEE Signal Process. Mag.* **25**(2), 21–30 (2008)
11. J Ma, G Plonka, MY Hussaini, Compressive video sampling with approximate message passing decoding. *IEEE Trans. Circuits Syst. Video Technol.* **22**(9), 1354–1364 (2012)
12. V Cevher, A Sankaranarayanan, M Duarte, D Reddy, R Baraniuk, R Chellappa, Compressive sensing for background subtraction, in *Pro. European Conference on Computer Vision (ECCV)*, 2008
13. H Jiang, W Deng, Z Shen, Surveillance video processing using compressive sensing. *Inverse Probl. Imaging* **6**(2), 201–214 (2012)
14. F Yang, H Jiang, Z Shen, W Deng, D Metaxas, Adaptive low rank and sparse decomposition of video using compressive sensing, in *Proc. IEEE International Conference on Image Processing (ICIP)*, 2013, pp. 1016–1020
15. J Wright, A Ganesh, K Min, Y Ma, Compressive principal component pursuit. *Inf. Inference* **2**(1), 32–68 (2013)
16. X Shu, N Ahuja, Imaging via three-dimensional compressive sampling (3DCS), in *IEEE International Conference on Computer Vision (ICCV)*, 2011, pp. 439–446
17. B Bao, G Liu, C Xu, S Yan, Inductive robust principal component analysis. *IEEE Trans. Image Process.* **21**(8), 3794–3800 (2012)
18. EJ Candes, X Li, Y Ma, J Wright, Robust principal component analysis? *J. ACM* **58**(1), 1–37 (2009)
19. E Borenstein, E Sharon, S Ullman, Combining top-down and bottom-up segmentation, in *Proc. Conference on Computer Vision and Pattern Recognition Workshop*, 2004, p. 46
20. M Shimizu, S Yoshimura, M Tanaka, M Okutomi, Super-resolution from image sequence under influence of hot-air optical turbulence, in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008, pp. 1–8
21. O Oreifej, G Shu, T Pace, M Shah, A two-stage reconstruction approach for seeing through water, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 1153–1160
22. O Oreifej, X Li, M Shah, Simultaneous video stabilization and moving object detection in turbulence. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(2), 450–462 (2013)
23. C Stauffer, WEL Grimson, Adaptive background mixture models for real-time tracking, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 1999, p. 252
24. W Shandong, O Oreifej, M Shah, Action recognition in videos acquired by a moving camera using motion decomposition of Lagrangian particle trajectories, in *IEEE International Conference on Computer Vision (ICCV)*, 2011, pp. 1419–1426
25. S Wu, BE Moore, M Shah, Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 2054–2060
26. W Yin, S Morgan, J Yang, Y Zhang, Practical compressive sensing with Toeplitz and circulant matrices, in *Visual Communications and Image Processing Huangshan China*, 2010
27. H Yao, Z Debing, Y Jieping, L Xuelong, H Xiaofei, Fast and accurate matrix completion via truncated nuclear norm regularization. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(9), 2117–2130 (2013)
28. X Zhou, C Yang, Y Weichuan, Moving object detection by detecting contiguous outliers in the low-rank representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(3), 597–610 (2013)

29. Z Zivkovic, Improved adaptive Gaussian mixture model for background subtraction. *Int. Conf. Pattern Recog.* **2**, 28–31 (2004)
30. L Li, W Huang, IYH Gu, Q Tian, Statistical modeling of complex backgrounds for foreground object detection. *IEEE Trans. Image Process.* **13**(11), 1459–1472 (2004)
31. Y Sheikh, M Shah, Bayesian modeling of dynamic scenes for object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(11), 1778–1792 (2005)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
