Neurocomputing 328 (2019) 113–121

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Background subtraction with multi-scale structured low-rank and sparse factorization

Aihua Zheng, Tian Zou, Yumiao Zhao, Bo Jiang, Jin Tang, Chenglong Li*

School of Computer Science and Technology, Anhui University, Hefei, China

ARTICLE INFO

Article history:
Received 15 November 2017
Revised 29 January 2018
Accepted 22 February 2018
Available online 20 August 2018

Keywords:
Low-rank and sparse factorization
Structured constraint
Appearance consistency
Spatial compactness
Multi-scale

ABSTRACT

Low-rank and sparse factorization, which models the background as a low-rank matrix and the foreground as the contiguously corrupted outliers, exhibits excellent performance in background subtraction, in which the structured constraints of the foreground usually play a very essential role. In this paper, we propose a novel approach with multi-scale structured low-rank and sparse factorization for background subtraction. Different from the conventional methods that only enforce the smoothness between the spatial neighbors, we propose to explore the structured smoothness with both appearance consistency and spatial compactness in the low-rank and sparse factorization framework. Moreover, we integrate structural information at different scales into the formulation for robustness. We also design a low-rank decomposition scheme to improve the computational efficiency of the optimization algorithm. Extensive experiments on benchmark datasets GTFD and CDnet suggest that our approach achieves big superior performance against the state-of-the-art methods.

1. Introduction

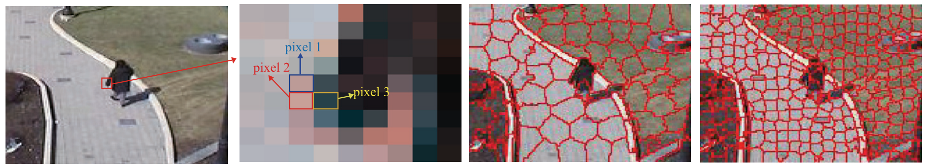
Moving object detection is a fundamental problem in computer vision, and plays a critical role in numerous vision applications, such as intelligent transportation [1], vehicle navigation [2] and scene understanding [3]. In the past decades, extensive algorithms made remarkable efforts to moving object detection, as referred in the interesting surveys [4–7], while background substraction has been recognized as one of the most competitive approaches. Conventional background substraction methods include GMM [8], VIBE [9] and their variants. Qin et al. [10] further introduced the Gabor filter into VIBE to reduce the “ghost” and then accelerated the updating procedure via GPU parallel computing. Zhang et al. [11] revealed the foreground and background imbalance problem and designed a two-stage compensation framework on the data and the algorithm level sequentially. Kim and Jung [7] comprehensively discussed the issue of illumination changes which is ubiquitous in the outdoor scenarios from the methodology and performance prospectives. However, it is still a challenging task due to undesired environment such as dynamic scene, illumination changes, severe weathers, occlusions and so on.

Meanwhile, deep learning based methods have been widely used in computer vision and pattern recognition. Some state-of-the-art networks have been designed for semantic segmentation, including FCN [12], DeepLab [13], MSCNN [14] and so on. Specifically, moving object detection tries to segment the moving foreground out of the background. Zhang et al. [15] designed a blockwise deep learning module to encode the image representation and then accelerated the detection via binary scene modeling in Hamming space. Heo et al. [16] combined the appearance network (a pre-trained VGG-16 network) and the motion network (a shallow network) for moving object detection. Chen et al. [17] designed an end-to-end architecture which first extracted high-level CNN features via an encoder–decoder network then modeled pixelwise temporal changes via an Attention ConvLSTM. Kim and Keles [18] developed a encoder–decoder neural network embedded the image into multi-scale feature space under a triplet framework during the encoder while learning the feature to image mapping via a transposed convolutional network during decoder.

Recently, the low-rank and sparse factorization framework has attracted a considerable attention. The basic idea is to factorize the given matrix of the accumulated frames into the low-rank background and sparse foreground as outliers. The pioneering work is Robust Principal Component Analysis (RPCA) [19,20] and its variants [21–23]. However, RPCA-based methods are sensitive to the outlier noise and tend to produce cavities due to the laciness of structural contiguous constraints. In order to enforce structured constraint on the foreground, DECOLOR [24] proposed to model the spatial contiguity among the neighboring pixels to preserve the spatial homogeneity of the foreground, Xin et al. [25] exploited

* Corresponding author.
E-mail addresses: lcl1314@foxmail.com, chenglongli@ahu.ed (C. Li).

https://doi.org/10.1016/j.neucom.2018.02.101
0925-2312/



(a) Original (b) Zoom-out (c) SLIC-100 (d) SLIC-300

Fig. 1. Demonstration of the multi-scale structured appearance consistency and the spatial compactness. Where (b) is the zoom-out view of the specified part of the person in (a), (c) and (d) are the superpixel segmentations via SLIC [30] where the number "100" and "300" indicate the number of superpixels.

the continuous structural information via efficient GFL [26] to strengthen the fusion among the adjacent pixels. GOSUS [27] proposed to impose the group sparsity on the pre-processed sliding windows or superpixels to encourage the perceptually meaningful groups. Yang et al. [28] proposed a subspace-based motion segmentation method by integrating the global sparse subspace optimization via PCA projection and local refinement via a simple error estimation. Javed et al. [29] presented an online matrix decomposition with max-norm regularization and structured sparsity constraints on each superpixel segment. However, they construct the structured constraint only between the spatial neighbors while ignoring the high-level spatial compactness of the foreground. Moreover, none of them take into account the multi-scale cues which can further promote the appearance consistency and spatial compactness on varying scales.

In this paper, we propose a novel multi-scale structured low-rank and sparse factorization for background subtraction. Comparing to the deep-based methods, we do not require laborious pre-training or the large training set. In addition, we have no need of saving a large pre-trained deep model. We enforce our structured constraints in terms of integrating the multi-scale appearance consistency and spatial compactness of the foreground into a unified model. First, we observe that, the foregrounds are generally consistent in appearance. As demonstrated in Fig. 1(b), we should penalty the neighboring pixels 1 and 3 with inconsistent appearance while enhancing the structure between neighboring pixels 1 and 2 with consistent appearances. Second, we observe that the foregrounds are homogeneous in the same concept of spatial region, such as the same superpixel as shown in Fig. 1(c) and (d). Therefore, we further encourage this structure of spatial compactness on the foregrounds in this paper.

In addition, we have observed that structure of the foreground presents diversely on varying scales. As compared in Fig. 1(c) and (d). Coarse scale (with smaller number of superpixels as Fig. 1(c)) imposes global structure of the pattern (such as the whole body imposed in a single superpixel), while fine scale captures local structure (such as the head imposed in a single superpixel in Fig. 1(d)). In this paper, we propose to integrate the multi-scale cues into an unified structured low-rank and sparse factorization model to capture the diverse structure of the foregrounds.

As for the optimization, the low-rank model always suffers from heavy computational burden due to singular value decomposition procedure in most of the existing batch mannered models. To relieve this issue, we introduce an alternative definition of the nuclear-norm with a efficient decomposition strategy in this paper.

As summarized, we make the following contributions for moving object detection and related applications. We propose an effective approach for background subtraction. Our approach takes both appearance consistency and spatial compactness in a multi-scale manner for the separation of background and moving objects. In this way, it is capable of capturing rich information among pixels and thus improves the detection performance significantly. We design an efficient algorithm to optimize the proposed model. In particular, we decompose the low-rank background matrix into two sub-matrices to avoid SVD operations in each iteration, and thus deliver an efficient solver to the proposed model. Extensive experiments on two benchmark datasets GTFD and CDnet demonstrate that, our approach can better preserve the boundary of the foregrounds and achieves big superior performance against the state-of-the-arts.

The rest of paper is organized as follow: Section 2 briefly reviews the related work on low-rank and sparse factorization for moving object detection. We elaborate our approach and the efficient optimization in Section 3. Section 4 demonstrates the experimental results on two benchmark datasets while Section 5 concludes our paper.

2. Related work

In the past years, extensive work have developed on low-rank and sparse factorization for moving object detection. One pioneering work is Robust Principal Component Analysis (RPCA) [19,20], which decomposes a given matrix/frames into a low-rank background matrix and sparse foreground matrix. Candès et al. [21] proposed to recover the low-rank and sparse components individually by solving a convenient convex program called Principal Component Pursuit (PCP). Zhou et al. [22] proposed to handle both small entrywise noises and gross sparse errors. Dou et al. [23] proposed a incremental learning model using K-SVD for dictionary learning. Ebadi et al. [31] constructed the image sequence into the low-rank background and a dynamic tree-structured sparse foreground. Javed et al. [32] introduced a motion-aware prior obtained by optical flow as the regularization of graphs into the low-rank component. To enforce the spatial smoothness, Zhou et al. [24] proposed to relax the requirement of sparse and random distribution of corruption by preserving l_0-penalty and modeling the spatial contiguity of the sequence. CLASS [33] proposed a collaborative framework to leverage the various size of the moving objects via introducing the global appearance consistency. However, they considered the spatial smoothness only according to the coordinates of the pixels while ignoring the appearance similarity. Xin et al. [25] introduced the intensity similarities to the neighboring pixels via regularization terms to enforce the appearance smoothness. Shakeri at al. [34] presented an online sequential framework via solving sequential low-rank approximation and contiguous outlier representation problem. However, these methods constructed the graph only based on pixel level which ignored the spatial compactness. Recently, Javed et al. [29] presented an online matrix decomposition using max-norm constraints on each superpixel segment. However, they only enforce the structured constraint between the spatial neighbors while ignoring the more informative structures of the foreground including the spatial compactness and the multi-scale cues.

3. Our approach

In this section, we will elaborate our multi-scale structured low-rank and sparse factorization model, followed by the alternating optimization algorithm.

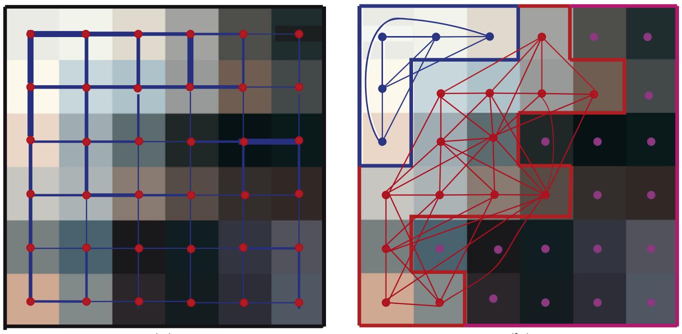

(a) (b)

Fig. 2. Illustration of generating the informative graphs. (a) Weights construction between the neighboring pixels where the thicker edges indicate the higher appearance similarities. (b) Graph construction between the pixel pairs within the same superpixel.

3.1. Problem formulation

We formulate the problem of background subtraction as structured low-rank and sparse factorization model. A video sequence $\mathbf{D} = [\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_n] \in \mathbb{R}^{m\times n}$ is composed of n frames by of m pixels per frame. $\mathbf{B} \in \mathbb{R}^{m\times n}$ is a background matrix, which denotes the underlying background images. Our goal is to discover the object mask $\mathbf{S}$ from data matrices $\mathbf{D}$, where $\mathbf{S}_{ij}$ is a binary matrix:

$$\mathbf{S}_{ij} = \begin{cases} 0, & \text{if } ij \text{ is background}, \\ 1, & \text{if } ij \text{ is foreground}. \end{cases} \tag{1}$$

We assume that the underlying background images with low-rank structure and the foregrounds with sparse and contiguous structure, which has been successfully applied in structured background modeling [24,35]. Furthermore, for the background region where $\mathbf{S}_{ij} = 0$, we assume that $\mathbf{D}_{ij} = \mathbf{B}_{ij} + \epsilon_{ij}$, where $\boldsymbol{\epsilon}_{ij}$ denotes i.i.d. Gaussian noise. Based on the above assumptions, we have:

$$\begin{aligned} &\min_{\mathbf{B}, \mathbf{S}_{ij}\in\{0,1\}} \alpha \parallel vec(\mathbf{S}) \parallel_0 \\ &s.t.\ \mathbf{S}_\perp \circ \mathbf{D} = \mathbf{S}_\perp \circ (\mathbf{B} + \epsilon),\ rank(\mathbf{B}) \le r, \end{aligned} \tag{2}$$

where α is a penalized factor, $\|\mathbf{X}\|_0$ indicates the l_0 norm of a vector. "$\circ$" denotes element-wise multiplication of two matrices, $\mathbf{S}_\perp$ denotes complementary matrix of $\mathbf{S}$, i.e, $\mathbf{S}_\perp + \mathbf{S} = 1$. r is a constant that suppresses the structure complexity of the background model.

3.1.1. Appearance consistency

In order to preserve the spatial structure of the objects, [24,35] constructed the graph based on the neighboring pixels. However, as we discussed in Fig. 1(b), it is essential to penalise the neighbors with inconsistent appearances while enhancing the ones with consistent appearances [25,29]. Therefore, we enforce the appearance consistency into the structure of the informative graphs (as shown in Fig. 2(a)) by:

$$||\mathbf{C}\ vec(\mathbf{S})||_1 = \sum_{(ij,kl)\in\mathcal{N}_1} w_{ij,kl}\ |\mathbf{S}_{ij} - \mathbf{S}_{kl}|, \tag{3}$$

where, $\parallel \mathbf{X} \parallel_1 = \sum_{ij} |\mathbf{X}_{ij}|$ denotes the l_1-norm, $\mathcal{N}_1$ denotes the edge set connecting spatially neighboring pixels, $(ij, kl) \in \mathcal{N}_1$ when pixel ij and kl are spatially connected. $\mathbf{C}$ is the node-edge incidence matrix denoting the connecting relationship among pixels, and $vec(\mathbf{S})$ is a vectorize operator on matrix $\mathbf{S}$. $w_{ij,\ kl}$ indicates the appearance similarity between the pixels which is defined as:

$$w_{ij,kl} = \exp\frac{-||d_{ij} - d_{kl}||_2^2}{2\delta^2}, \tag{4}$$

where d_{ij} and d_{kl} represent the intensity of pixel ij and kl respectively and δ is a tunning parameter which is set as 30 in this paper.

Based on this construction, as shown in Fig. 1(a), the higher probability that a pair of pixels belongs to the same pattern (with closer appearance similarity), the stronger correlation between this pair, which can enforce their appearance consistency.

3.1.2. Spatial compactness

It is observed that, the pixels from the same superpixel, which is a perceptually compact unit with consistent color and texture, are basically derived from the same pattern (background/ foreground) as shown in Fig. 1(c) and (d). In order to enforce the structure of spatial compactness, we first generate the superpixel segmentation via SLIC [30], which is simple and efficient comparing to LRW [36] and GWT [37]. Then, we construct the fully connected graph between the pixels within each superpixel as illustrated in Fig. 2(b). We introduce this structured constraint into the model via:

$$||\mathbf{A}\ vec(\mathbf{S})||_1 = \sum_{(ij,pq)\in\mathcal{N}_2} |\mathbf{S}_{ij} - \mathbf{S}_{pq}|, \tag{5}$$

where, $\mathcal{N}_2$ indicates edge set connecting all the pixel pairs within each superpixel and $\mathbf{A}$ is the node-edge incidence matrix denoting the connecting relationship among pixels. Noted that we don't enforce additional appearance similarities onto the graph within a superpixel since the supperpixel is a conceptual group with similar appearances. In a sense, the compact structure in a superpixel can promote the appearance consistency simultaneously.

3.1.3. Multi-scale integration

Noted that the scale of superpixel segmentation controls the diversity of the foreground structure. As shown in Fig. 1(c) and (d), the coarse/fine scale with smaller/larger number of superpixels represents the global/local structure of the foreground. In order to capture the diverse foreground structures on varying scales, we further integrate above structured constraints into a multi-scale fashion. Considering the $\mathbf{S}^k, k = 1, \ldots, K$ indicates the foreground support matrix on the kth scale, we encourage the foreground supports from varying scales close to the true foreground $\mathbf{S}$. As concluded, we formulate the multi-scale structured foreground integration as:

$$\begin{aligned} &\min_{\mathbf{B}^k, \{\mathbf{S}^k_{ij}, \mathbf{S}_{ij}\}\in\{0,1\}} \sum_{k=1}^{K} \alpha \parallel vec(\mathbf{S}^k) \parallel_0 + ||\mathbf{E}^k\ vec(\mathbf{S}^k)||_1 + \eta \parallel \mathbf{S}^k - \mathbf{S} \parallel_F^2 \\ &\quad s.t.\ \mathbf{S}^k_\perp \circ \mathbf{D} = \mathbf{S}^k_\perp \circ (\mathbf{B}^k + \epsilon),\ rank(\mathbf{B}^k) \le r, \end{aligned} \tag{6}$$

with:

$$||\mathbf{E}^k\ vec(\mathbf{S}^k)||_1 = \beta||\mathbf{C}^k\ vec(\mathbf{S}^k)||_1 + \gamma||\mathbf{A}^k\ vec(\mathbf{S}^k)||_1, \tag{7}$$

where β, γ and η are the tuning parameters leveraging the contribution of appearance consistency, spatial compactness and the multi-scale integration respectively.

3.2. Model optimization

Eq. (6) is a NP-hard problem, to make it tractable, we relax the rank operator on $\mathbf{B}$ with the nuclear norm which has been proven as an effective convex surrogate of the rank operator [38]. Therefore, Eq. (6) can be reformulated as:

$$\begin{aligned} &\min_{\mathbf{B}^k, \{\mathbf{S}^k_{ij}, \mathbf{S}_{ij}\}\in\{0,1\}} \sum_{k=1}^{K} \frac{1}{2}\ ||\mathbf{S}^k{}_\perp \circ (\mathbf{D} - \mathbf{B}^k)||_F^2 + \alpha \parallel vec(\mathbf{S}^k) \parallel_0 \\ &\quad + ||\mathbf{E}^k\ vec(\mathbf{S}^k)||_1 + \lambda \parallel \mathbf{B}^k \parallel_* + \eta \parallel \mathbf{S}^k - \mathbf{S} \parallel_F^2, \end{aligned} \tag{8}$$

where λ is a balance parameter. $||\cdot||_*$ and $||\cdot||_F$ indicate the nuclear norm and the Frobenius norm of a matrix, respectively.

We adopt an alternating algorithm by separating Eq. (8) over $\mathbf{B}^k$, $\mathbf{S}^k$ and $\mathbf{S}$ via solving the following three subproblems.

3.2.1. Solving $\mathbf{B}^k$

Due to the high computational complexity of singular value decomposition via optimizing the nuclear norm, we decompose the background $\mathbf{B}^k$ into two sub-matrices as $\mathbf{B}^k = \mathbf{M}^k\mathbf{N}^k$ inspired by Mazumder et al. [39] and reformat Eq. (8) as:

$$\min_{\{\mathbf{S}^k,\mathbf{S}_{ij}\}\in\{0,1\},\mathbf{M}^k,\mathbf{N}^k} \sum_{k=1}^{K} \frac{1}{2} ||\mathbf{S}^k{}_{\perp} \circ (\mathbf{D} - \mathbf{M}^k\mathbf{N}^k)||_F^2 + \alpha \parallel vec(\mathbf{S}^k) \parallel_0 + ||\mathbf{E}^k\ vec(\mathbf{S}^k)||_1 + \eta \parallel \mathbf{S}^k - \mathbf{S} \parallel_F^2 + \frac{\lambda}{2} \parallel \mathbf{M}^k \parallel_F^2 + \frac{\lambda}{2} \parallel \mathbf{N}^k \parallel_F^2 . \tag{9}$$

Given an current foreground mask $\mathbf{S}^k$, estimating $\mathbf{B}^k$ by minimizing Eq. (8) turns out to be the matrix completion problem. This is to learn a low-rank background matrix from partial observations:

$$\arg\min_{\mathbf{M}^k,\mathbf{N}^k} = ||\mathbf{P}^k - \mathbf{M}^k\mathbf{N}^k||_F^2 + \lambda(\left\|\mathbf{M}^k\right\|_F^2 + \left\|\mathbf{N}^k\right\|_F^2), \tag{10}$$

where $\mathbf{P}^k = \mathbf{S}^k{}_{\perp} \circ \mathbf{D} + \mathbf{S}^k \circ \mathbf{B}^k$, which can be minimized along one coordinate direction at each iteration. We expand this procedure as follows:

First, fix the other variables and update $\mathbf{M}^k$ by solving the problem:

$$\arg\min_{\mathbf{M}^k} ||\mathbf{P}^k - \mathbf{M}^k\mathbf{N}^k||_F^2 + \lambda||\mathbf{M}^k||_F^2, \tag{11}$$

which has a closed-form solution given as:

$$\mathbf{M}^k = \mathbf{P}^k{\mathbf{N}^k}^{\top}(\mathbf{N}^k{\mathbf{N}^k}^{\top} + \lambda\mathbf{I})^{-1}. \tag{12}$$

Then, $\mathbf{N}^k$ plays a symmetric role to $\mathbf{M}^k$ which can be updated by solving:

$$\arg\min_{\mathbf{N}^k} ||\mathbf{P}^k - \mathbf{M}^k\mathbf{N}^k||_F^2 + \lambda||\mathbf{N}^k||_F^2, \tag{13}$$

with a closed-form solution given as:

$$\mathbf{N}^k = ({\mathbf{M}^k}^{\top}\mathbf{M}^k + \lambda\mathbf{I})^{-1}{\mathbf{M}^k}^{\top}\mathbf{P}^k. \tag{14}$$

Finally, $\mathbf{B}^k$ can be achieved via:

$$\mathbf{B}^k = \mathbf{M}^k\mathbf{N}^k. \tag{15}$$

3.2.2. Solving $\mathbf{S}^k$

Given a current estimate of the background position matrix $\mathbf{B}^k$, Eq. (8) can be transferred into following optimization function:

$$\min_{\mathbf{S}^k} \frac{1}{2}||\mathbf{S}^k{}_{\perp} \circ (\mathbf{D} - \mathbf{B}^k)||_F^2 + \alpha \parallel vec(\mathbf{S}^k) \parallel_0 + ||\mathbf{E}^k vec(\mathbf{S}^k)||_1 + \eta \parallel \mathbf{S}^k - \mathbf{S} \parallel_F^2 . \tag{16}$$

The energy function Eq. (16) can be rewritten in line with the standard form of a first-order Markov Random Fields [40] as:

$$\begin{aligned}
&\frac{1}{2} ||\mathbf{S}^k{}_{\perp} \circ (\mathbf{D} - \mathbf{B}^k)||_F^2 + \alpha \parallel vec(\mathbf{S}^k) \parallel_0 + ||\mathbf{E}^k\ vec(\mathbf{S}^k)||_1 \\
&\quad + \eta \parallel \mathbf{S}^k - \mathbf{S} \parallel_F^2 \\
&= \frac{1}{2}\sum_{i,j}(\mathbf{D}_{ij} - \mathbf{B}_{ij}^k)^2(1 - \mathbf{S}_{ij}^k) + \alpha\sum_{i,j}\mathbf{S}_{ij}^k + ||\mathbf{E}^k\ vec(\mathbf{S}^k)||_1 \\
&\quad + \eta\sum_{ij}(1 - 2\mathbf{S}_{ij})\mathbf{S}_{ij}^k \\
&= \sum_{i,j}(\alpha - \frac{1}{2}(\mathbf{D}_{ij} - \mathbf{B}_{ij}^k)^2 + \eta(1 - 2\mathbf{S}_{ij}))\mathbf{S}_{ij}^k + ||\mathbf{E}^k\ vec(\mathbf{S}^k)||_1 \\
&\quad + \frac{1}{2}\sum_{i,j}(\mathbf{D}_{ij} - \mathbf{B}_{ij}^k)^2.
\end{aligned} \tag{17}$$

$\mathbf{S}_{ij}^k$ is a constant, $\frac{1}{2}\sum_{i,j}(\mathbf{D}_{ij} - \mathbf{B}_{ij}^k)^2$ is also a constant with fixed $\mathbf{B}^k$. Meanwhile, $\mathbf{S}_{ij}$ which can be achieved via Eq. (19) is also a constant. Known Markov unary term and pairwise smoothing term, one can easily obtain the optimal foreground matrix though graph cuts method [41,42] since $\mathbf{S}_{ij}^k \in \{0, 1\}$ is discrete.

3.2.3. Solving $\mathbf{S}$

Once attained the current $\mathbf{S}^k$ and $\mathbf{B}^k$ for each scale, Updating $\mathbf{S}$ turns to be:

$$\min_{\mathbf{S}} \sum_{k=1}^{K} \parallel \mathbf{S} - \mathbf{S}^k \parallel_F^2, \tag{18}$$

which has a closed-form solution given as:

$$\mathbf{S} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{S}^k. \tag{19}$$

A sub-optimal solution can be obtained by alternating optimizing $\mathbf{B}^k$, $\mathbf{S}^k$ and $\mathbf{S}$ and the algorithm is summarised in Algorithm 1.

Algorithm 1 Optimization Algorithm to Eq. (8).

Input: $\mathbf{D} = [\mathbf{I}_1, \mathbf{I}_2, ..., \mathbf{I}_n] \in \mathbb{R}^{m\times n}$.
Initialize $\mathbf{B}^k = \mathbf{D}$, $\mathbf{S} = \mathbf{S}^k = \mathbf{0}$, $\tau = 1e-4$, $maxIter = 20$.
Output: $\mathbf{S}, \mathbf{B}^k, \mathbf{S}^k$.

1: **For** $k = 1$ to K do
2: Updating $\mathbf{B}^k$: optimizing energy function Eq. (10) via:
(1) updating $\mathbf{M}^k$: $\mathbf{M}^k \leftarrow \mathbf{P}^k{\mathbf{N}^k}^T(\mathbf{N}^k{\mathbf{N}^k}^{\top} + \lambda\mathbf{I})^{-1}$,
(2) updating $\mathbf{N}^k$: $\mathbf{N}^k \leftarrow ({\mathbf{M}^k}^{\top}\mathbf{M}^k + \lambda\mathbf{I})^{-1}{\mathbf{M}^k}^T\mathbf{P}^k$,
(3) updating $\mathbf{B}^k$: $\mathbf{B}^k \leftarrow \mathbf{M}^k\mathbf{N}^k$.
3: Updating $\mathbf{S}^k$: using graph cuts to optimize energy function Eq. (16) by :
$\mathbf{S}^k \leftarrow arg\,min_{\mathbf{S}^k}\sum_{i,j}(\alpha - \frac{1}{2}(\mathbf{D}_{ij} - \mathbf{B}_{ij}^k)^2 + \eta(1 - 2\mathbf{S}_{ij}))\mathbf{S}_{ij}^k + ||\mathbf{E}^k\ vec(\mathbf{S}^k)||_1 + \frac{1}{2}\sum_{i,j}(\mathbf{D}_{ij} - \mathbf{B}_{ij}^k)^2$.
4: **End For.**
5: Updating $\mathbf{S}$: $\mathbf{S} \leftarrow \frac{1}{K}\sum_{k=1}^{K}\mathbf{S}^k$.
6: Check the convergence condition: if the maximum objective change between two consecutive iterations is less than τ or the maximum number of iterations reaches *maxIter*, then terminate the loop.

---

4. Experiments

We evaluate our approach on GTFD [35] and CDnet14 [43] against five state-of-the-arts including COROLA [34], DECOLOR [24], GMM [8], VIBE [9] and PCP [21]. To keep things fair, we choose the default parameters released by the authors for corresponding methods and resize the resolution of all the frames into 160×120.

4.1. Evaluation settings

4.1.1. Datasets

GTFD [35] dataset consists of 25 video sequence pairs in both visual and thermal modality with various challenges including intermittent motion, low illumination, bad weather, intense shadow, dynamic scene and background clutter etc. We evaluate the proposed method on visual modality videos.

CDnet14 [43] is a large scale dataset consisting of 11 different categories with 55 video sequences. We evaluate our

Table 1
Evaluated parameters on GTFD dataset.

Param	Setting	F-measure	Param	Setting	F-measure	Param	Setting	F-measure
α	$14.4\sigma^2$	0.63	β	0.14	0.63	γ	0.014	0.64
	$16.2\sigma^2$	0.69		0.21	0.69		0.028	0.69
	$18\sigma^2$	0.66		0.28	0.65		0.042	0.63
η	0.006	0.65	γ	7.5	0.63	$\{\mathcal{A}_1, \mathcal{A}_2\}$	{450, 650}	0.67
	0.007	0.69		8.5	0.69		{650, 950}	0.69
	0.008	0.63		9.5	0.65		{450, 950}	0.68

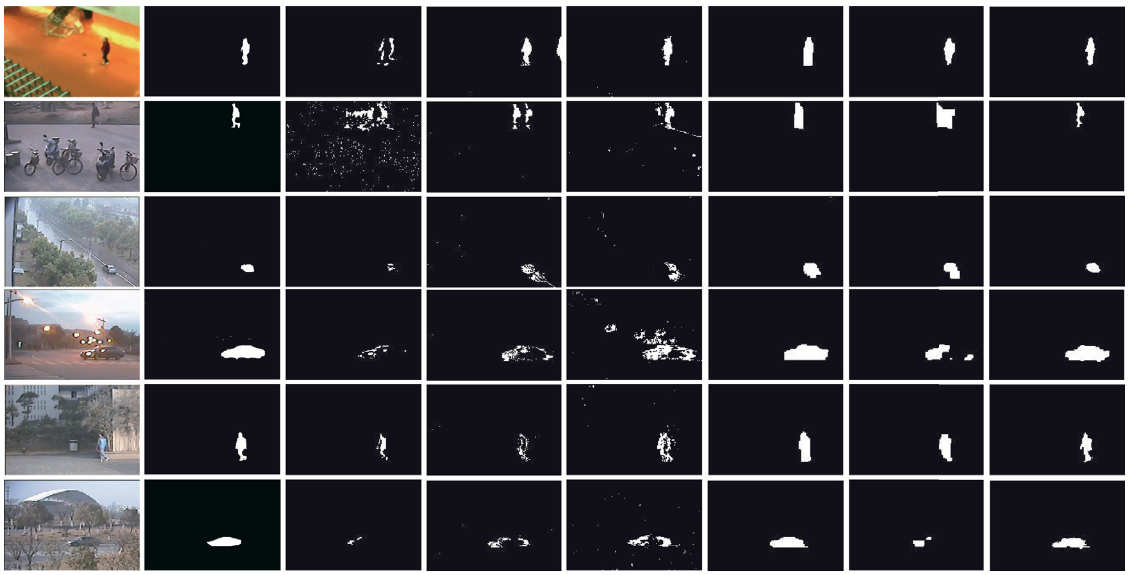


Fig. 3. Sample results of our method against the state-of-the-arts on six videos from GTFD dataset.

method on 10 videos from 5 challenging categories including *Baseline* (Office, PETS2006), *intermittentObjectMotion* (Sofa, winterDriveway), *nightVideos* (TramStation, WinterStreet), *Shadow* (CopyMachine, BusStation) and *Thermal* (Park, Corridor).

4.1.2. Evaluation criterion

The Precision, Recall, F-measure are first comprehensively evaluated, which are defined as following:

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}}, \qquad \text{F-measure} = 2\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}. \tag{20}$$

where the *TP, FP, FN* represent the true positive, false positive and false negative, respectively.

Furthermore, the Mean Absolute Error (MAE) is evaluated to measure the disagreement between the detected results and the groundtruth:

$$MAE = \frac{1}{N \times \mathcal{F}} \sum_{i=1}^{\mathcal{F}} \sum_{p \in DR, \acute{p} \in GT} XOR(p, \acute{p}) \tag{21}$$

where N denotes and resolution of the frame and $\mathcal{F}$ denotes the number of the frames in the video clip. *DR* and *GT* indicate the "Detection Result" and the "Ground Truth" respectively. $XOR(*)$ denotes the logic operator "*exclusive OR*". $p, \acute{p} \in \{0, 1\}$ denotes the background/foreground pixels.

4.1.3. Parameters

There are six important parameters in our model. We adjust one parameter while fixing other parameters and then obtain best performance for our approach. In our model of Eq. (8), the parameter α which controls the sparsity structure of the foreground masks is set as $\alpha = 16.2\sigma^2$, where σ^2 is estimated online by the mean variance of $\{\mathbf{D_{ij}} - \mathbf{B_{ij}}\}$. The parameter β, γ, η and λ control the relative contribution of each term in Eqs. (7) and (8), respectively. We empirically set as $\{\beta, \gamma, \eta, \lambda\} = \{0.21, 0.028, 0.007, 8.5\}$. The number of superpixels on two different scales are $\{\mathcal{A}_1, \mathcal{A}_2\} = \{650, 950\}$ with $K = 2$. We evaluate the parameter variations on GTFD dataset and report the corresponding F-measure in Table 1. It is worth to note that our method is insensitive to the parameters.

4.2. Qualitative results

Figs. 3 and 4 demonstrate the detected results on a certain frame of video clips from GTFD [35] and CDnet14 [43] respectively. From which we can see, our method can produce more precise boundary information and better suppress the influence of the noise. PCP is not able to capture contiguous structure of the foreground since only enforcing the sparse structure of the foreground. GMM and VIBE work on the original pixel space therefore they are quite sensitive to the noise and introduce "ghost". DECOLOR is much more robust but fails to sketch the contours of the objects. COROLA, as the state-of-the-art online detection model, tends to produce the cavities on the objects due to the lack of structure constraint of spatial compactness.

4.3. Quantitative results

Tables 2 and 3 report Precision, Recall, F-measure, and MAE on public GTFD dataset [35] and CDnet14 [43] respectively. It is clear to see that our method significantly outperforms the state-of-the-arts in precision. Although the recall of our method looks lower

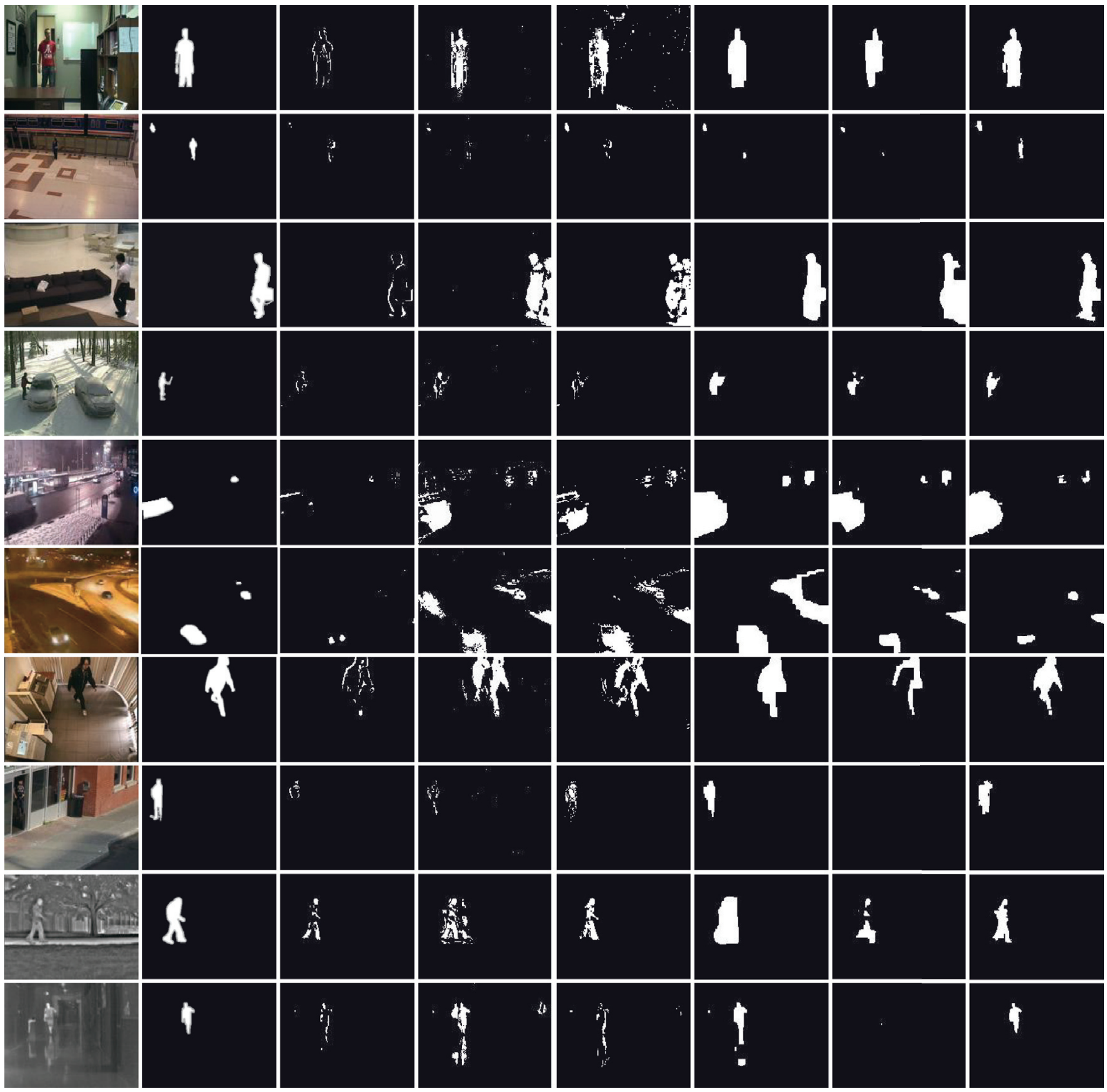


Fig. 4. Sample results of our method against the state-of-the-arts on ten videos from CDnet dataset.

Table 2
The average Precision, Recall, F-measure and MAE values on GTFD dataset.

Algorithm	PCP [21]	VIBE [9]	GMM [8]	DECOLOR [24]	COROLA [34]	OURS
Precision	0.29	0.40	0.51	0.54	0.53	**0.65**
Recall	0.18	0.47	0.64	**0.82**	0.62	0.80
F-measure	0.22	0.39	0.51	0.59	0.52	**0.69**
MAE	0.0155	0.0169	0.0125	0.0061	0.0135	**0.0048**

than DECOLOR [24], from Fig. 3 we can see, DECOLOR [24] tends to produce coarse boundary which always leads to high recall. The more balanced criteria between precision and recall, F-measure, together with the MAE verify the promising performance of our method.

4.4. Component analysis

In order to validate the contribution of our multi-scale structure integration of appearance consistency and spatial compactness, we evaluate several variants of our model on GTFD and report the results on Table 4, where OURS: the proposed model; OURS-I: our model without spatial compactness structure by setting $\gamma = 0$; OURS-II: our model without appearance consistency structure by setting all $w_{ij,kl} = 1$; OURS-III: our model on the single scale (without multi-scale structure) by setting $\eta = 0$ with the number of superpixels $\mathcal{A} = 950$. From Table 4 we can see that: 1) Both appearance consistency and spatial compactness structures play important roles for moving object detection. 2) The multi-scale integration can further boost the performance. Together with the visualized examples on Fig. 5 we can see that: 1) After introducing the spatial compactness structure (comparing OURS to OURS-I), the model can better capture the global spatial coherence and suppress the influence of the noises. 2) After introducing the appearance consistency structure (comparing OURS to OURS-II), the model can better preserve the local contour of the objects. 3)

Table 3
Comparison of Precision (P), Recall (R), F-measure (F) and MAE on videos from CDnet14 dataset.

Methods		PCP [21]	VIBE [9]	GMM [8]	DECOLOR [24]	COROLA [34]	OURS
	P	0.65	0.75	0.81	0.87	0.81	**0.96**
Office	R	0.08	0.73	0.54	**0.83**	0.74	0.82
	F	0.15	0.73	0.63	0.82	0.76	**0.84**
	P	0.69	0.74	0.88	0.64	0.59	**0.89**
PETS2006	R	0.13	0.22	0.27	0.23	0.15	**0.67**
	F	0.22	0.33	0.40	0.30	0.23	**0.75**
	P	0.48	0.47	0.85	0.87	0.82	**0.87**
Sofa	R	0.05	0.36	0.30	0.44	0.36	**0.44**
	F	0.09	0.40	0.43	0.56	0.47	**0.56**
	P	0.27	0.62	0.66	0.64	0.71	**0.73**
WinterDriveway	R	0.05	0.32	0.17	0.48	0.35	**0.71**
	F	0.09	0.41	0.26	0.52	0.37	**0.71**
	P	**0.50**	0.27	0.32	0.41	0.49	0.49
TramStation	R	0.19	0.65	0.54	**0.98**	0.82	0.84
	F	0.27	0.37	0.39	0.57	0.60	**0.61**
	P	0.51	0.21	0.22	0.20	0.45	**0.61**
WinterStreet	R	0.21	0.69	0.68	**0.97**	0.60	0.57
	F	0.29	0.31	0.31	0.34	0.50	**0.56**
	P	0.28	0.48	0.69	0.79	0.77	**0.95**
CopyMachine	R	0.06	0.80	0.68	**0.99**	0.76	0.92
	F	0.10	0.57	0.68	0.87	0.75	**0.93**
	P	**0.92**	0.78	0.87	0.86	0.81	0.86
BusStation	R	0.12	0.34	0.36	**0.78**	0.28	0.76
	F	0.21	0.46	0.50	0.79	0.33	**0.80**
	P	**0.99**	0.67	0.92	0.65	0.89	0.91
Park	R	0.27	0.37	0.45	**0.99**	0.37	0.61
	F	0.42	0.47	0.60	**0.78**	0.48	0.71
	P	0.80	0.46	0.48	0.57	0.66	**0.89**
Corridor	R	0.20	0.68	0.25	0.99	0.16	**0.87**
	F	0.30	0.52	0.30	0.70	0.24	**0.88**
	P	0.61	0.55	0.67	0.65	0.70	**0.82**
Average	R	0.14	0.52	0.42	**0.77**	0.46	0.72
	F	0.21	0.46	0.45	0.63	0.47	**0.74**
	MAE	0.0368	0.0404	0.0349	0.0284	0.0259	**0.0069**

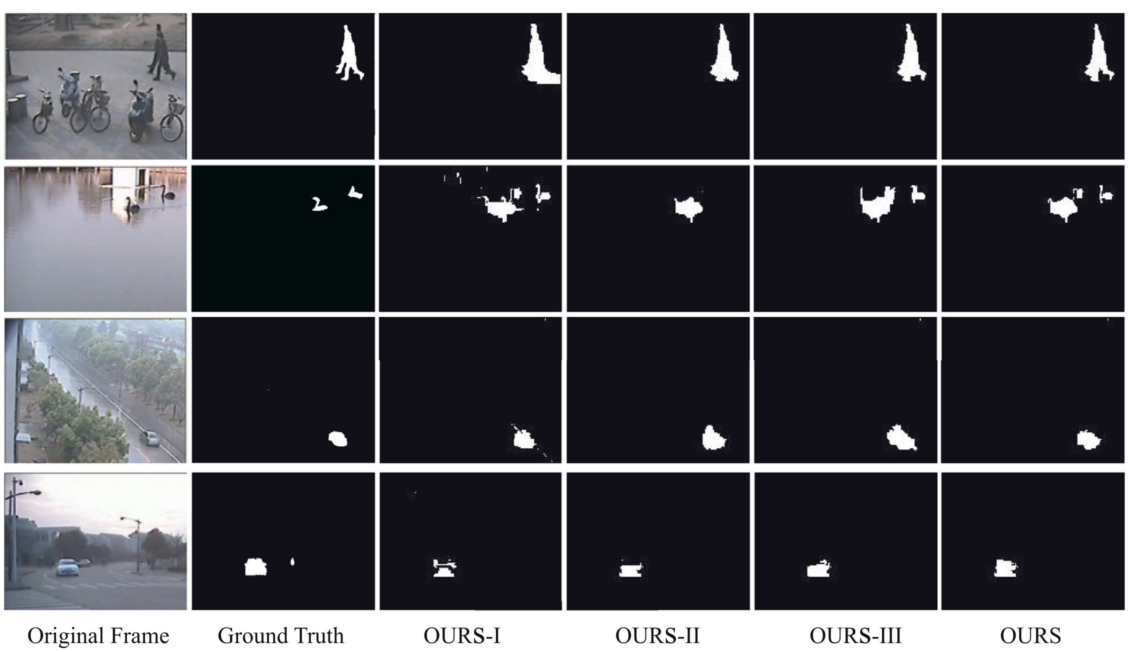


Fig. 5. Example results of our method and its variants on four video sequences from GTFD dataset.

Table 4
Component analysis of our method and its variants on the GTFD dataset.

Algorithm	Precision	Recall	F-measure
OURS	**0.65**	0.80	**0.69**
OURS-I	0.58	0.81	0.64
OURS-II	0.62	0.78	0.66
OURS-III	0.60	**0.82**	0.66

By multi-scale integration, our model yields to better global and local consistency in both spatial and appearance aspects.

4.5. Computational complexity

Our method are implemented on the mixing platform of C++ and MATLAB without any code optimization on a desktop with an

Table 5
Computational complexity (in fps) and F-measure comparison on GTFD dataset.

	PCP	VIBE	GMM	DECOLOR	COROLA	OURS	OURS-SS	OURS-SS+SI
Code Type	C++	C++	MATLAB	MATLAB & C++	MATLAB & C++	MATLAB & C++	MATLAB & C++	MATLAB & C++
FPS	57	60	65	2.59	11	1.13	2.14	1.55
F-measure	0.22	0.39	0.51	0.59	0.52	0.69	0.69	0.70

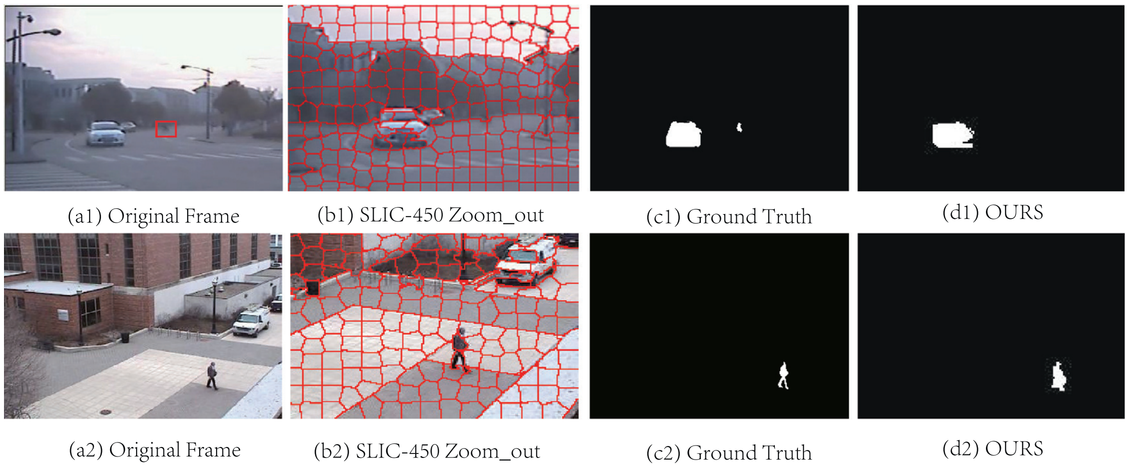


Fig. 6. Illustration of limitation and failure example of our approach from video “car3” and “pedestrain7” from GTFD dataset [35].

Intel i7 3.4 GHz CPU and 32 GB RAM. Runtime together with the F-measure of our method against other methods on GTFD dataset is presented in Table 5, where OURS-SS: our model excludes the time for superpixel segmentation, OURS-SS+SI: OURS-SS model solving **B** via conventional SOFT-INPUT [44] instead of our decomposition strategy. From which we can see: 1) Though GMM, VIBE and PCP run much faster than ours, these methods perform significantly worse. 2) Our method performs slight slower than DECOLOR and COROLA. The main reason is the time for superpixel segmentation and more connections on the informative graphs. 3) The matrix decomposition strategy can speed up the optimization comparing to the conventional SOFT-INPUT [44] without losing accuracy. Our method achieves significantly superior accuracy. We believe that it can be accelerated by further code optimization.

4.6. Limitation

We also encounter unsatisfactory results and detection failure as shown in Fig. 6. (1) The unfaithful superpixel segmentation of the car in Fig. 6(b1) and the pedestrian in Fig. 6(b2) will introduce false contour of the moving object as shown in Fig. 6(d1) and (d2). (2) The tiny object (the bicycle man indicated in Fig. 6(a1)) moving in the comparative dark environment fails to be detected as shown in Fig. 6(d1). The first limitation could be refined by more robust superpixel segmentation method. The second failure could be significantly alleviated via multi-modal resources like thermal images.

5. Conclusion

In this paper, we have proposed a novel method for moving object detection via multi-scale structured low-rank and sparse factorization. We first encourage the structure of the neighboring pixels with close appearance. Then we explore the spatial compactness structure among the pixels within the same superpixel. Finally we integrate the multi-scale coherence with different number of superpixels into a unified framework. We optimize the proposed model via an efficient alternating algorithm. Extensive experiments against state-of-the-arts on the public datasets suggest that, the proposed method can better preserve the boundary of the objects and is robust to the noise. In future work, we will focus on extending our model to online or streaming fashion for real-life applications.

Acknowledgment

This study was partially funded by the National Nature Science Foundation of China (61502006, 61702002, 61602001, 61671018), the Natural Science Foundation of Anhui Province (1508085QF127), and Co-Innovation Center for Information Supply & Assurance Technology, Anhui University.

References

[1] Q. Ling, J. Yan, F. Li, Y. Zhang, A background modeling and foreground segmentation approach based on the feedback of moving objects in traffic surveillance systems, Neurocomputing, 133 (2014) 32–45.
[2] Y. Chong, W. Chen, Z. Li, W.H. Lam, C. Zheng, Q. Li, Integrated real-time vision-based preceding vehicle detection in urban roads, Neurocomputing, 116 (2013) 144–149.
[3] L. Song, F. Jiang, Z. Shi, R. Molina, A.K. Katsaggelos, Toward dynamic scene understanding by hierarchical motion pattern mining, IEEE Trans. Intell. Transp. Syst. 15 (3) (2014) 1273–1285.
[4] K.A. Joshi, D.G. Thakore, A survey on moving object detection and tracking in video surveillance system, Int. J. Soft Comput. Eng. 2 (3) (2012) 44–48.
[5] T. Bouwmans, Traditional and recent approaches in background modeling for foreground detection: an overview, Comput. Sci. Rev. 11 (2014) 31–66.
[6] T. Bouwmans, A. Sobral, S. Javed, S.K. Jung, E.-H. Zahzah, Decomposition into low-rank plus additive matrices for background/foreground separation: a review for a comparative evaluation with a large-scale dataset, Comput. Sci. Rev. 23 (2017) 1–71.
[7] W. Kim, C. Jung, Illumination-invariant background subtraction: comparative review, models, and prospects, IEEE Access 5 (2017) 8369–8384.
[8] P. KaewTraKulPong, R. Bowden, An improved adaptive background mixture model for real-time tracking with shadow detection, in: Video-based Surveillance Systems, Springer, 2002, pp. 135–144.
[9] O. Barnich, M. Van Droogenbroeck, Vibe: a universal background subtraction algorithm for video sequences, IEEE Trans. Image Process. 20 (6) (2011) 1709–1724.
[10] L. Qin, B. Sheng, W. Lin, W. Wu, R. Shen, Gpu-accelerated video background subtraction using gabor detector, J. Vis. Commun. Image Represent. 32 (2015) 1–9.
[11] X. Zhang, C. Zhu, H. Wu, Z. Liu, Y. Xu, An imbalance compensation framework for background subtraction, IEEE Trans. Multim. 19 (11) (2017) 2425–2438.
[12] Y. Li, H. Qi, J. Dai, X. Ji, and Y. Wei, “Fully convolutional instance-aware semantic segmentation,” CoRR, vol. abs/1611.07709, 2016.

[13] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs., IEEE Trans. Pattern Anal. Mach. Intell. PP (99) (2016) 1–14.
[14] Y. Wang, Z. Luo, P.-M. Jodoin, Interactive deep learning method for segmenting moving objects, Pattern Recognit. Lett. 96 (2017) 66–75.
[15] Y. Zhang, X. Li, Z. Zhang, F. Wu, L. Zhao, Deep learning driven blockwise moving object detection with binary scene modeling, Neurocomputing, 168 (2015) 454–463.
[16] B. Heo, K. Yun, J.Y. Choi, Appearance and motion based deep learning architecture for moving object detection in moving camera, in: IEEE International Conference on Image Processing (ICIP), 2017, pp. 1–5.
[17] Y. Chen, J. Wang, B. Zhu, M. Tang, H. Lu, Pixel-wise deep sequence learning for moving object detection, IEEE Trans. Circuits Syst. Video Technol. PP (99) (2017) 1–13.
[18] L. Ang Lim and H. Yalim Keles. Foreground segmentation using a triplet convolutional neural network for multiscale feature encoding. CoRR, abs/1801.02225, 2018.
[19] F.D.L. Torre, M.J. Black, A framework for robust subspace learning, Int. J. Comput. Vis. 54 (1–3) (2003) 117–142.
[20] B. Jiang, C. Ding, J. Tang, Graph-laplacian pca: Closed-form solution and robustness, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 3492–3498.
[21] E.J. Candès, X. Li, Y. Ma, J. Wright, Robust principal component analysis? J. ACM (JACM), 58 (3) (2011) 1–36.
[22] Z. Zhou, X. Li, J. Wright, E. Candes, Y. Ma, Stable principal component pursuit, in: IEEE International Symposium on Information Theory, IEEE, 2010, pp. 1518–1522.
[23] J. Dou, J. Li, Q. Qin, Z. Tu, Moving object detection based on incremental learning low rank representation and spatial constraint, Neurocomputing, 168 (C) (2015) 382–400.
[24] X. Zhou, C. Yang, W. Yu, Moving object detection by detecting contiguous outliers in the low-rank representation, IEEE Trans. Pattern Anal. Mach. Intell. 35 (3) (2013) 597–610.
[25] B. Xin, Y. Tian, Y. Wang, W. Gao, Background subtraction via generalized fused lasso foreground modeling, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 4676–4684.
[26] B. Xin, Y. Kawahara, Y. Wang, W. Gao, Efficient generalized fused lasso and its application to the diagnosis of alzheimer's disease., in: Proceedings of the AAAI Conference on Artificial Intelligence, 2014, pp. 2163–2169.
[27] J. Xu, V.K. Ithapu, L. Mukherjee, J.M. Rehg, V. Singh, Gosus: Grassmannian online subspace updates with structured-sparsity, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2013, pp. 3376–3383.
[28] M.Y. Yang, S. Feng, H. Ackermann, B. Rosenhahn, Global and local sparse subspace optimization for motion segmentation, ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci. 2 (2015) 475–482.
[29] S. Javed, S.H. Oh, A. Sobral, T. Bouwmans, S.K. Jung, Background subtraction via superpixel-based online matrix decomposition with structured foreground constraints, in: IEEE International Conference on Computer Vision Workshop, 2016, pp. 930–938.
[30] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Süsstrunk, Slic superpixels compared to state-of-the-art superpixel methods, IEEE Trans. Pattern Anal. Mach. Intell. 34 (11) (2012) 2274–2282.
[31] S.E. Ebadi, V.G. Ones, E. Izquierdo, Dynamic tree-structured sparse rpca via column subset selection for background modeling and foreground detection, in: 2016 IEEE International Conference on Image Processing (ICIP), IEEE, 2016, pp. 3972–3976.
[32] S. Javed, S.K. Jung, A. Mahmood, T. Bouwmans, Motion-aware graph regularized rpca for background modeling of complex scenes, in: International Conference on Pattern Recognition (ICPR), 2016, pp. 120–125.
[33] A. Zheng, M. Xu, B. Luo, Z. Zhou, C. Li, Class: collaborative low-rank and sparse separation for moving object detection, Cognit. Comput. 9 (2) (2017) 180–193.
[34] M. Shakeri, H. Zhang, Corola: a sequential solution to moving object detection using low-rank approximation, Comput. Vis. Image Underst. 146 (2016) 27–39.
[35] C. Li, X. Wang, L. Zhang, J. Tang, H. Wu, L. Lin, Weighted low-rank decomposition for robust grayscale-thermal foreground detection, IEEE Trans. Circuits Syst. Video Technol. 27 (4) (2017) 725–738.
[36] J. Shen, Y. Du, W. Wang, X. Li, Lazy random walks for superpixel segmentation, IEEE Trans. Image Process. 23 (4) (2014) 1451.
[37] Q. Miao, T. Liu, J. Song, M. Gong, Y. Yang, Guided superpixel method for topographic map processing, IEEE Trans. Geosci. Remote Sens. 54 (11) (2016) 6265–6279.
[38] B. Recht, M. Fazel, P.A. Parrilo, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, SIAM Revi. 52 (3) (2010) 471–501.
[39] R. Mazumder, T. Hastie, R. Tibshirani, Spectral regularization algorithms for learning large incomplete matrices, J. Mach. Learn. Res. 11 (2010) 2287–2322.
[40] S.Z. Li, Markov Random Field Modeling in Image Analysis, Springer London, 2009.
[41] Y. Boykov, O. Veksler, R. Zabih, Fast approximate energy minimization via graph cuts, IEEE Trans. Pattern Anal. Mach. Intell. 23 (11) (2001) 1222–1239.
[42] V. Kolmogorov, R. Zabin, What energy functions can be minimized via graph cuts? IEEE Trans. Pattern Anal. Mach. Intell. 26 (2) (2004) 147–159.
[43] Y. Wang, P.M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, P. Ishwar, Cdnet 2014: an expanded change detection benchmark dataset, in: Computer Vision and Pattern Recognition Workshops (CVPRW), 2014, pp. 393–400.
[44] R.T. Rahul Mazumder, Trevor Hastie, Spectral regularization algorithms for learning large incomplete matrices, J. Mach. Learn. Res. 11 (11) (2010) 2287–2322.

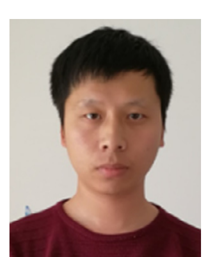
Aihua Zheng received her B.Eng. degrees and finished her Master-Docter combined program in computer science and technology from Anhui University of China in 2006 and 2008 respectively. And received her Ph.D. degree in computer science from the University of Greenwich of UK in 2012. She is currently a Lecturer in Anhui University. Her main research areas are visual based signal processing and pattern recognition.

Tian Zou received her B.E. degree in information and computing science in 2016 from Anqing Normal University, Anqing, China. She is currently pursuing the M.S. degree in computer science and technology at Anhui University. Her current research is moving object detection.

Yumiao Zhao received her B.E. degree in digital media technology in 2016 from Anqing Normal University, Anqing, China. She is currently pursuing the M.S. degree in computer science and technology at Anhui University. Her current research is object detection.

Bo Jiang received the B.S. degrees in mathematics and applied mathematics and the M.Eng. and Ph.D. degrees in computer science from Anhui University of China in 2009, 2012 and 2015, respectively. He is currently an associated professor in computer science at Anhui University. His current research interests include image feature extraction and matching, data representation and learning.

Jin Tang received the B.Eng. degree in automation and the Ph.D. degree in computer science from Anhui University, Hefei, China, in 1999 and 2007, respectively. He is a Professor with the School of Computer Science and Technology, Anhui University. His research interests include computer vision, pattern recognition, and machine learning.



Chenglong Li received the M.S. and Ph.D. degrees from the School of Computer Science and Technology, Anhui University, Hefei, China, in 2013 and 2016, respectively. From 2014 to 2015, he was a visiting student with the School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China. He is currently a lecturer at the School of Computer Science and Technology, Anhui University, and also a postdoctoral research fellow at the Center for Research on Intelligent Perception and Computing (CRIPAC), National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences (CASIA), China. He was a recipient of the ACM Hefei Doctoral Dissertation Award in 2016.