Moving Object Detection via Integrating Spatial Compactness and Appearance Consistency in the Low-Rank Representation

Minghe Xu, Chenglong Li, Hanqin Shi, Jin Tang, and Aihua Zheng[✉]

School of Computer Science and Technology, Anhui University,
No. 111 Jiulong Road, Hefei 230601, China
{xuminghe001,lcl1314,shihq726}@foxmail.com, {tj,ahzheng214}@ahu.edu.cn

Abstract. Low-rank and sparse separation models have been successfully applied to background modeling and achieved promising results on moving object detection. It is still a challenging task in complex environment. In this paper, we propose to enforce the spatial compactness and appearance consistency in the low-rank and sparse separation framework. Given the data matrix that accumulates sequential frames from the input video, our model detects the moving objects as sparse outliers against the low-rank structure background. Furthermore, we explore the spatial compactness by enforcing the consistency among the pixels within the same superpixel. This strategy can simultaneously promote the appearance consistency since the superpixel is defined as the pixels with homogenous appearance nearby the neighborhood. The extensive experiments on public GTD dataset suggest that, our model can better preserve the boundary information of the objects and achieves superior performance against other state-of-the-arts.

Keywords: Low-rank representation · Smoothness constraint
Spatial compactness · Appearance consistency

1 Introduction

Moving object detection is a fundamental problem in video analysis, and plays a critical role in numerous vision applications, such as intelligent transportation [1], vehicle navigation [25] and scene understanding [17]. Over the years, many approaches have been proposed for moving object detection while background subtraction has been recognized as one of the most competitive approaches.

Conventional background modelling methods include single Gaussian distribution [23], Mixture of Gaussian [7,21], and their variations, VIBE [2] and fuzzy concepts based methods [3]. However, these methods model the background for each pixel independently and lack of the relations between the consecutive frames, thus they are very sensitive to noises and occlusions.

M. Xu—Master student.

Recently, the low-rank and sparse separation framework has emerged by decomposing the video sequence into low-rank background and sparse foregrounds (moving objects). One pioneering work is Robust Principal Component Analysis (RPCA) [12,14,22], which decomposes a given matrix/frames into a low-rank background matrix and sparse foreground matrix. Candès et al. [6] proposed to recover the low-rank and sparse components individually by solving a convenient convex program called Principal Component Pursuit (PCP). Zhou et al. [27] proposed to handle both small entrywise noises and gross sparse errors. Dou et al. [10] proposed a incremental learning based LRR model using K-SVD for dictionary learning. Zhou et al. [26] 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. In order to enforce the appearance consistency onto the spatial neighboring relationship, Xin et al. [24] introduced the intensity similarities to the neighboring pixels via regularization terms for both the foreground and background matrices. However, these methods constructed the graph only based on pixel level which ignored the spatial compactness. Recently, Javed et al. [11] proposed a superpixel-based online matrix decomposition method which separate the low-rank background and sparse foregrounds on the superpixel level. However, the performance may excessively rely on the superpixel prior which often produces unfaithful segmentation.

As we observed, the objects are generally spatially compact and consistent in appearance which means the pixels in the same concept of spatial region with close appearance tend to belong to the same pattern (foreground/background). Based on this observation, our main effort is to explore the spatial compactness and the appearance consistency of the objects based on the general framework of low-rank and sparse separation. Specifically, we first encourage the appearance consistency for the object by weighting the neighboring pixel pairs with the appearance similarity. Furthermore, we enforce the global spatial compactness on the superpixel level by constructing the informative graphs for the pixels within the same superpixel. Noted that the superpixel strategy can also promote the appearance consistency since a superpixel is defined as the perceptually consistent unit in appearance.

2   Our Approach

In this section, we will present our model by elaborating the enforcement of spatial compactness and appearance consistency in the low-rank and sparse separation framework, followed by the alternating optimization algorithm.

2.1   Problem Formulation

In this paper, we formulate the problem of foreground detection as a low-rank and sparse separation model. A video sequence $\mathbf{D} = [\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n] \in \mathbb{R}^{m \times n}$ 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 are linearly correlated and the foregrounds are sparse and contiguous, which has been successfully applied in background modeling [16,26]. Furthermore, for the background region where $\mathbf{S}_{ij} = 0$, we assume that $\mathbf{D}_{ij} = \mathbf{B}_{ij} + \epsilon_{ij}$, where ϵ_{ij} denotes i.i.d. Gaussian noise. Based on the above assumptions, we have:

$$\min_{\mathbf{B},\mathbf{S}_{ij}\in\{0,1\}} \alpha\|vec(\mathbf{S})\|_0$$
$$s.t. \ \mathbf{S}_\perp \circ \mathbf{D} = \mathbf{S}_\perp \circ (\mathbf{B} + \epsilon), \ rank(\mathbf{B}) \le r, \tag{2}$$

where α is a penalized factor, and $\|\mathbf{X}\|_0$ indicates the l_0 norm of a vector. The operator "$\circ$" denotes element-wise multiplication of two matrices, $\mathbf{S}_\perp$ denotes the region of $\mathbf{S}_{ij} = 0$, and r is a constant that suppresses the complexity of the background model.

Appearance consistency. Due to the non-convexity of l_0 norm of the matrix $\mathbf{S}$, a common practice is to introduce a contiguous constraint to form a MRF [8] model which can be solved by graph cuts [4,15]. In order to preserve the spatial smoothness of the objects, [16,26] constructed the graph based on the neighboring pixels. However, it is necessary to enforce the appearance similarity onto the neighboring pixels for the informative graphs [11,24]. Therefore, we construct the smoothness by:

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

where, $\|\mathbf{X}\|_1 = \sum_{\mathbf{ij}} |\mathbf{X}_{\mathbf{ij}}|$ denotes the l_1-norm, ε denotes the edge set connecting spatially neighboring pixels, $(ij, kl) \in \varepsilon$ 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 $\mathbf{S}$. Among them, consider the first term $\|\mathbf{C} \ vec(\mathbf{S})\|_1$ in Eq. (7) represents the difference between the adjacent pixels $w_{ij,kl}$ indicates the adaptive weighting factor between the pixels and is defined as:

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

where d_{ij} and d_{kl} represent the intensity of pixel ij and kl respectively and σ is a tunning parameter. Based on this construction, as shown in Fig. 1(a), the higher probability that a pair of pixels belongs to the same segment (with close intensity), the stronger correlation between this pair, which can further enforce the appearance consistency between neighboring pixels.
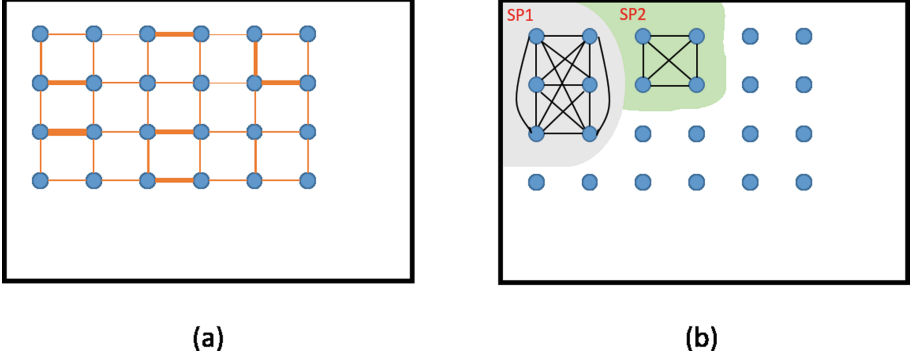
Fig. 1. Illustration of generating the informative graphs. (a) Constructing the weighted graphs for the neighboring pixels where the thicker links between pixel pairs indicate the higher appearance similarity. (b) Constructing graphs between the pixel pairs within the same superpixel. (Color figure online)

Spatial compactness. It is observed that, the pixels from the same superpixel, which is a perceptually consistent unit in color and texture, are basically derived from the same concept (background/foreground). In order to enforce this spatial compactness, we further construct the fully connected graph between the pixels within each superpixel (as shown in Fig. 1(b)) generated by the lazy random walks (LRW) [20] and introduce the spatial compactness into the model via:

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

where, $\mathcal{N}$ 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. It can also promote the appearance consistency since the superpixel consists of the consistent unit in color and texture.

As concluded, we can integrate our formulation by enforcing the spatial compactness and appearance consistency into Eq. (2) as:

$$\min_{\mathbf{B},\mathbf{S}_{ij}\in\{0,1\}} \alpha||vec(\mathbf{S})||_0 + \ \mu||\mathbf{E} \ vec(\mathbf{S})||_1,$$
$$s.t. \ \mathbf{S}_\perp \circ \mathbf{D} = \mathbf{S}_\perp \circ (\mathbf{B} + \epsilon), \ rank(\mathbf{B}) \leq r, \tag{6}$$

with:

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

where μ, β and γ are tuning parameters.

2.2   Model Optimization

Equation (6) is a NP-hard problem, to make Eq. (6) tractable, we relax the rank operator on $\mathbf{B}$ with the nuclear norm, the nuclear norm has proven to be an

effective convex surrogate of the rank operator [19]. Therefore, Eq. (6) can be reformulated as:

$$\min_{\mathbf{B}, \mathbf{S}_{ij} \in \{0,1\}} \frac{1}{2} \, ||P_{\mathbf{S}_\perp}(\mathbf{D} - \mathbf{B})||_F^2 + \alpha ||vec(\mathbf{S})||_0 + \mu ||\mathbf{E} \, vec(\mathbf{S})||_1 + \lambda ||\mathbf{B}||_*, \quad (8)$$

where λ is a balance parameter. $|| \cdot ||_*$ and $|| \cdot ||_F$ indicate the nuclear norm of a matrix and the Frobenius norm of a matrix, respectively. $P_{\mathbf{S}_\perp}(\mathbf{X})$ is the complement to $P_{\mathbf{S}}(\mathbf{X})$ which is the orthogonal projection of matrix $\mathbf{X}$ denoted by:

$$P_{\mathbf{S}}(\mathbf{X})(i,j) = \begin{cases} 0, & if \ \mathbf{S}_{ij} = 0, \\ \mathbf{X}_{ij}, & if \ \mathbf{S}_{ij} = 1, \end{cases} \quad (9)$$

Therefore, we adopt an alternating algorithm by separating Eq. (8) over $\mathbf{B}$ and $\mathbf{S}$ in the following two steps.

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

$$\min_{\mathbf{B}} \frac{1}{2} \, ||P_{\hat{\mathbf{S}}_\perp}(\mathbf{D} - \mathbf{B})||_F^2 + \lambda ||\mathbf{B}||_*, \quad (10)$$

The optimal $\mathbf{B}$ in Eq. (13) can be computed by the SOFT-IMPUTE [18] algorithm. Which based on the following Lemma [5]:

Lemma 1. Given a matrix $\mathbf{Z}$, the solution to the optimization problem

$$\min_{\mathbf{X}} \frac{1}{2} \, ||\mathbf{Z} - \mathbf{X}||_F^2 + \lambda ||\mathbf{X}||_*, \quad (11)$$

is given by $\hat{\mathbf{X}} = \Theta_\lambda(\mathbf{Z})$, where Θ_λ means the singular value thresholding

$$\Theta_\lambda(\mathbf{Z}) = \mathbf{U}\Sigma_\lambda\mathbf{V}^T, \quad (12)$$

Here, $\Sigma_\lambda = diag[(d_1 - \lambda)_+, \cdots, (d_r - \lambda)_+]$, $\mathbf{U}\Sigma_\lambda\mathbf{V}^T$ is the SVD of $\mathbf{Z}$, $\Sigma = diag[d_1 - d_r]$ and $t_+ = max(t, 0)$. Rewriting Eq. (10), we have:

$$\min_{\mathbf{B}} \frac{1}{2} \, ||P_{\hat{\mathbf{S}}_\perp}(\mathbf{D} - \mathbf{B})||_F^2 + \lambda ||\mathbf{B}||_*,$$
$$= \min_{\mathbf{B}} \frac{1}{2} \, ||[P_{\hat{\mathbf{S}}_\perp}(\mathbf{D}) + P_{\hat{\mathbf{S}}}(\mathbf{B})] - \mathbf{B})||_F^2 + \lambda ||\mathbf{B}||_*, \quad (13)$$

According to Lemma 1, given an arbitrary initialization $\hat{\mathbf{B}}$, the optimal solution can be obtained by iteratively using Eq. (14):

$$\hat{\mathbf{B}} \longleftarrow \Theta_\lambda(P_{\hat{\mathbf{S}}_\perp}(\mathbf{D}) + P_{\hat{\mathbf{S}}}(\hat{\mathbf{B}})), \quad (14)$$

S− *subproblem.* Given an current estimate of the background position matrix $\hat{\mathbf{B}}$, Eq. (8) can be transferred into following optimization functions:

$$\min_{\mathbf{S}} \frac{1}{2} \, ||P_{\mathbf{S}_\perp}(\mathbf{D} - \hat{\mathbf{B}})||_F^2 + \alpha ||vec(\mathbf{S})||_0 + \mu ||\mathbf{E} \, vec(\mathbf{S})||_1, \quad (15)$$

Algorithm 1. Optimization Algorithm to Eq. 8

Require: $\mathbf{D} = [\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_n] \in \mathbb{R}^{m \times n}$.
  Set $\mathbf{B} = \mathbf{D}$, $\mathbf{S} = \mathbf{0}$, $\tau = 1e - 4$, $maxIter = 20$.
Ensure: $\hat{\mathbf{S}}, \hat{\mathbf{B}}$.
 1: Using SOFT-IMPUTE algorithm to optimize energy function Eq. (10), by comput-
    ing $\hat{\mathbf{B}}$: $\hat{\mathbf{B}} \longleftarrow \Theta_\lambda(P_{\hat{\mathbf{S}}_\perp}(\mathbf{D}) + P_{\hat{\mathbf{S}}}(\hat{\mathbf{B}}))$
 2: **if** $rank(\hat{\mathbf{B}}) \leq K$ **then**
 3:    tuning parameters λ, returns run to step 1.
 4: **end if**
 5: Using graph cuts algorithm to optimize energy function Eq. (15) by computing $\hat{\mathbf{S}}$:
    $\hat{\mathbf{S}} = arg\ min_{\mathbf{S}} \sum_{i,j}(\alpha - \frac{1}{2}(\mathbf{D}_{ij} - \hat{\mathbf{B}}_{ij}))^2 \mathbf{S}_{ij} + \mu||\mathbf{E}\ vec(\mathbf{S})||_1$
 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.

The energy function Eq. (15) can be rewritten in line with the standard form of a first-order Markov Random Fields [8] as:

$$\frac{1}{2}\,||P_{\mathbf{S}_\perp}(\mathbf{D} - \hat{\mathbf{B}})||_F^2 + \alpha||vec(\mathbf{S})||_0 + \mu||\mathbf{E}\ vec(\mathbf{S})||_1$$

$$= \frac{1}{2}\sum_{i,j}(\mathbf{D}_{ij} - \hat{\mathbf{B}}_{ij})^2(1 - \mathbf{S}_{ij}) + \alpha\sum_{i,j}\mathbf{S}_{ij} + \mu||\mathbf{E}\ vec(\mathbf{S})||_1, \qquad (16)$$

$$= \sum_{i,j}(\alpha - \frac{1}{2}(\mathbf{D}_{ij} - \hat{\mathbf{B}}_{ij}))^2 \mathbf{S}_{ij} + \mu||\mathbf{E}\ vec(\mathbf{S})||_1 + \frac{1}{2}\sum_{i,j}(\mathbf{D}_{ij} - \hat{\mathbf{B}}_{ij})^2.$$

When $\hat{\mathbf{B}}$ is fixed, $\frac{1}{2}\sum_{i,j}(\mathbf{D}_{ij} - \hat{\mathbf{B}}_{ij})^2$ is constant. Meanwhile, $\mathbf{S}_{ij}$ beside the $(\beta - \frac{1}{2}(\mathbf{D}_{ij} - \hat{\mathbf{B}}_{ij}))^2$ is also constant. Known Markov unary term and pairwise smoothing term, one can easily obtain the optimal foreground matrix though graph cuts method [4,15] since $\mathbf{S}_{ij} \in \{0, 1\}$ is discrete.

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

3 Experiments

We evaluate our method against the state-of-the-arts on the public challenging GTD dataset [16]. It consists of 25 video sequence pairs in both visual and thermal modality. In this paper, we evaluate the proposed method on visual modality videos. The GTD dataset [16] contains fifteen different scenes and various challenges including intermittent motion, low illumination, bad weather, intense shadow, dynamic scene and background clutter etc.

3.1 Evaluation Settings

Parameters. In our model of Eq. (8), the parameter λ controls the complexity of the background model which is first roughly estimated by the rank of the

background model. The parameter α which controls the sparsity of the foreground masks is set as $\alpha = 16.2\sigma^2$, where σ^2 is estimated online by the mean variance of $\{\mathbf{D_{ij}} - \hat{\mathbf{B}}_{\mathbf{ij}}\}$. The parameter μ controls spatial smoothness between pixels that satisfies the constructed informative graphs, and is set as $\mu = 0.205$. The parameter β and γ control the relative contribution of each term in Eq. (7), respectively. We determine β and γ by adjusting its ratio to α, and empirically set as $\{\beta, \gamma\} = \{2.7\alpha, 0.13\alpha\}$. Moreover, we set $\sigma = 25$ in Eq. (3), and set the number of superpixel patches $\mathcal{A} = 650$.

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

$$
\begin{aligned}
\text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\
\text{Recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\
\text{F-measure} &= 2\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.
\end{aligned}
\tag{17}
$$

where TP = True Positives, indicating the foreground pixels correctly labeled as foreground. FP = False Positives, referring the background pixels incorrectly labeled as foreground. TN = True Negatives, corresponding to background pixels correctly labeled as background. FN = False Negatives, referring to foreground pixels incorrectly labeled as background [9]. F-measure is a comprehensive measurement to balance the argument between precision and recall.

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

$$
MAE = \frac{1}{N \times \mathcal{F}} \sum_{i=1}^{\mathcal{F}} \sum_{p \in DR, \acute{p} \in GT} XOR(p, \acute{p})
\tag{18}
$$

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.

3.2 Comparison Results

We compare our approach with four state-of-the-art moving object detection algorithms including DECOLOR [26], GMM [13], VIBE [2] and PCP [6]. To keep things fair, we choose the default parameters released by the authors for corresponding methods.

Qualitative Results. Figure 2 demonstrates the detected results on a certain frame of six video clips from GTD dataset [16]. From which we can see, our method can produce finer boundary information and better suppress the influence of the noise.
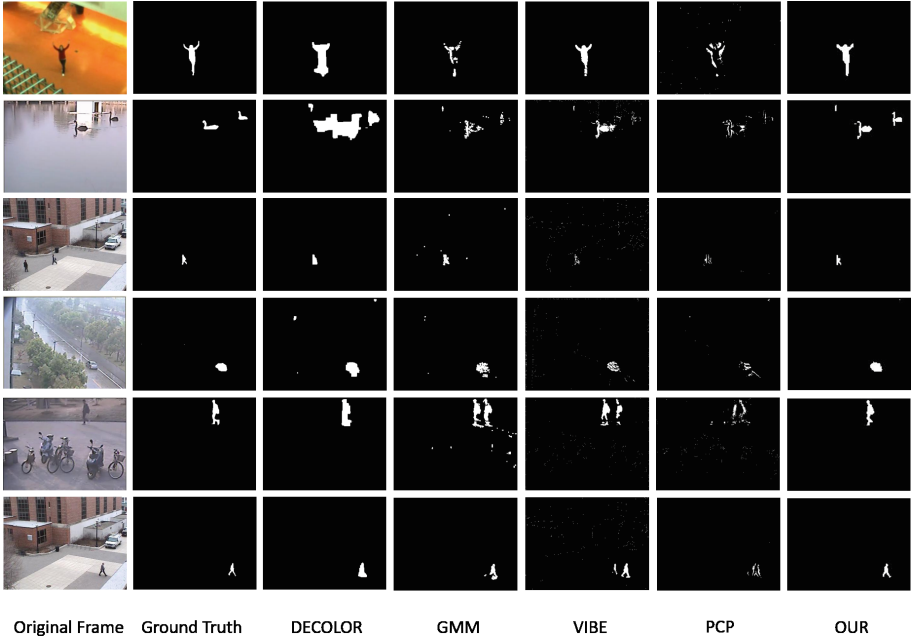
Fig. 2. Sample results of our method against the state-of-the-arts on six video sequences from GTD dataset.

Quantitative Results. Table 1 reports precision, recall, F-measure, and MAE on public GTD dataset [16]. We can see our method significantly outperforms the state-of-the-arts in precision, F-measure, and MAE. Although the recall of our method looks lower than DECOLOR [26], from Fig. 2 we can see, DECOLOR [26] tends to produce coarse boundary which always leads to high recall. The F-measure which is the comprehensive criteria between precision and recall together with the MAE verify the promising performance of our method.

Table 1. The precision, recall, F-measure and MAE values on GTD public dataset, where the bold fonts of results indicate the best performance.

Algorithm	DECOLOR	GMM	VIBE	PCP	OUR
Precision	0.54	0.51	0.40	0.29	**0.62**
Recall	**0.82**	0.64	0.47	0.18	0.79
F-measure	0.59	0.51	0.39	0.22	**0.67**
MAE	0.006	0.0125	0.0169	0.0155	**0.005**

3.3 Component Analysis

In order to validate the spatial compactness and appearance consistency via superpixel constraint, we evaluate several variations of our model and report the results on Table 2 and visualize several detection results on Fig. 3, where Ours: the proposed model; Our-I: our model without spatial compactness by setting γ to 0; Our-II: our model without appearance consistancy by setting all $w_{ij,kl}$ to 1; Our-III: our model without spatial compactness and the appearance consistency by setting all $w_{ij,kl} = 1$ and $\gamma = 0$. From Table 2 we can see that: Our-II significantly beats Our-III and Our outperforms Our-I in Recall and F-measure, which suggest that superpixel constraint plays importance role on moving object detection. From Fig. 3 we can see that: After introducing the spatial compactness and appearance consistency via superpixel constraint, our method can better preserve the boundary information and suppress the noise.

Original Frame   Ground Truth   OUR   OUR-I   OUR-II   OUR-III
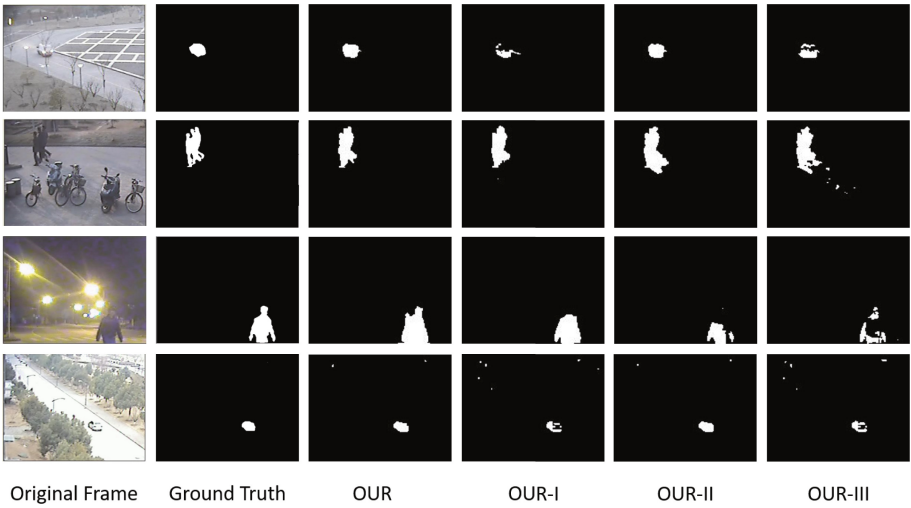
Fig. 3. Example results of our method and its variants on four video sequences from GTD dataset.

Table 2. Average precision, recall, and F-measure of our method and its variants on the GTD dataset. The bold fonts of results indicate the best performance.

Algorithm	Precision	Recall	F-measure
Ours	**0.62**	**0.79**	**0.67**
Ours-I	**0.62**	0.77	0.64
Ours-II	0.57	0.78	0.62
Ours-III	0.51	0.66	0.55

4   Conclusion

In this paper, we have proposed a novel method for moving object detection under the low-rank and sparse separation framework. We have first emphasized the neighboring pixels with close appearance. We have further explored the spatial compactness and appearance consistency between the pixels within the same superpixel. Extensive experiments against state-of-the-arts on the public video sequences suggest that, the proposed method can better preserve the boundary of the objects and robust to the noise. In future work, we will focus on extending our model to online or streaming fashion for real-life applications.

References

1. Al-Sultan, S., Al-Bayatti, A.H., Zedan, H.: Context-aware driver behavior detection system in intelligent transportation systems. IEEE Trans. Veh. Technol. **62**(9), 4264–4275 (2013)
2. Barnich, O., Van Droogenbroeck, M.: ViBe: a universal background subtraction algorithm for video sequences. IEEE Trans. Image Process. **20**(6), 1709–1724 (2011)
3. Bouwmans, T.: Background subtraction for visual surveillance: a fuzzy approach. In: Handbook on Soft Computing for Video Surveillance, pp. 103–134 (2012)
4. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Anal. Mach. Intell. **23**(11), 1222–1239 (2001)
5. Cai, J.F., Candès, E.J., Shen, Z.: A singular value thresholding algorithm for matrix completion. SIAM J. Optim. **20**(4), 1956–1982 (2010)
6. Candès, E.J., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? J. ACM (JACM) **58**(3), 1–36 (2011)
7. Chauhan, A.K., Krishan, P.: Moving object tracking using Gaussian mixture model and optical flow. Int. J. Adv. Res. Comput. Sci. Softw. Eng. **3**(4), 243–246 (2013)
8. Darbon, J.: Global optimization for first order Markov random fields with submodular priors. Discret. Appl. Math. **157**(16), 3412–3423 (2009)
9. Davis, J., Goadrich, M.: The relationship between precision-recall and ROC curves. In: International Conference on Machine Learning, pp. 233–240 (2006)
10. Dou, J., Li, J., Qin, Q., Tu, Z.: Moving object detection based on incremental learning low rank representation and spatial constraint. Neurocomputing **168**(C), 382–400 (2015)
11. Javed, S., Oh, S.H., Sobral, A., Bouwmans, T., Jung, S.K.: Background subtraction via superpixel-based online matrix decomposition with structured foreground constraints. In: IEEE International Conference on Computer Vision Workshop, pp. 930–938 (2016)
12. Jiang, B., Ding, C., Tang, J.: Graph-Laplacian PCA: closed-form solution and robustness. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3492–3498 (2013)

13. KaewTraKulPong, P., Bowden, R.: An improved adaptive background mixture model for real-time tracking with shadow detection. In: Remagnino, P., Jones, G.A., Paragios, N., Regazzoni, C.S. (eds.) Video-Based Surveillance Systems, pp. 135–144. Springer, Heidelberg (2002). https://doi.org/10.1007/978-1-4615-0913-4_11

14. Ke, Q., Kanade, T.: Robust L1 norm factorization in the presence of outliers and missing data by alternative convex programming. In: 2005 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 739–746. IEEE (2005)

15. Kolmogorov, V., Zabin, R.: What energy functions can be minimized via graph cuts? IEEE Trans. Pattern Anal. Mach. Intell. **26**(2), 147–159 (2004)

16. Li, C., Wang, X., Zhang, I., Tang, J., Wu, H., Lin, L.: WELD: Weighted low-rank decomposition for robust grayscale-thermal foreground detection. IEEE Trans. Circuits Syst. Video Tech. **1**(1), 1–14 (2016)

17. Lin, D., Fidler, S., Urtasun, R.: Holistic scene understanding for 3D object detection with RGBD cameras. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1417–1424 (2013)

18. Mazumder, R., Hastie, T., Tibshirani, R.: Spectral regularization algorithms for learning large incomplete matrices. J. Mach. Learn. Res. **11**, 2287–2322 (2010)

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

20. Shen, J., Du, Y., Wang, W., Li, X.: Lazy random walks for superpixel segmentation. IEEE Trans. Image Process. **23**(4), 1451 (2014)

21. Stauffer, C., Grimson, W.E.L.: Adaptive background mixture models for real-time tracking. In: 1999 Proceedings of the IEEE International Conference on Computer Vision, vol. 2, pp. 246–252 (1999)

22. Torre, F.D.L., Black, M.J.: A framework for robust subspace learning. Int. J. Comput. Vis. **54**(1–3), 117–142 (2003)

23. Wren, C.R., Azarbayejani, A., Darrell, T., Pentland, A.P.: Pfinder: real-time tracking of the human body. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 780–785 (1997)

24. Xin, B., Tian, Y., Wang, Y., Gao, W.: Background subtraction via generalized fused Lasso foreground modeling. In: 2015 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4676–4684 (2015)

25. Zhou, H., Kong, H., Wei, L., Creighton, D., Nahavandi, S.: Efficient road detection and tracking for unmanned aerial vehicle. IEEE Trans. Intell. Transp. Syst. **16**(1), 297–309 (2015)

26. Zhou, X., Yang, C., Yu, W.: Moving object detection by detecting contiguous outliers in the low-rank representation. IEEE Trans. Pattern Anal. Mach. Intell. **35**(3), 597–610 (2013)

27. Zhou, Z., Li, X., Wright, J., Candes, E., Ma, Y.: Stable principal component pursuit. In: 2010 IEEE International Symposium on Information Theory, pp. 1518–1522. IEEE (2010)