Multispectral Foreground Detection via Robust Cross-Modal Low-Rank Decomposition

Aihua Zheng, Yumiao Zhao, Chenglong Li(✉), Jin Tang, and Bin Luo

School of Computer Science and Technology, Anhui University, Hefei 230601, China
{ahzheng214,tj,luobin}@ahu.edu.cn, ymiaozhao@foxmail.com, lcl1314@foxmail.com

Abstract. In this paper, we propose a novel approach which pursues cross-modal low-rank decomposition for robust multi-spectral foreground detection. For each spectrum, we employ the idea of low-rank and sparse decomposition to detect sparse moving objects against background with low-rank structure for its robustness to noises. Unlike simply combining multi-modal detecting results or compulsively enforcing a shared foreground mask in existing methods, we propose to pursue the cross modality consistency among heterogeneous modalities by introducing a soft cross-modality consistent constraint to the multi-modal low-rank decomposition model. Extensive experiments on the benchmark dataset GTFD suggest that our approach achieves superior performance over the state-of-the-art algorithms.

Keywords: Cross-modality consistency · Foreground detection
Low-rank decomposition

1 Introduction

Foreground detection is a fundamental research topic in computer vision and essential in many related scenarios, such as video surveillance [25], behavior analysis [5], visual tracking [15] and object retrieval [10] et al. Despite of the great progress in the past decades, it is still a challenging task due to the complex factors, such as background clutter, illumination, bad weather, et al.

Extensive methods have been proposed for single-modality foreground detection over the past decades. The representative methods include Gaussian Mixture Models (GMM) [24], non-parameter algorithms [1], multiple features based methods [23], low-rank decomposition models [13,32] and convolutional neural network methods [4,17]. However, single visual sensor suffers from the aforementioned challenging scenarios. Recently, some literatures integrated the complementary thermal infrared sensor to effectively boost the performance in challenging scenarios. Han et al. [9] proposed a hierarchical scheme to automatically align synchronous grayscale and thermal frames, and probabilistically combined

R. Hong et al. (Eds.): PCM 2018, LNCS 11164, pp. 819–829, 2018.
https://doi.org/10.1007/978-3-030-00776-8_75

cues from registered grayscale-thermal frames for human silhouette detection. Davis et al. [6] proposed a new background-subtraction technique fusing contours from thermal and grayscale videos for object detection in urban settings. Zhao et al. [31] integrated the infrared and visible images with different strategies on salient and non-salient regions, then employed GMM to achieve the background subtraction.

Recently, some literatures focused on the multi-modal foreground detection on low-rank decomposition framework for its robustness to noises [14,28]. Li et al. [14] proposed a weighted low-rank decomposition method by learning the shared foreground mask matrix for different modalities to achieve adaptive fusion of different source data. Yang et al. [28] proposed a fast grayscale-thermal foreground detection via collaboratively separating and integrating the foregrounds from different modalities in low-rank decomposition framework. However, the hard consistency [14] with shared foreground among different modalities may be overstrict. Inspired by the fact that different images can be perceived from multi-view features [11], we argue that the different modalities are heterogeneous with different properties as shown in Fig. 1. Furthermore, the independency between modalities [28] ignored the complementary benefits from different modalities as shown in Fig. 1. Where the visible spectrum disturbed by low illumination benefits from thermal source in Fig. 1(a), and the thermal one disturbed by glass and thermal crossover benefits from visible one in Fig. 1(b). Therefore, we argue to pursuing the cross modality consistency among the heterogeneous modalities to capture these benefits.

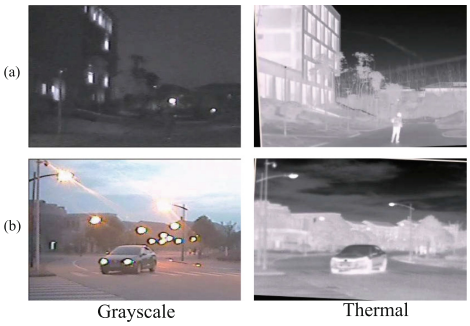

Fig. 1. Sample of the multi-modal image pairs from GTFD dataset, the grayscale and thermal modalities are heterogeneous with different properties.

Based on above discussion, we propose a novel and robust multispectral foreground detection approach to capture cross modality consistency among the heterogeneous modalities in a unified low-rank decomposition framework, we first accumulate sequential frames as two input matrices from the grayscale and thermal videos. The underlying background images are linearly correlated in each modality when ignoring the sparse and heterogeneous foregrounds and outliers. After introducing the appearance consistency and spatial compactness constraint among the neighborhood in each heterogeneous modality, we propose to

construct the cross-modal graph to pursue the cross-modality consistency among the heterogeneous foregrounds into a unified low-rank decomposition framework. Finally, we jointly optimize the proposed multi-modal low-rank decomposition to generate the heterogeneous background models and the foreground masks simultaneously.

2 Our Algorithm

Given a grayscale-thermal video pair, we solve the multi-modal foreground detection based on the low-rank decomposition framework in a batch manner.

2.1 Model Formulation

Given the k-th modal video, we accumulate n frames into a matrix by reshaping each frame into a column vector, *i.e.*, $\mathbf{D}^k = [\mathbf{d}_1^k, \mathbf{d}_2^k, ..., \mathbf{d}_n^k] \in R^{m\times n}$, with $k = 1, \cdots, K$ and m is the number of pixel on each frame. Herein, the grayscale-thermal data in this paper is the special case with $K = 2$. First, we assume that the underlying background images are linearly correlated in each modality video and the foregrounds are sparse and contiguous. This assumption has been successfully applied in background modeling [7,32].

Heterogeneous Decomposition. As we discussed above, the different modalities are heterogeneous with different properties. Therefore, we decompose the input matrices into heterogeneous foreground/background for each modality as: $\mathbf{D}^k = \mathbf{B}^k + \mathbf{S}^k$, where $\mathbf{B}^k \in R^{m\times n}$ is the low-rank background matrix, and $\mathbf{S}^k \in R^{m\times n}$ denotes the sparse heterogeneous foreground matrix of the k-th modality, which can be formulated as:

$$\begin{aligned} \min_{\mathbf{B}^k, \mathbf{S}^k} \frac{1}{2}||f_{\mathbf{S}_\perp^k}(\mathbf{D}^k - \mathbf{B}^k)||_F^2 + \beta||vec(\mathbf{S}^k)||_0, \\ s.t.\ rank(\mathbf{B}^k) \le r^k,\ k = 1, 2, ..., K, \end{aligned} \tag{1}$$

where β is a balance parameter. $vec(\cdot)$ is a vectorize operator on a matrix. $||\cdot||_F$ and $||\cdot||_0$ indicate the Frobenius norm of a matrix and the l_0 norm of a vector, respectively. r^k is a constant that suppresses the complexity of the background model in each modality. $f_{\mathbf{S}}(\mathbf{X})$ represents the orthogonal projection of a matrix $\mathbf{X}$ onto the linear space of matrices supported by $\mathbf{S}$:

$$f_{\mathbf{S}}(\mathbf{X})(i,j) = \begin{cases} 0, & \mathbf{S}_{ij} = 0, \\ \mathbf{X}_{ij}, & \mathbf{S}_{ij} = 1. \end{cases} \tag{2}$$

and $f_{\mathbf{S}_\perp}(\mathbf{X})$ is its complementary projection, *i.e.*, $f_{\mathbf{S}}(\mathbf{X}) + f_{\mathbf{S}_\perp}(\mathbf{X}) = \mathbf{X}$.

Appearance Consistency and Spatial Compactness. We observe that the neighbouring pixels have high probability with similarity appearance, which has

been successfully applied in foreground detection [26,27]. Based on this consideration, we encourage the appearance consistency by constructing adaptive weights ${w^k}_{ij,pq}$ into the spatial compactness constraint:

$$\begin{aligned} ||\mathbf{C}^k\ vec(\mathbf{S}^k)||_1 &= \sum_{(ij,kl)\in\varepsilon^k} {w^k}_{ij,pq}\ |\mathbf{S}^k_{ij} - \mathbf{S}^k_{pq}|; \\ {w^k}_{ij,pq} &= \exp\frac{-||{d^k}_{ij} - {d^k}_{pq}||_2^2}{2\theta^2}. \end{aligned} \tag{3}$$

where, $||\mathbf{X}||_1 = \sum_{ij} |\mathbf{X}_{ij}|$ denotes the l_1-norm, ε^k denotes the edge set connecting spatially neighboring pixels in the k-th modality. $\mathbf{C}^k$ is the node-edge incidence matrix denoting the connecting relationship among pixels in the k-th modality, ${d^k}_{ij}$ and ${d^k}_{pq}$ represent the intensity of pixel ij and pq in the k-th modality respectively and θ is a tunning parameter.

Cross-Modality Consistency. Different from the existing multispectral foreground detection methods that consider the information from individual modality are independent, we further propose to enforce the cross-modality consistency among the multispectral data. Meanwhile, to deal with occasional perturbation or malfunction of individual sources, we construct the cross-modality graph among the quad on one modality (thermal image) for each pixel from the other modality (grayscale image). This constraint is defined as:

$$\sum_{k=2,(ij,mn)\in\mathcal{F}}^{K} ||\mathbf{S}^k_{ij} - \mathbf{S}^{k-1}_{mn}||_F^2, \tag{4}$$

where $\mathcal{F}$ denotes edge set connecting spatially cross-modality pixels in the k-th modality (as shown in Fig. 2). Equation (4) encourages the pixel $\mathbf{S}^{k-1}_{ij}$ and its quad neighbors on the other modality $[\mathbf{S}^k_{ij}, \mathbf{S}^k_{(i+1)j}, \mathbf{S}^k_{(i+1)(j+1)}, \mathbf{S}^k_{i(j+1)}]$ belonging to the same pattern. Therefore, our model can be rewritten as:

$$\begin{aligned} &\min_{\mathbf{B}^k,\mathbf{S}^k} \frac{1}{2}||f_{\mathbf{S}^k_\perp}(\mathbf{D}^k - \mathbf{B}^k)||_F^2 + \ \beta||vec(\mathbf{S}^k)||_0 + \mu||\mathbf{C}^k\ vec(\mathbf{S}^k)||_1 \\ &+\ \gamma \sum_{k=2,(ij,mn)\in\mathcal{F}}^{K} ||\mathbf{S}^k_{ij} - \mathbf{S}^{k-1}_{mn}||_F^2, \quad s.t.\ rank(\mathbf{B}^k) \le r^k,\ k = 1, 2, ..., K, \end{aligned} \tag{5}$$

Equation (5) is a NP-hard problem, to make Eq. (5) tractable, we relax the rank operator on $\mathbf{B}^k$ with the nuclear norm, which has proven to be an effective convex surrogate of the rank operator [22]. Meanwhile, we impose the low-rank constraints on the joint background matrix that concatenates all matrices of different modalities together to optimize them collaboratively. The formulation of collaborative low-rank representation model is proposed as follows:

Algorithm 1. Optimization Procedure to Eq. (6)

Require: $\mathbf{D}^k,\ (k = 1, ..., K)$.
Set $\mathbf{B}^k = \mathbf{D}^k\ (k = 1, 2, ..., K),\ \mathbf{S}^k = \mathbf{0},\ maxIter = 20,\ \epsilon = 1e-4$.
Ensure: $\mathbf{B}^k,\ \mathbf{S}^k,\ (k = 1, 2, ..., K)$.
1: **for** $i = 1 : maxIter$ **do**
2: Parallelly update $\mathbf{B}^k$ by Eq. (7);
3: **if** $rank(\hat{\mathbf{B}}^k) \leq r^k$ **then**
4: tuning parameters λ , return to step 2.
5: **end if**
6: Update $\{\mathbf{S}^k\}$ by Eq. (9);
7: Check the convergence condition: if the maximum objective change between two consecutive iterations is less than ϵ, then terminate the loop.
8: **end for**

$$\begin{aligned}\min_{\mathbf{B},\mathbf{S}^k} \sum_{k=1}^{K} \frac{1}{2}||f_{\mathbf{S}_\perp^k}(\mathbf{D}^k - \mathbf{B}^k)||_F^2 + \ \beta||vec(\mathbf{S}^k)||_0 + \mu||\mathbf{C}^k\ vec(\mathbf{S}^k)||_1 \\ + \lambda||\mathbf{B}^k||_* + \gamma \sum_{k=2,(ij,mn)\in\mathcal{F}} ||\mathbf{S}_{ij}^k - \mathbf{S}_{mn}^{k-1}||_F^2,\end{aligned} \tag{6}$$

where γ and λ are balance parameters, $||\cdot||_*$ denotes the nuclear norm of a matrix.

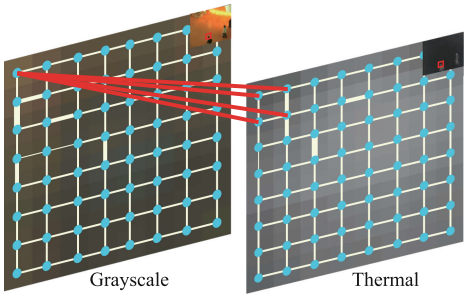

Fig. 2. The graph construction of cross-modality consistency, each pixels in grayscale image are connected to the corresponding four neighborhoods in thermal image.

2.2 Optimization

Equation (6) can be efficiently solved by the alternating optimization algorithm.

B–*subproblem*. Given an current estimate of the foreground mask $\hat{\mathbf{S}}^k$, estimating $\mathbf{B}^k$ by minimizing Eq. (6) turns to be the matrix completion problem:

$$\min_{\mathbf{B}^k} \sum_{k=1}^{K} \frac{1}{2}||f_{\hat{\mathbf{S}}_\perp^k}(\mathbf{D}^k - \mathbf{B}^k)||_F^2 + \lambda \parallel \mathbf{B}^k \parallel_*, \tag{7}$$

This is to learn a low-rank background matrix from partial observations, which can be computed via SOFT-IMPUTE [19] by iteratively using Eq. (8):

$$\hat{\mathbf{B}}^k \longleftarrow \Theta_\lambda(P_{\hat{\mathbf{S}}_\perp}(\mathbf{D}^k) + P_{\hat{\mathbf{S}}^k}(\hat{\mathbf{B}}^k)), \tag{8}$$

S–*subproblem*. Given an current estimate of the background position matrix $\hat{\mathbf{B}}^k$, Eq. (6) can be transferred into following optimization function:

$$\begin{aligned} \min_{\mathbf{S}^k} \sum_{k=1}^{K} \frac{1}{2} \, ||f_{\mathbf{S}^k_\perp}(\mathbf{D}^k - \hat{\mathbf{B}}^{\mathbf{k}})||_F^2 + \ \beta||vec(\mathbf{S}^k)||_0 + \mu||\mathbf{C}^k \ vec(\mathbf{S}^k)||_1 \\ + \gamma \sum_{k=2,(ij,mn)\in\mathcal{F}}^{K} ||\mathbf{S}^k_{ij} - \mathbf{S}^{k-1}_{mn}||_F^2 \end{aligned} \tag{9}$$

The energy function Eq. (9) can be rewritten in line with the standard form of a first-order Markov Random Fields [16] as:

$$\begin{aligned} &\min_{\mathbf{S}^k} \frac{1}{2} \sum_{k=1}^{K} \sum_{ij} (\mathbf{D}^k_{ij} - \hat{\mathbf{B}}^k_{ij})^2 (1 - \mathbf{S}^k{}_{ij}) + \beta \sum_{ij} \mathbf{S}^k_{ij} + \mu||\mathbf{C}^k \ vec(\mathbf{S}^k)||_1 \\ &\quad + \gamma \sum_{k=2,(ij,mn)\in\mathcal{F}}^{K} ||\mathbf{S}^k_{ij} - \mathbf{S}^{k-1}_{mn}||_F^2 = \min_{\mathbf{S}^k} \sum_{ij} [\beta - \frac{1}{2} \sum_{k=1}^{K} (\mathbf{D}^k_{ij} - \hat{\mathbf{B}}^k_{ij})^2] \\ &+ \mu||\mathbf{C}^k \ vec(\mathbf{S}^k)||_1 + \gamma \sum_{k=2,(ij,mn)\in\mathcal{F}}^{K} ||\mathbf{S}^k_{ij} - \mathbf{S}^{k-1}_{mn}||_F^2 + \mathscr{C} \end{aligned} \tag{10}$$

where $\mathscr{C} = \frac{1}{2} \sum_{k=1}^{K} \sum_{ij} (\mathbf{D}^k_{ij} - \hat{\mathbf{B}}^k_{ij})^2$ is a constant with respect to $\mathbf{S}^k$. The Eq. (10) can be efficiently solved by graph cut algorithm [2,12].

A sub-optimal solution can be obtained by alternating optimization to $\{\mathbf{B}^k\}$, $\{\mathbf{S}^k\}$ as summarized in Algorithm 1. The convergence of our model can be guaranteed obviously, as each sub-problem converges to a optimal solution.

3 Experiments

We evaluate our method against the state-of-the-arts on the public challenging GTFD dataset [14]. It consists of 25 video sequence pairs with grayscale and thermal modalities captured from fifteen different scenes, including laboratory rooms, campus roads, playgrounds and water pools, etc. The main challenges include intermittent motion, low illumination, bad weather, intense shadow, dynamic scene, background clutter.

3.1 Parameters

There are five parameters in our method, we adjust one parameter while fixing other parameters and then obtain better performance for our approach. The parameter β controls the sparsity of the foreground masks. We typically set $\beta = 4.5\sigma^2$, where σ is estimated online by the mean variance of $\{\mathbf{D}^k - \mathbf{B}^k\}$. The parameter μ controls the spatial smoothness to punish the neighboring pixels with different labels. The parameter γ controls the cross-modality consistency of the foreground masks to promote the pixels with same label from different modality. The parameter r constrains the complexity of the background model. The parameter θ is the tunning parameter for the appearance consistency. The final parameters are empirically set as $\{\beta, \mu, \gamma, r, \theta\} = \{4.5\sigma^2, 0.5\beta, 0.5\beta, \sqrt{n}, 10\}$, where n is the total number of pixels.

3.2 Comparison Results

We compare our approach with some state-of-the-art foreground detection algorithms, including grayscale, thermal and grayscale-thermal detection methods. Following the protocols in [14,28], we choose the detection result under grayscale scenarios as the final foreground.

Quantitative Results. Figure 3 demonstrates several detected results from GTFD dataset. From which we can see, the cross-modality consistent constraints can better preserve the foreground structures from both modalities, and achieve promising performance in both grayscale and thermal modalities even if there are misalignment among the image pairs. Furthermore, our method can produce more compact structured foregrounds.

Qualitative Results. Table 1 reports comparison results on precision, recall, F-measure together with the running speed on public GTFD dataset. We can conclude that: (1) Our method substantially outperforms other grayscale-thermal methods in precision, recall and F-measure, verifying the contribution of the proposed cross-modality consistent constraints. (2) Although WELD [14] and CLoD [28] achieve satisfying performance after fusing the grayscale and thermal results, but perform much worse in each single modality than ours. (3) The running speed of our method is lower than CLoD [28], but with much higher precision, recall and F-measure. Therefore, our method keeps a good balance between the efficiency and accuracy.

3.3 Component Analysis

To justify the component contributions of the proposed approach, we evaluate several variants of our model and report the results in Table 2, where Ours: the proposed model; Ours-I: our model without cross-modality consistency by setting γ to 0 in Eq. (6); Ours-II: our model without appearance consistency by setting adaptive weighting factor ${w^k}_{ij,kl}$ to 1 in Eq. (6); Ours-III: our model without spatial smoothness and appearance consistency by setting μ to 0.

Table 1. Average Precision (P), Recall (R), and F-measure (F) of our method against state-of-the-arts. The bold fonts of results indicate the best performance.

Algorithm	Grayscale			Thermal			Grayscale-Thermal			Code type	FPS
	P	*R*	*F*	*P*	*R*	*F*	*P*	*R*	*F*		
ASOM [18]	0.18	0.07	0.06	0.16	0.07	0.08	–	–	–	C++	111.11
FCFT [30]	0.39	0.20	0.22	0.25	0.22	0.20	–	–	–	C++	38.46
APKV [20]	0.38	0.42	0.36	0.42	0.20	0.24	–	–	–	Matlab, C++	0.03
ViBe [1]	0.41	0.49	0.41	0.41	0.47	0.39	–	–	–	C++	318.47
TTD [21]	0.59	0.29	0.32	0.58	0.38	0.40	–	–	–	Matlab	0.07
PCP [3]	0.28	0.18	0.21	0.49	0.40	0.43	–	–	–	Matlab	20.42
GMM [24]	0.48	0.65	0.52	0.48	0.65	0.50	–	–	–	C++	93.37
SAC [7]	0.42	0.74	0.41	0.47	0.71	0.53	–	–	–	Matlab	1.15
DECOLOR [32]	0.54	0.84	0.59	0.52	0.82	0.59	–	–	–	Matlab, C++	1.98
MAMR [29]	0.57	0.67	0.60	0.59	0.63	0.59	–	–	–	Matlab, C++	3.37
GMM-GT [24]	–	–	–	–	–	–	0.53	0.60	0.53	C++	34.04
JSC [8]	–	–	–	–	–	–	0.17	0.43	0.18	Matlab	10.21
WELD [14]	0.58	0.80	0.64	0.50	0.63	0.50	**0.64**	0.81	0.67	Matlab, C++	2.43
CLoD [28]	0.53	0.71	0.55	**0.63**	0.62	0.57	0.62	0.80	0.66	Matlab, C++	45.66
Ours	**0.66**	**0.86**	**0.71**	**0.65**	**0.85**	**0.70**	**0.66**	**0.86**	**0.71**	Matlab,C++	3.51

Table 2. Average Precision (P), Recall (R), and F-measure (F) of our method and its variants. The bold fonts of results indicate the best performance.

Algorithm	Grayscale			Thermal		
	P	R	F	P	R	F
Ours	**0.66**	**0.86**	**0.71**	**0.65**	**0.85**	**0.70**
Ours-I	0.59	0.73	0.60	0.60	0.70	0.61
Ours-II	0.64	**0.86**	0.70	0.63	**0.85**	0.69
Ours-III	0.62	0.85	0.68	0.61	**0.85**	0.67

The evaluation results demonstrate that: (1) Each component plays important roles in our model. (2) The cross-modality consistency contributes most by comparing Ours-I to Ours, which consequentially verifies the significance of the proposed model. Note that the higher recall in Ours-II and Ours-III results from the coarse boundary of the detected foregrounds.

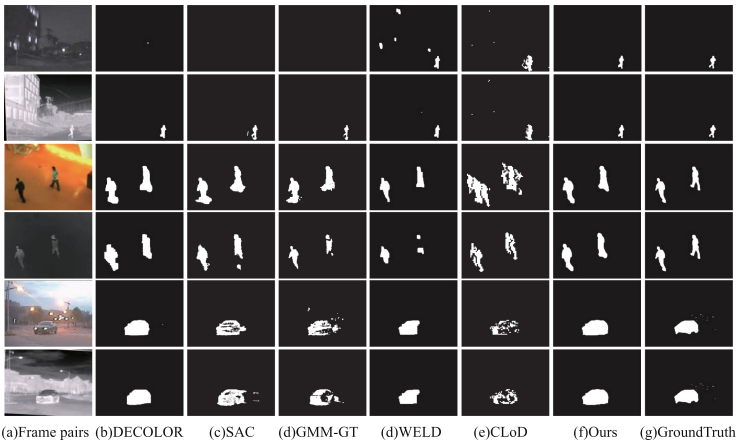

Fig. 3. Sample results of our method against other methods. The odd rows indicate the grayscale frames and the corresponding detection results generated by grayscale methods, and the even rows denote the thermal frames and the corresponding detection results generated by thermal methods.

4 Conclusion

In this paper, we have proposed novel multispectral foreground detection approach by exploring the cross-modality consistency in the low-rank and sparse decomposition framework. Extensive experiments on the GTFD dataset suggest that our approach achieved superior performance against other state-of-the-art approaches. In future work, we will develop prior models on foreground or background into our framework to further improve the robustness, and extend our algorithm into a streaming or an online fashion.

Acknowledgment. This work was partially supported by the National Natural Science Foundation of China (61502006, 61702002, 61472002 and 61671018) and the Natural Science Foundation of Anhui Higher Education Institutions of China (KJ2017A017).

References

1. Barnich, O., Droogenbroeck, M.V.: ViBe: a universal background subtraction algorithm for video sequences. IEEE Trans. Image Process. **20**, 1709–1724 (2011)
2. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Anal. Mach. Intell. **23**, 1222–1239 (2001)
3. Candes, E., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? J. ACM **58**, 175–181 (2011)
4. Chen, Y., Wang, J., Zhu, B., Tang, M., Lu, H.: Pixel-wise deep sequence learning for moving object detection. IEEE Trans. Circuits Syst. Video Technol., 1 (2017)
5. Cho, S.H., Kang, H.B.: Abnormal behavior detection using hybrid agents in crowded scenes. Pattern Recogn. Lett. **44**, 64–70 (2014)

6. Davis, J., Sharma, V.: Background-subtraction using contour-based fusion of thermal and visible imagery. Comput. Vis. Image Underst. **106**, 162–182 (2007)
7. Guo, X., Wang, X., Yang, L., Cao, X., Ma, Y.: Robust foreground detection using smoothness and arbitrariness constraints. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8695, pp. 535–550. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10584-0_35
8. Han, G., Cai, X., Wang, J.: Object detection based on combination of visible and thermal videos using a joint sample consensus background model. J. Softw. **8**, 987–994 (2013)
9. Han, J., Bhanu, B.: Fusion of color and infrared video for moving human detection. Pattern Recogn. **40**, 1771–1784 (2007)
10. Hong, R., Hu, Z., Wang, R., Wang, M., Tao, D.: Multi-view object retrieval via multi-scale topic models. IEEE Trans. Image Process. **25**, 5814–5827 (2016)
11. Hong, R., Zhang, L., Zhang, C., Zimmermann, R.: Flickr circles: aesthetic tendency discovery by multi-view regularized topic modeling. IEEE Trans. Multimedia **18**, 1555–1567 (2016)
12. Kolmogorov, V., Zabin, R.: What energy functions can be minimized via graph cuts? IEEE Trans. Pattern Anal. Mach. Intell. **2**, 147–159 (2004)
13. Li, C., Lin, L., Zuo, W., Wang, W., Tang, J.: An approach to streaming video segmentation with sub-optimal low-rank decomposition. IEEE Trans. Image Process. **25**, 1947–1960 (2016)
14. Li, C., Wang, X., Zhang, L., Tang, J., Wu, H., Lin, L.: Weighted low-rank decomposition for robust grayscale-thermal foreground detection. IEEE Trans. Circuits Syst. Video Technol. **27**, 725–738 (2017)
15. Li, C., Wu, X., Bao, Z., Tang, J.: ReGLe: spatially regularized graph learning for visual tracking. In: ACM, pp. 252–260 (2017)
16. Li, S.Z.: Markov Random Field Modeling in Image Analysis. Springer, London (2009). https://doi.org/10.1007/978-1-84800-279-1
17. Lim, L.A., Keles, H.Y.: Foreground segmentation using a triplet convolutional neural network for multiscale feature encoding. arXiv preprint arXiv:1801.02225 (2018)
18. Lucia, M., Alfredo, P.: A self-organizing approach to background subtraction for visual surveillance applications. IEEE Trans. Image Process. **17**, 1168–1177 (2008)
19. Mazumder, R., Hastie, T., Tibshirani, R.: Spectral regularization algorithms for learning large incomplete matrices. J. Mach. Learn. Res. **11**, 2287–2322 (2010)
20. Narayana, M., Hanson, A., Learned-miller, E.: Background modeling using adaptive pixelwise kernel variances in a hybrid feature space. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2012)
21. Oreifej, O., Li, X., Shah, M.: Simultaneous video stabilization and moving object detection in turbulence. IEEE Trans. Pattern Anal. Mach. Intell. **35**, 450–462 (2013)
22. Recht, B., Fazel, M., Parrilo, P.A.: Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM Rev. **52**, 471–501 (2010)
23. St-Charles, P.L., Bilodeau, G.A., Bergevin, R.: Subsense: a universal change detection method with local adaptive sensitivity. IEEE Trans. Image Process. **24**, 359–373 (2014)
24. 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, pp. 246–252 (1999)
25. Wang, X.: Intelligent multi-camera video surveillance: a review. Pattern Recogn. Lett. **34**, 3–19 (2013)

26. Xin, B., Tian, Y., Wang, Y., Gao, W.: Background subtraction via generalized fused lasso foreground modeling. arXiv preprint, pp. 4676–4684 (2015)
27. Xu, M., Li, C., Shi, H., Tang, J., Zheng, A.: Moving object detection via integrating spatial compactness and appearance consistency in the low-rank representation. In: Yang, J., et al. (eds.) CCCV 2017. CCIS, vol. 773, pp. 50–60. Springer, Singapore (2017). https://doi.org/10.1007/978-981-10-7305-2_5
28. Yang, S., Luo, B., Li, C., Wang, G., Tang, J.: Fast grayscale-thermal foreground detection with collaborative low-rank decomposition. IEEE Trans. Circuits Syst. Video Technol., 1 (2017)
29. Ye, X., Yang, J., Sun, X., Li, K., Hou, C., Wang, Y.: Foreground-background separation from video clips via motion-assisted matrix restoration. IEEE Trans. Circuits Syst. Video Technol. **25**, 1721–1734 (2015)
30. Zhang, H., Xu, D.: Fusing color and texture features for background model. In: Wang, L., Jiao, L., Shi, G., Li, X., Liu, J. (eds.) FSKD 2006. LNCS (LNAI), vol. 4223, pp. 887–893. Springer, Heidelberg (2006). https://doi.org/10.1007/11881599_110
31. Zhao, B., Li, Z., Liu, M., Cao, W., Liu, H.: Infrared and visible imagery fusion based on region saliency detection for 24-hours-surveillance systems. In: Proceeding of the IEEE International Conference on Robotics and Biomimetics (2013)
32. 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**, 597–610 (2013)