Information Sciences 369 (2016) 548–563

Contents lists available at ScienceDirect

Information Sciences

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

A novel spatio-temporal saliency approach for robust dim moving target detection from airborne infrared image sequences

Yansheng Li [a], Yongjun Zhang [a,*], Jin-Gang Yu [b], Yihua Tan [c], Jinwen Tian [c], Jiayi Ma [d]

[a] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China
[b] Department of Computer Science and Engineering, University of Nebraska-Lincoln, Lincoln, NE 68588, USA
[c] School of Automation, Huazhong University of Science and Technology, Wuhan 430074, China
[d] Electronic Information School, Wuhan University, Wuhan 430072, China

ARTICLE INFO

Article history:
Received 27 December 2015
Revised 11 July 2016
Accepted 17 July 2016
Available online 18 July 2016

Keywords:
Infrared dim moving target detection
Spatio-temporal saliency
Regularized feature reconstruction
The local adaptive contrast operation
The transmission operation

ABSTRACT

Dim moving target detection from infrared image sequences, which lags behind the visual perception ability of humans, has attracted considerable interest from researchers due to its crucial role in airborne surveillance systems. This paper proposes a novel spatio-temporal saliency model to cope with the infrared dim moving target detection problem. Based on a closed-form solution derived from regularized feature reconstruction, a local adaptive contrast operation is proposed, whereby the spatial saliency map and the temporal saliency map can be calculated on the spatial domain and the temporal domain. In order to depict the motion consistency characteristic of the moving target, this paper also proposes a transmission operation to generate the trajectory prediction map. The fused result of the spatial saliency map, the temporal saliency map, and the trajectory prediction map is called the "spatio-temporal saliency map" in this paper, from which the target of interest can be easily segmented. A diverse test dataset comprised of three infrared image sequences under different backgrounds was collected to evaluate the proposed model; and extensive experiments confirmed that the proposed spatio-temporal saliency model can achieve much better detection performance than the state-of-the-art approaches.

1. Introduction

Automatic infrared small target detection plays an important role in infrared search and track systems [5,12,13,19,45]. For military applications, it is necessary to warn of the incoming target at a very long distance [18]. As the infrared sensor is far away from the target of interest, the imaging resolution is diminished and the size of the target in the infrared image is very small. Furthermore, the infrared radiant energy of the target decays greatly after long distance propagation, which results in a very low signal-to-noise ratio (SNR) for the target [20]. Also, the target is often buried in the background clutter because of the complex surrounding environment. In this situation, the target in the image appears to be "dim," [21] which indicates that the size of the target is very small and the SNR is very low. When both the imaging resolution and the SNR of the

* Corresponding author.
E-mail address: zhangyj@whu.edu.cn (Y. Zhang).

http://dx.doi.org/10.1016/j.ins.2016.07.042
0020-0255/

target are less than desirable, the target may be perceptually invisible in one single image. Generally, successful local feature descriptors [1,8,9] in the natural image processing domain are good at encoding texture and shape features, but they are not competent for infrared small target detection as infrared images lack the available structure features and the projected infrared small target does not contain the available shapes. In addition, though some super-resolution methods have also been proposed to deal with the very low-resolution problem [17], they mainly aim at obtaining visually appealing results and are unsuitable for subsequent detection task. Even under these extreme circumstances, the target still appears to be salient in the image sequences. Hence, the temporal context is the key to improving the dim target detection performance. However, extraction of the temporal information from infrared sequences is not easy as common feature descriptors [27,25] and image matching approaches [28,29] are not often used directly in motion compensation of infrared sequences. Compared to moving target detection from natural images [38,43], the crucial difficulties of dim moving target detection from infrared images are embodied in the noisy background and the dim target. Although a number of studies [3,7,23,39] addressed dim moving target detection recently, how to robustly extract a dim moving target remains a challenging problem.

Numerous infrared small moving target detection approaches were proposed in the past few decades, which can be classified according to their input into two categories based on whether or not the approach utilized the temporal information: 1) methods using one single image and 2) methods using multiple adjacent images. In the first category, only one single image is available for implementing infrared small target detection. Generally, an infrared small target presents itself as a Gaussian light blob in the image, which is the crucial distinction between the infrared small target and the background clutter. The approaches based on this discrimination rule include the mathematical morphological-based methods [2,30,40], the LS-SVM-based approach [44], the facet-based methods [32,41,48], the image layering-based method [26], the local contrast based method [5], the low rank recovery-based method [12], and the visual attention-based approaches [15,32,33]. Generally, these methods can efficiently extract the target when the received image is substantially ideal. However, these methods do not work well when the SNR of the target is very low and the complexity of the background clutter is very high. In this situation, it is necessary to adopt temporal cues. In [3,10], target detection and tracking were jointly considered; and [18] integrated a temporal detector, which noticeably improved the target detection rate with an acceptable false alarm rate. Deng and Zhu [7] proposed a local increment coding approach to suppress complex background clutter. Chen et al. [4] utilized the bi-dimensional empirical model to address the dim moving target detection problem. Sun et al. [39] utilized the patches from the previous or after images to reconstruct the current image, and the residual image between the current image and the reconstructed image also was utilized to highlight the dim moving target. Based on the background and target dictionary, the likelihood that each patch belonged to the target could be solved by the sparse representation on the spatio-temporal domain [22,39]. Inspired by the visual attention process of humans, Li et al. [23] proposed a hierarchical approach to extract moving targets in which motion cues were utilized to generate candidate regions. In summary, although progress has been achieved in dim moving target detection, more exploitation of motion analysis is needed because adequate utilization of the temporal information is the key to further improving the moving target detection performance.

It is well known that the existing computational models related to dim moving target detection still lag far behind the visual perceptual ability of humans. The natural next step is to further pursue the imitation of the biological visual perception process to design infrared dim moving target detection algorithms. In the visual cortex, visual information is processed and organized along two different streams: 1) the ventral stream for appearance perception and 2) the dorsal stream for motion perception [11]. In the literature, the two-stream hypothesis has been widely employed in many computer vision tasks such as action recognition [37], scene recognition [36], and spatio-temporal saliency modeling [24]. Similar to these computer vision tasks [24,36,37], object detection can be solved by spatial saliency modeling, which aims to imitate the perception function of the primary areas of the ventral stream [16]. Moving object detection can be improved by spatio-temporal saliency modeling, which aims to imitate the joint perception function of the ventral stream and the dorsal stream [24,35]. Several existing studies [15,32,33] made use of the capabilities of spatial saliency models to cope with the infrared small target detection problem. In order to comprehensively implement the motion analysis using image sequences, it is logical to design computational models in a spatio-temporal manner [14,49]. Accordingly, infrared small moving target detection can be similarly solved through spatio-temporal saliency modeling. However, the existing spatio-temporal saliency models [24,35] were established for high-resolution natural images and cannot be directly utilized in the infrared small moving target detection task as infrared images generally lack rich structure features. To the best of our knowledge, there are currently no spatio-temporal saliency models specifically designed for infrared dim moving target detection.

Generally, one infrared moving small target reflects three characteristics: spatial singularity, temporal singularity, and motion consistency. The singularity characteristic can be interpreted as the isolation property, and the motion consistency characteristic means that one moving target would consistently appear in the field of view for some time. More specifically, the spatial singularity characteristic reflects the isolation property in the spatial domain (i.e., in the current image). The temporal singularity characteristic depicts the isolation property in the temporal domain (i.e., in two adjacent images). The motion consistency characteristic reveals the history information of the infrared moving target, which is reflected in multiple history images.

In order to narrow the gap between the unstable detection performance from computational methods and the perfect interpretation performance of humans, this paper proposes a novel spatio-temporal saliency model. More specifically, based on a closed-form solution derived from regularized feature reconstruction, we developed a local adaptive contrast operation whereby the spatial saliency map and the temporal saliency map can be calculated on the spatial domain and the temporal domain, respectively. In order to depict the motion consistency characteristic of the moving target, this paper also proposes

a transmission operation to generate the trajectory prediction map. The motion continuity constraint allows manipulation of the proposed transmission in the local domain, which makes the calculation process efficient. In addition, the proposed transmission operation works in a recursive way, which is well suited to mine the history information. The spatial saliency map, the temporal saliency map, and the trajectory prediction map, which reflect the singularity characteristic in the spatial domain, the singularity characteristic in the temporal domain, and the motion consistency characteristic, respectively, then are fused and become the "spatio-temporal saliency map," from which the target can be easily segmented. Hence, the proposed spatio-temporal saliency model can specifically depict the spatial singularity characteristic, the temporal singularity characteristic, and the motion consistency characteristic of an infrared dim moving target.

A representative test set comprised of three infrared sequences taken under different backgrounds (i.e., sky-sea, ground, and sky) were used to confirm the validity of the proposed approach. When compared with the state-of-the-art approaches, the proposed approach showed very promising results. As a whole, the contributions of this paper can be summarized as follows:

- Based on regularized feature reconstruction, the proposed approach utilizes a closed-form local adaptive contrast operation which performed better background clutter suppression compared to the existing local contrast operations.
- The proposed approach offers a simple but efficient transmission operation to catch the motion consistency characteristic of the moving target; and based on the existing literature, this is the first time that the motion consistency characteristic has been mined and utilized in infrared dim moving target detection.
- The proposed novel spatio-temporal saliency model potentially has other applications, such as spatio-temporal interest point detection [46] and action recognition [34].

The remainder of this paper is organized as follows. Section 2 introduces the proposed local adaptive contrast operation. Section 3 introduces the proposed infrared dim moving target detection approach through spatio-temporal saliency modeling. Section 4 presents the experimental results, which include a comprehensive analysis of the proposed infrared dim moving target detection approach and a comparison of the proposed approach to the existing state-of-the-art approaches. Finally, Section 5 presents the conclusions of this paper.

2. The proposed local adaptive contrast operation

Before the complete spatio-temporal saliency method is introduced in Section 3, this section first discusses the saliency measure since it plays a key role in saliency modeling. More specifically, the traditional local contrast measure (LCM), which has been proposed and utilized in infrared small target enhancement in the past, is discussed. Due to the drawbacks of LCM, we then propose the feature distinction-based LCM and the feature reconstruction-based LCM.

2.1. The classical local contrast operation

Due to its biological plausibility, the contrast measure has been widely utilized to measure saliency [6,24,31]. The LCM, in particular, has been successfully utilized in small target enhancement [5] because the Gaussian shape and light of the infrared small target mainly reflect in a local context. As depicted in [5], LCM is calculated based on the statistical difference between the central patch and its neighboring patches. Let P^c denote the central patch and its neighboring patches by $\{P_i^s | i = 1, 2, \cdots, n\}$ where n is the number of neighboring patches. Additionally, V^c is the vector, which is the vectorization result of P^c, and $\{V_i^s | i = 1, 2, \cdots, n\}$ denotes the vectorization results of $\{P_i^s | i = 1, 2, \cdots, n\}$; more specifically, LCM can be expressed by:

$$\mathrm{LCM} = \min_i L^c \times \frac{L^c}{L_i^s} i{=}1,2,\cdots,n \tag{1}$$

where L^c is the maximum value of the central vector V^c, and L_i^s is the mean value of the neighbor vector V_i^s.

From the definition of LCM in Eq. (1), it can be seen that LCM is able to highlight the target with a high level of brightness as well as large local contrast. It is well known that pixel-size electronic noises (PSEN) with high brightness usually exist in the infrared images and the noises [42]. Hence, LCM may introduce a high false alarm rate because PSEN also can generate a large response [15]. LCM also suffers the challenge of complex background clutter (e.g., clouds) because utilizing only the statistical value of the patch in the calculation of LCM does not make full use of the patch feature.

2.2. The local contrast operation based on feature distinction

In order to take full advantage of the feature, the local contrast operation based on feature distinction (LCFD) was taken as the saliency measure to indicate a small target. Similar to the contrast measure defined in [6,24,31], based on the aforementioned central vector V^c and neighboring vectors $\{V_i^s | i = 1, 2, \cdots, n\}$ utilized in the LCM calculation, LCFD can be specifically calculated by:

$$\mathrm{LCFD} = \sum_{i=1}^{n} \left\| V^c - V_i^s \right\|^2 \tag{2}$$

Compared to LCM, LCFD can successfully cope with PSEN in infrared image sequences. As defined in Eq. (1), LCM is calculated based on the statistical difference between the central patch and its neighboring patches. LCM can be easily affected by PSEN because the statistical value of one patch would be changed only if one pixel electronic noise appears in the patch. However, LCFD is the summation of the Euclidean distances between the feature vector of the central patch and the feature vectors of the neighboring patches. Although PSEN exists in one patch, only one element of the feature vector with a few dozen dimensions is changed. Hence, the influence of PSEN is suppressed in LCFD.

However, LCFD outputs a large response in a complex background, such as textured clouds in a sky background or structured roads in a ground background because it depicts only the sum of the differences between the central raw feature vector and the neighboring raw feature vectors with the same weight. Due to this drawback, LCFD therefore cannot effectively suppress the complex backgrounds.

2.3. The local adaptive contrast operation based on feature reconstruction

Considering the above weakness of LCFD, the local adaptive contrast operation was explored. More specifically, the linear combination of the neighboring feature vectors was utilized to reconstruct the central feature vector, and the Euclidean distance between the reconstructed feature vector and the central feature vector was taken as the local contrast measure.

Assuming that the central feature vector $V^c \in R^{m\times 1}$ is a column vector where m is the number of pixels in the aforementioned patch P^c, $V=[V_1^s, V_2^s, \cdots, V_n^s] \in R^{m\times n}$ is the union of the neighboring feature vectors, and $\boldsymbol{w} \in R^{n\times 1}$ denotes the combination coefficients, the regularized feature reconstruction is formalized by:

$$\min_{w} \| V^c - V \cdot \boldsymbol{w} \|^2 + \lambda \cdot \| \boldsymbol{w} \|^2 \tag{3}$$

where λ is the regularization term that controls the trade-off between the bias and the variance of the fitting model, which is the combination coefficients $\boldsymbol{w}$ in this circumstance.

Through optimizing the objective function defined in Eq. (3), a closed-form solution of the linear combination coefficients $\boldsymbol{w}$ was obtained:

$$\boldsymbol{w}=\left(V^T \cdot V + \lambda \cdot I\right)^{-1} \cdot V^T \cdot V^c \tag{4}$$

where $I \in R^{n\times n}$ is the identity matrix.

Based on the union matrix V and the linear combination coefficients $\boldsymbol{w}$, the reconstruct feature vector V^r of the central feature vector V^c can be depicted by:

$$\begin{aligned} V^r &= V \cdot \boldsymbol{w} \\ &= V \cdot (V^T \cdot V + \lambda \cdot I)^{-1} \cdot V^T \cdot V^c \end{aligned} \tag{5}$$

Furthermore, the local adaptive contrast operation based on regularized feature reconstruction (LACRFR) can be expressed by:

$$\begin{aligned} \text{LACRFR} &= \| V^c - V^r \|^2 \\ &= \left\| V^c - V \cdot (V^T \cdot V + \lambda \cdot I)^{-1} \cdot V^T \cdot V^c \right\|^2 \end{aligned} \tag{6}$$

LACRFR is called the local adaptive contrast operation because it can be rewritten as Eq. (7) depending on the linear combination coefficients $\boldsymbol{w} = [\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_n]'$, which can be adaptively learned from the regularized feature reconstruction function in Eq. (3).

$$\text{LACRFR} = \| V^c - V \cdot \boldsymbol{w} \|^2 \tag{7}$$

As depicted in Eq. (6), LACRFR is calculated by the Euclidean distance between the vectorization feature vector of the central patch and the reconstructed feature vector, which is the reconstruction result of the feature vector of the central patch using the vectorization feature vectors of neighboring patches. As long as one neighboring patch is similar to the central patch, the reconstruction feature would be similar to the vectorization feature vector of the central patch. Coincidentally, infrared small target reflects an obvious isolation property. Hence, LACRFR is well suited for indexing an infrared small target. When the central patch is located in the plain background region, the response value of LACRFR is very small as the reconstruction feature vector is similar to the feature vector of the central patch. This result demonstrates that LACRFR improves the background suppression performance. When the central patch is located in the line structures, the reconstruction feature vector is also similar to the feature vector of the central patch because there are at least two neighboring patches similar to the central patch. In this situation, the response value of LACRFR is also very small, which indicates that LACRFR can suppress the structure clutter. When the central patch is located at the dim target, all the neighboring patches are different from the central patch. In this situation, the reconstruction feature vector is different from the feature vector of the central patch and the response value of LACRFR is large, which demonstrates that LACRFR can enhance the infrared dim target. Hence, LACRFR can robustly suppress complex backgrounds and enhance a small target.

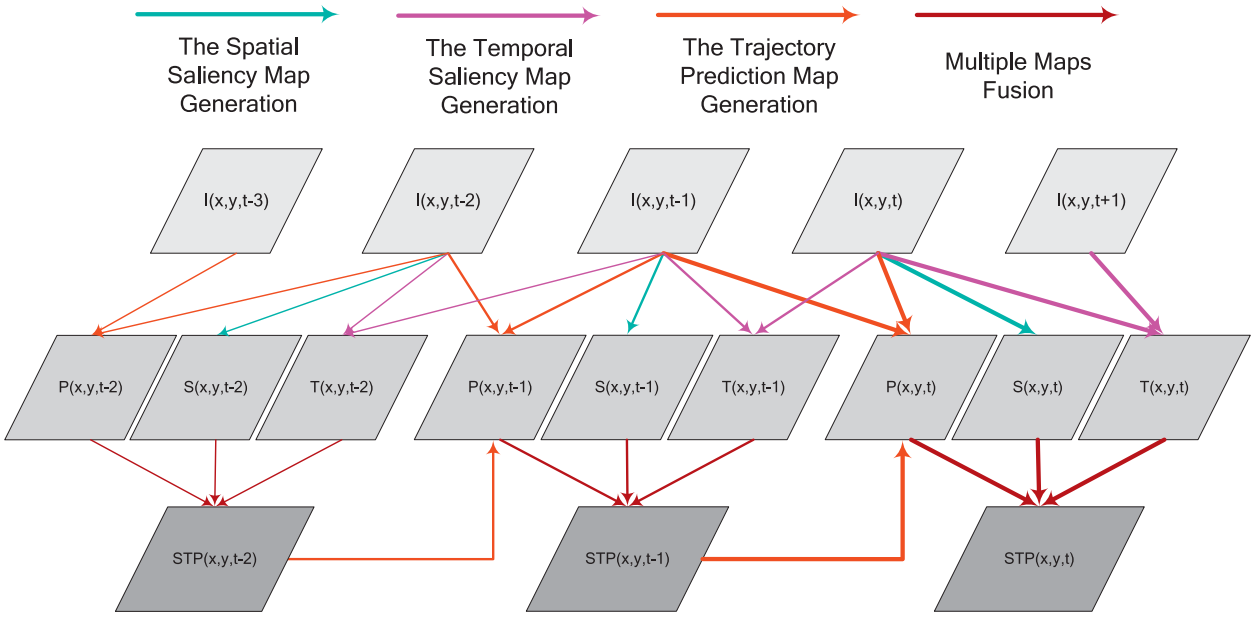

Fig. 1. The work flow of the proposed spatio-temporal saliency approach. In Fig. 1, (x, y) denotes the image coordinate and t denotes the time coordinate.

The qualitative and quantitative performance comparison of LCM, LCFD, and LACRFR is discussed in Section 4 as well as the influence of the regularization term λ of LACRFR on the final performance.

3. The proposed infrared dim moving target detection approach

In this section, the overall architecture of the proposed spatio-temporal saliency approach is introduced. Then, the spatial saliency map generation process, the temporal saliency map generation process, the trajectory prediction map generation process, and the fusion process for generating the final spatio-temporal saliency map are presented. Concluding this section, the infrared dim moving target detection approach, which is based on the proposed spatio-temporal saliency approach, is introduced.

3.1. The overall architecture of the proposed spatio-temporal saliency approach

Fig. 1 illustrates the work flow of the proposed spatio-temporal saliency approach. In Fig. 1, $I(x, y, t)$, $S(x, y, t)$, $T(x, y, t)$, $P(x, y, t)$, and $STP(x, y, t)$ represent the original image, the spatial saliency map, the temporal saliency map, the trajectory prediction map, and the spatio-temporal saliency map, respectively. The corresponding results of the intermediate steps of the proposed approach are illustrated in Fig. 2.

As shown in Fig. 1, the proposed approach works in a recursive style. The spatial saliency map $S(x, y, t)$ can be computed using just the current image $I(x, y, t)$ while the calculation of the temporal saliency map $T(x, y, t)$ needs to utilize the current image $I(x, y, t)$ and the backward image $I(x, y, t+1)$. With the aid of the history spatio-temporal saliency map $STP(x, y, t-1)$, the current trajectory prediction map $P(x, y, t)$ can be calculated using patches matching between the current image $I(x, y, t)$ and the forward image $I(x, y, t-1)$. Corresponding to the current image $I(x, y, t)$, its spatio-temporal saliency map $STP(x, y, t)$ can be calculated by fusing the spatial saliency map, the temporal saliency map, and the trajectory prediction map. Fig. 2 illustrates the proposed approach's ability to exploit the stability of the recursive process. At the start of the test sequence, the trajectory prediction map is initialized with the same value as depicted in $P(x, y, 2)$ in Fig. 2. After a few rounds, the trajectory prediction map $P(x, y, t)$ can produce a perfect prediction of the target. Benefiting from $P(x, y, t)$, the spatio-temporal saliency map $STP(x, y, t)$ can effectively indicate the dim moving target in the current image $I(x, y, t)$.

In the following paragraphs, the crucial steps of the proposed approach include the spatial saliency map generation, the temporal saliency map generation, the trajectory prediction map generation, and the fusion of multiple maps, are discussed in detail. The proposed spatio-temporal saliency model takes the patches as the primary units, therefore, a visual illustration of the division of the original images is provided. As depicted in Fig. 3, one given image can be represented by the small patches with the size $\varsigma \times \varsigma$ and the overlap length between two adjacent patches is ℓ.

It is noted that LACRFR was taken as the saliency measure to generate the saliency maps. LCM and LCFD also can be similarly utilized to output saliency maps.

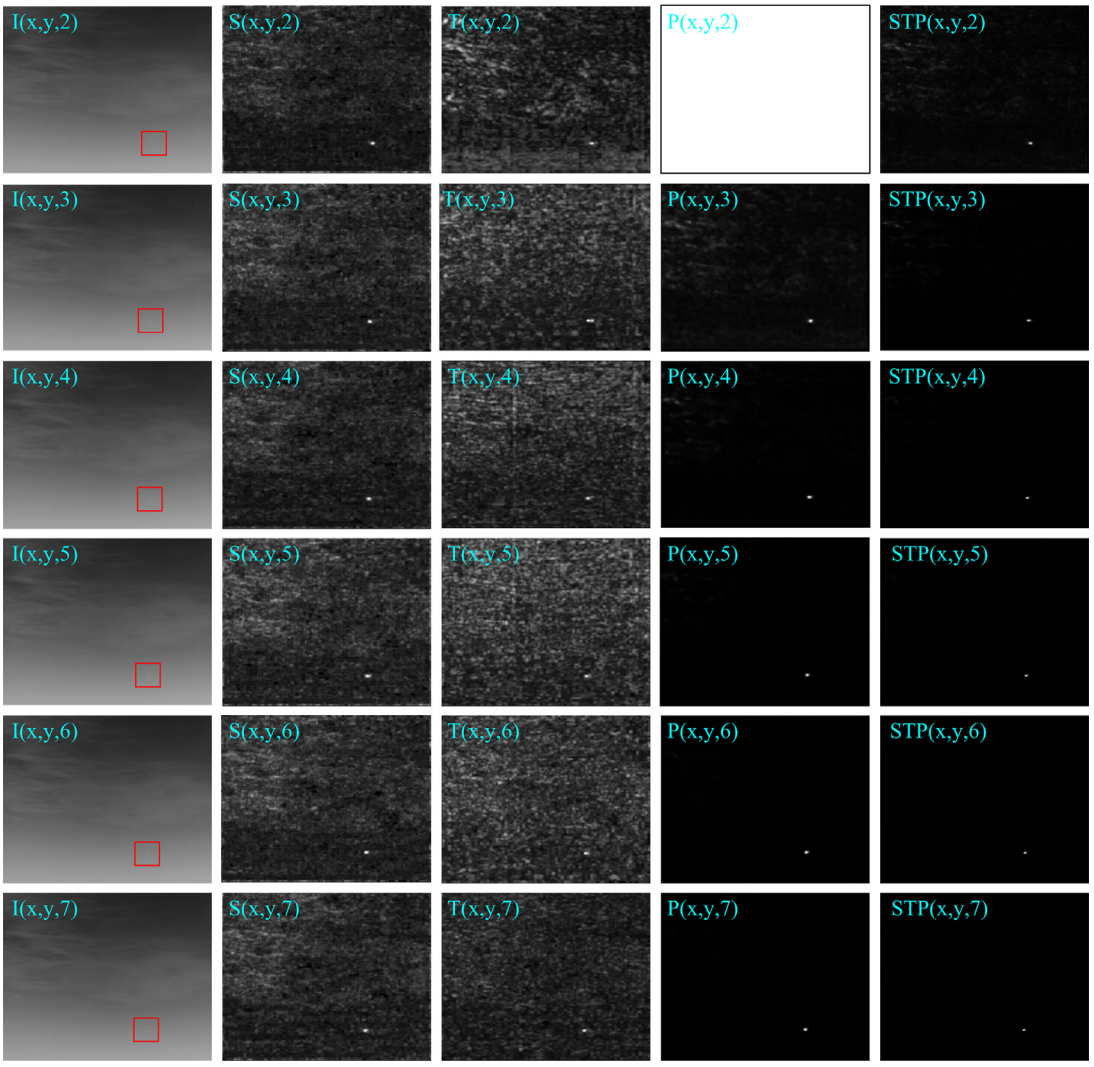

Fig. 2. The visual results of the intermediate steps of the proposed spatio-temporal saliency approach.

3.2. The spatial saliency map generation process

The spatial saliency map aims to mine the spatial singularity characteristic. The calculation of the spatial saliency map $S(x, y, t)$ depends only on the current image $I(x, y, t)$. Based on the saliency measure (i.e., the local contrast operation) introduced in Section 2, the elaborate calculation process follows.

As depicted in Fig. 3, $I(x, y, t)$ is first divided into multiple small patches, which are the primary units in the calculation of the saliency map. As illustrated in Fig. 4, the red rectangle denotes the current sliding patch and the yellow rectangle denotes its local domain. Taking the current sliding patch as the central patch and its neighboring patches shown in $I(x, y, t)$ of Fig. 4, the corresponding central vector can be expressed by $V^{t,c}$ and its neighboring vectors can be expressed by $\{V_i^{t,s} | i = 1, 2, \cdots, n\}$ as illustrated in Fig. 4. Hence, the saliency intensity of the current sliding patch is indicated by the contrast value between the central vector $V^{t,c}$ and its neighboring vectors $\{V_i^{t,s} | i = 1, 2, \cdots, n\}$. Utilizing the proposed LACRFR, the spatial saliency $S(x, y, t)$ corresponding to the current sliding window whose coordinate range is $[h_\alpha : h_\alpha + \varsigma - 1, w_\alpha : w_\alpha + \varsigma - 1]$ can be expressed by:

$$S(i, j, t) = \left\| V^{t,c} - V^{t,s} \cdot \left((V^{t,s})^T \cdot V^{t,s} + \lambda \cdot I \right)^{-1} \cdot (V^{t,s})^T \cdot V^{t,c} \right\|^2 \tag{8}$$

where $i \in [h_\alpha, h_\alpha + \varsigma - 1]$, $j \in [w_\alpha, w_\alpha + \varsigma - 1]$, and $V^{t,s} = [V_1^{t,s}, V_2^{t,s}, \cdots, V_n^{t,s}]$ denote the spatial neighboring vectors of the current sliding patch.

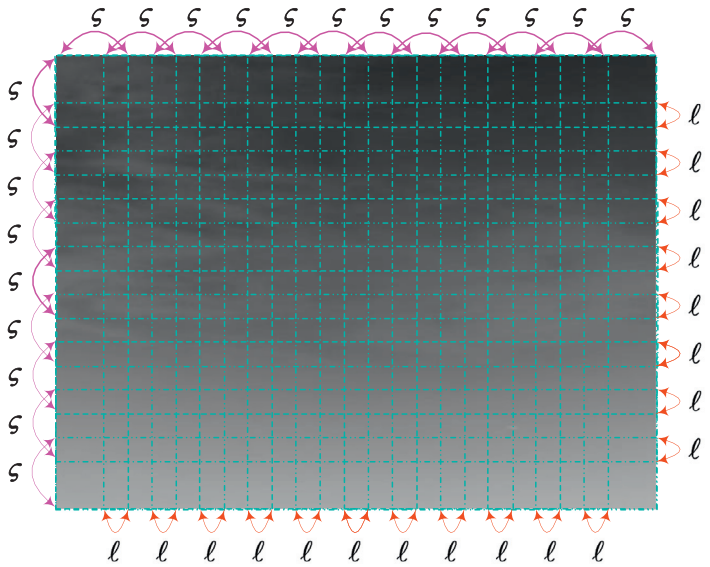

Fig. 3. Illustration of the generation of the primary units.

As previously mentioned, the divided patches are overlapped. Hence, the spatial saliency value $S(x, y, t)$ in the overlapped region is the average of the saliency responses of the overlapped patches.

3.3. The temporal saliency map generation process

The temporal saliency map is appropriate for mining the temporal singularity characteristic; and the calculation of the temporal saliency map $T(x, y, t)$ depends on the current image $I(x, y, t)$ and the backward image $I(x, y, t+1)$, which is discussed below.

The temporal singularity corresponds to the local contrast between the local patch in the current image $I(x, y, t)$ and its temporal neighboring patches in the backward image $I(x, y, t+1)$. As depicted in Fig. 4, the red rectangle in $I(x, y, t)$ denotes the local patch in $I(x, y, t)$, and all of the patches in the yellow rectangle in $I(x, y, t+1)$ represent its temporal neighboring patches in $I(x, y, t+1)$. Based on vector $V^{t,c}$ of the current sliding window in $I(x, y, t)$ and its temporal neighboring vectors $\{V_i^{t+1,c+s} | i = 1, 2, \cdots, n+1\}$, the temporal saliency $T(x, y, t)$ corresponding to the current sliding patch whose coordinate range is $[h_\alpha : h_\alpha + \varsigma - 1, w_\alpha : w_\alpha + \varsigma - 1]$ can be expressed by:

$$T(i, j, t) = \left\| V^{t,c} - V^{t+1,c+s} \cdot \left((V^{t+1,c+s})^T \cdot V^{t+1,c+s} + \lambda \cdot I \right)^{-1} \cdot (V^{t+1,c+s})^T \cdot V^{t,c} \right\|^2 \tag{9}$$

where $i \in [h_\alpha, h_\alpha + \varsigma - 1]$, $j \in [w_\alpha, w_\alpha + \varsigma - 1]$, and $V^{t+1,c+s} = [V_1^{t+1,c+s}, V_2^{t+1,c+s}, \cdots, V_{n+1}^{t+1,c+s}]$ denotes the temporal neighboring vectors of the current sliding vector.

Similar to the calculation process of the spatial saliency map, the temporal saliency value $T(x, y, t)$ in the overlapped region is also the average of the saliency responses of the overlapped patches.

3.4. The trajectory prediction map generation process

The trajectory prediction map mainly takes charge of exploiting the motion consistency. More specifically, the calculation of the trajectory prediction map $P(x, y, t)$ depends on the current image $I(x, y, t)$, the forward image $I(x, y, t-1)$, and the previous spatio-temporal saliency map $STP(x, y, t-1)$.

Here, the transmission operation for searching the matching patch in $I(x, y, t-1)$ of the current sliding patch in $I(x, y, t)$ is first introduced. The current sliding patch in $I(x, y, t)$ is defined as before, and all of the patches in the yellow rectangle in $I(x, y, t-1)$ denote the candidate set for matching as depicted in Fig 4. As illustrated in Fig. 4, the feature vector of the current sliding patch is $V^{t,c}$ and the feature vectors of the candidate set are $\{V_i^{t-1,c+s} | i = 1, 2, \cdots, n+1\}$. By calculating the Euclidean distances between $V^{t,c}$ and $\{V_i^{t-1,c+s} | i = 1, 2, \cdots, n+1\}$ and selecting the smallest distance, the best matching patch in $I(x, y, t-1)$ for the current sliding patch in $I(x, y, t)$ can be determined. Assuming that the coordinate range of the matching patch in the local domain of $I(x, y, t-1)$ is $[h_\beta : h_\beta + \varsigma - 1, w_\beta : w_\beta + \varsigma - 1]$, the trajectory prediction map $P(x, y, t)$ corresponding to the current sliding patch whose coordinate range is $[h_\alpha : h_\alpha + \varsigma - 1, w_\alpha : w_\alpha + \varsigma - 1]$ can be

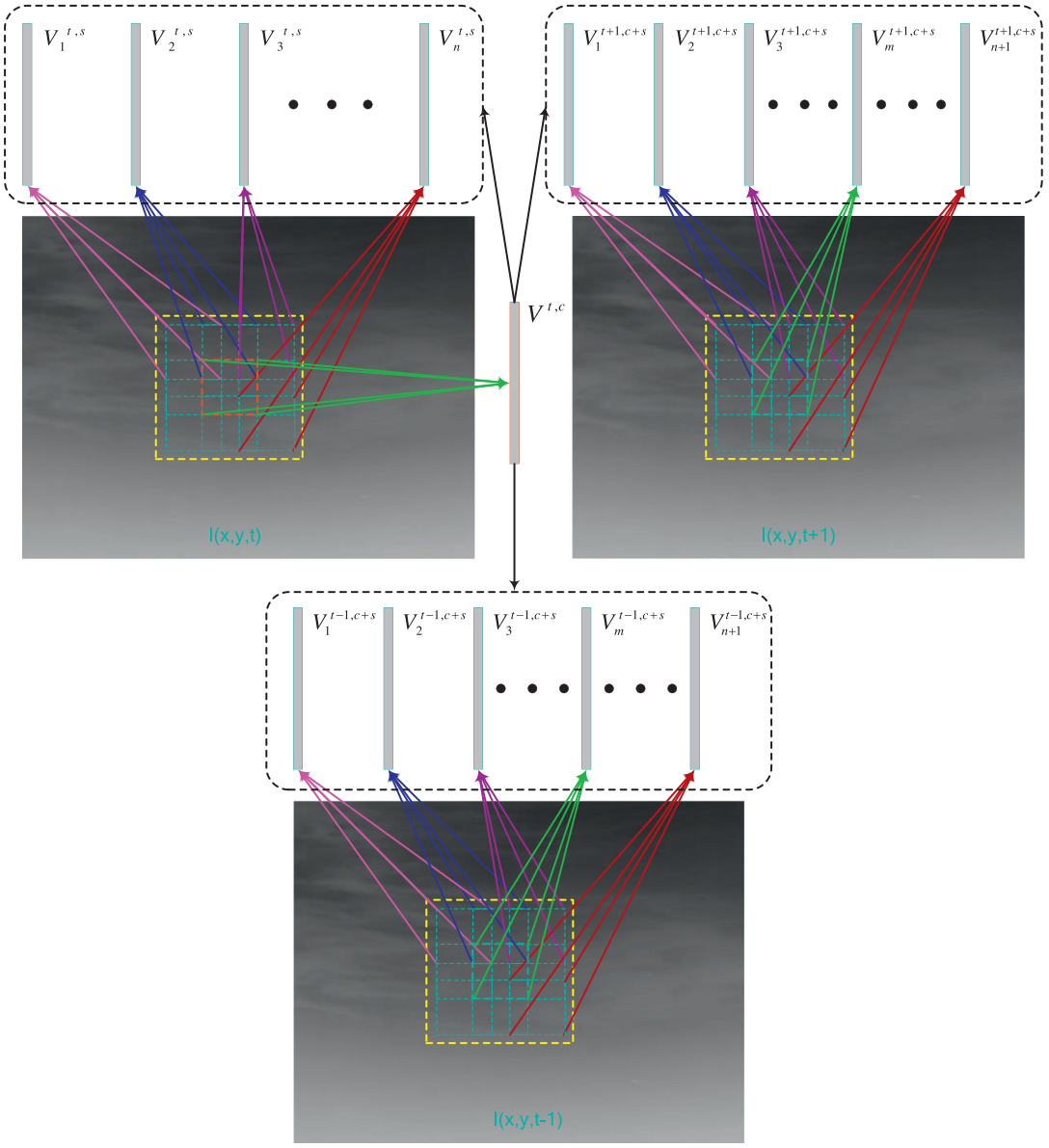

Fig. 4. Illustration of the calculation of the spatial saliency map, the temporal saliency map, and the trajectory prediction map. The central patch with the red rectangle annotating the local domain with the yellow rectangle labeled denotes the current sliding patch in $I(x, y, t)$. In $I(x, y, t)$, the patches excluding the central patch in the yellow rectangle stand for the spatial neighboring patches of the central patch in $I(x, y, t)$. In $I(x, y, t+1)$, and all of the patches in the yellow rectangle are the temporal neighboring patches of the central patch in $I(x, y, t)$. All of the patches in the yellow rectangle in $I(x, y, t-1)$ comprise the correspondence set for determining the best matching patch.

expressed by:

$$P(i, j, t) = \sum_{x=h_\beta}^{h_\beta+\varsigma-1} \sum_{y=w_\beta}^{w_\beta+\varsigma-1} STP(x, y, t-1) \tag{10}$$

where $i \in [h_\alpha, h_\alpha + \varsigma - 1]$, $j \in [w_\alpha, w_\alpha + \varsigma - 1]$, $STP(x, y, t-1)$ is the spatio-temporal saliency map at the previous moment. Its calculation process is presented in Section 3.5.

Similar to the calculation process of the spatial saliency map and temporal saliency map, the trajectory prediction value $P(x, y, t)$ in the overlapped region is also the average of the prediction values of the overlapped patches.

The process defined in Eq. (10) is the aforementioned transmission operation that can simply and effectively depict the motion consistency characteristic which plays a key role in infrared dim moving target detection. This supposition is verified in the experimental section.

Algorithm 1. The proposed dim moving target detection approach.

Input: The original image sequence $I(x, y, t) t = 1, 2, \cdots, n$.
Output: The target detection result $R(x, y, t) t = 2, 3, \cdots, n-1$.
Initialization: $STP(x, y, 1) = 1.0$.
1: **for** t=2 : n-1
2: Compute the spatial saliency map $S(x, y, t)$ through Eq. (8). In addition, the calculation process in Eq. (8) only depends on $I(x, y, t)$.
3: Calculate the temporal saliency map $T(x, y, t)$ through Eq. (9), which takes $I(x, y, t)$ and $I(x, y, t+1)$ as the input.
4: Calculate the trajectory prediction map $P(x, y, t)$ through Eq. (10). The calculation process defined in Eq. (10) depends on $I(x, y, t)$, $I(x, y, t-1)$, and $STP(x, y, t-1)$.
5: Calculate the spatio-temporal saliency map $STP(x, y, t)$ through both Eq. (11) and Eq. (12), which take $S(x, y, t)$, $T(x, y, t)$, and $P(x, y, t)$ as the input.
6: Segment the spatio-temporal saliency map $STP(x, y, t)$ based on the adaptive threshold defined by Eq. (13), and the final binary detection results are shown in $R(x, y, t)$.
7: **end for**

3.5. The spatio-temporal saliency fusion process

Previous Sections B, C, and D discussed in detail the calculation process of the spatial saliency map, the temporal saliency map, and the trajectory prediction map, respectively. As previously mentioned, $S(i, j, t)$ and $T(i, j, t)$ depict the spatial singularity characteristic and the temporal singularity characteristic that belong to the same singularity characteristic. In addition, $P(i, j, t)$ depicts the motion consistency characteristic. According to the differences between the types of characteristics, the singularity characteristic is first combined and then particularly combined with the motion consistency characteristic.

There are three strategies available for the combining process of the spatial saliency map and the temporal saliency map: the addition fusion strategy, the multiplication fusion strategy, and the maximum fusion strategy. The multiplicative fusion strategy exhibited superior performance compared to the two other strategies in outputting a more balanced fusion result [24,35]. Therefore, the spatial saliency map and the temporal saliency map are first combined by the multiplication fusion strategy:

$$ST(i, j, t) = N(S(i, j, t)) \cdot N(T(i, j, t)) \tag{11}$$

where $N(\cdot)$ denotes the normalization operation and $ST(i, j, t)$ reflects the singularity characteristic in the spatio-temporal domain.

In order to utilize the constraint from the motion consistency characteristic, $ST(i, j, t)$ is further combined with $P(i, j, t)$. Normally, the multiplication fusion strategy is implemented without a constant. The strategy defined in Eq. (11) coincides with the traditional multiplication fusion operator [24,35]. In most cases, the traditional multiplication fusion strategy can output a balanced fusion result. However, when one map contains 0, the multiplication fusion strategy presents as tendentious and the fusion result also will be 0. This issue is especially fatal when one approach is designed in a recursive manner. The sequential results will be incorrect and are not recoverable once the error occurs at one time. In order to prevent 0 from appearing in the intermediate results, $ST(i, j, t)$ is combined with $P(i, j, t)$ using the multiplication fusion strategy with a biased constant in Eq. (12). More specifically, the final spatio-temporal saliency map $STP(i, j, t)$ can be expressed by:

$$STP(i, j, t) = (N(ST(i, j, t)) + \mathbb{C}) \cdot (N(P(i, j, t)) + \mathbb{C}) \tag{12}$$

where $N(\cdot)$ denotes the normalization operation, $\mathbb{C}$ is an empirical constant and is set to 0.01 in this implementation. Due to the recursive manner in which the proposed approach operates, the introduction of constant $\mathbb{C}$ enhances the error tolerability of the recursive result.

3.6. The proposed dim moving target detection approach

As previously mentioned, the spatial saliency map and the temporal saliency map can reflect the spatial and temporal singularity characteristics, and the trajectory prediction map can reflect the motion consistency characteristic. These characteristics correctly show the intrinsic difference between the dim moving target and the complex background clutter. Hence, the final spatio-temporal saliency map defined by Eq. (12) robustly indicates the dim moving target. The proposed spatio-temporal saliency prediction module is expected to make the signal-to-clutter ratio (SCR) Gain as large as possible. In the generated spatio-temporal saliency map $STP(i, j, t)$, $|f_{STP}^t - u_{STP}|/\sigma_{STP}$ would be very large where f_{STP}^t, μ_{STP}, and σ_{STP} denote the response value of the target, the mean value of the background region, and the standard deviation value of the background region in the spatio-temporal saliency map. Hence, the target is easily segmented from the spatio-temporal saliency map $STP(i, j, t)$ based on the adaptive threshold [5,12,15]:

$$Th = \mu_{STP} + k \cdot \sigma_{STP} \tag{13}$$

where k is an empirical constant and is set to 10 in our implementation.

Based on the proposed spatio-temporal saliency approach, Algorithm 1 of the proposed dim moving target detection approach follows.

The comprehensive performance analysis of the proposed dim moving target detection approach and the comparison of the proposed approach to the state-of-the-art approaches are detailed in Section 4.

Table 1
The formed evaluation dataset.

The infrared sequence ID	The number of frames	The number of key frames	The type of background	The complexity of background	The category of target	The SNR of target
Sequence 1	100	19	Sky-sea	Relatively low	Ship	Low
Sequence 2	100	19	Ground	High	Vehicle	Relatively high
Sequence 3	125	24	Sky	Relatively low	Airplane	Relatively high

Table 2
The average evaluation scores of the intermediate results of the proposed approach.

		$S(x,y,t)$	$T(x,y,t)$	$P(x,y,t)$	$ST(x,y,t)$	$STP(x,y,t)$
Sequence 1	$\overline{\text{SCRGain}}$	302.4	196.5	885.2	516.9	**2136.3**
	$\overline{\text{BSF}}$	2.6	2.2	12.6	4.0	**23.8**
Sequence 2	$\overline{\text{SCRGain}}$	320.8	258.6	355.8	730.2	**779.9**
	$\overline{\text{BSF}}$	4.4	3.9	11.5	10.1	**21.6**
Sequence 3	$\overline{\text{SCRGain}}$	6.9	5.0	16.3	14.2	**39.6**
	$\overline{\text{BSF}}$	3.8	3.1	25.0	7.4	**36.8**

4. Experimental results

In this section, a diverse evaluation dataset and the corresponding evaluation metrics are first introduced. Then, the crucial procedures and parameters of our proposed dim moving target detection approach are verified experimentally. Finally, the comprehensive comparison with existing state-of-the-art approaches is presented.

4.1. Dataset and evaluation metrics

In order to fairly evaluate the performance of various dim moving target detection approaches, a representative dataset comprised of infrared sequences shot by the airborne platform under different backgrounds (sea-sky, ground, and sky) was formed. Table 1 describes the constructed dataset. The key frames were sequentially selected from the original sequences with fixed interval $l = 5$, and the moving targets in the key frames were manually annotated by their accurate boundaries for quantitative evaluation. In Sequence 1, the difficulty of infrared moving target detection is caused by the low SNR of the target. In Sequence 2, the detection is challenged by the complex background clutter. In Sequence 3, the cloud changes rapidly in the time axis, which makes motion analysis difficult. As a whole, the formed dataset covers a variety of situations encountered in infrared dim moving target detection. Hence, evaluation of the formed dataset fairly demonstrates the validity and performance of infrared dim moving target detection methods for airborne image sequences.

The common evaluation metrics [12,13,33], which include the SCR Gain, the background suppression factor (BSF), and the receiver operation characteristic (ROC) curve have been widely utilized in performance evaluation of infrared small target detection. More specifically, SCR Gain reflects the amplification of target signals relative to the backgrounds after and before processing, and BSF expresses the suppression level of the backgrounds. In addition, the ROC curve represents the varying relationship between the true positive rate (TPR) and the false positive rate (FPR).

In this paper, SCR Gain, BSF, and ROC curve were adopted for implementing the quantitative evaluation in the experimental section.

4.2. Overall performance of the proposed approach

In our implementation, the patch size ς and the overlap ℓ are set to 4 and 2. The number of local neighboring samples n is set to 8. In addition, the regularization term λ is set to 10^7 based on the sensitivity analysis in Section 4.3. The validity of the intermediate procedures of the proposed approach was first confirmed. Under the same parameter configuration, the proposed approach was tested on all the sequences in the constructed dataset. More specifically, the spatial saliency map $S(x,y,t)$, the temporal saliency map $T(x,y,t)$; the trajectory prediction map $P(x,y,t)$; the combination result $ST(x,y,t)$ of $S(x,y,t)$, and $T(x,y,t)$; the trajectory prediction map $P(x,y,t)$; and the spatio-temporal saliency map. The visual results are shown in Fig. 5, and the corresponding quantitative evaluation results are provided in Table 2 and Fig. 6.

As depicted in Fig. 5, $S(x,y,t)$ and $T(x,y,t)$ display the ability to reveal the small target in the preliminary results. In addition, the multiplication combination result $ST(x,y,t)$ of $S(x,y,t)$ and $T(x,y,t)$ is shown to benefit the results further by suppressing the background clutter and enhancing the target. Consistently, as illustrated in Table 2, both the SCR Gain and BSF of $ST(x,y,t)$ were larger than that of $S(x,y,t)$ and $T(x,y,t)$. From the visual and quantitative results in Figs. 5, 6 and Table 2, the target prediction performance of $P(x,y,t)$ was shown to be stable. Furthermore, the spatio-temporal saliency map $STP(x,y,t)$ produced the best target detection performance based on the visual comparison from Fig. 5 and the quantitative comparison from Table 2 and Fig. 6.

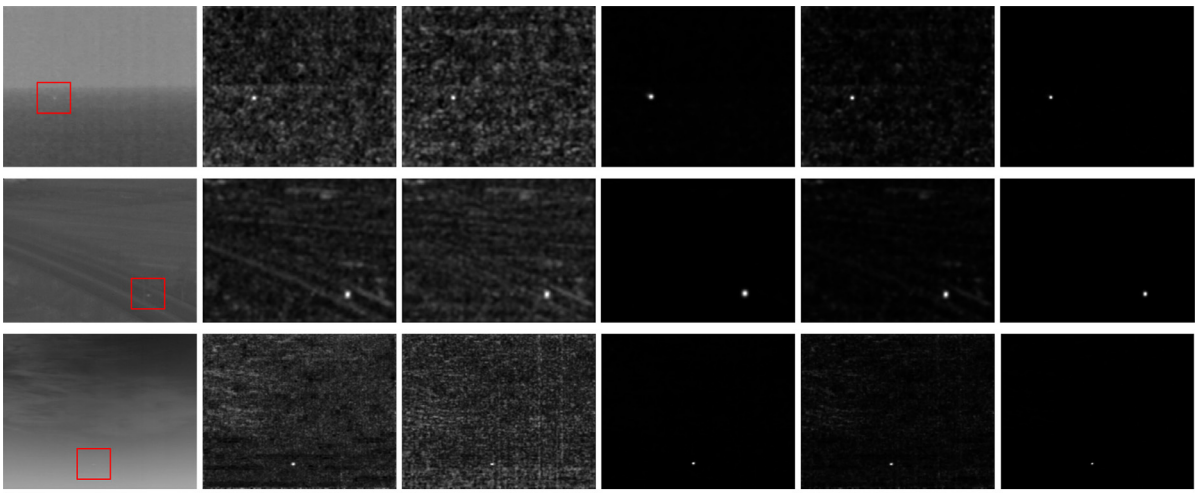

Fig. 5. Visual illustration of the intermediate results of the proposed approach. The first row, the second row, and the third row illustrate the results of the proposed approach on Sequence 1, Sequence 2, and Sequence 3, respectively; (a) denotes the original image in which the target is labeled by a red rectangle, (b) denotes the spatial saliency map, (c) denotes the temporal saliency map, (d) denotes the trajectory prediction map, (e) stands for the combination results of the spatial saliency map and the temporal saliency map, and (f) denotes the final spatio-temporal saliency map.

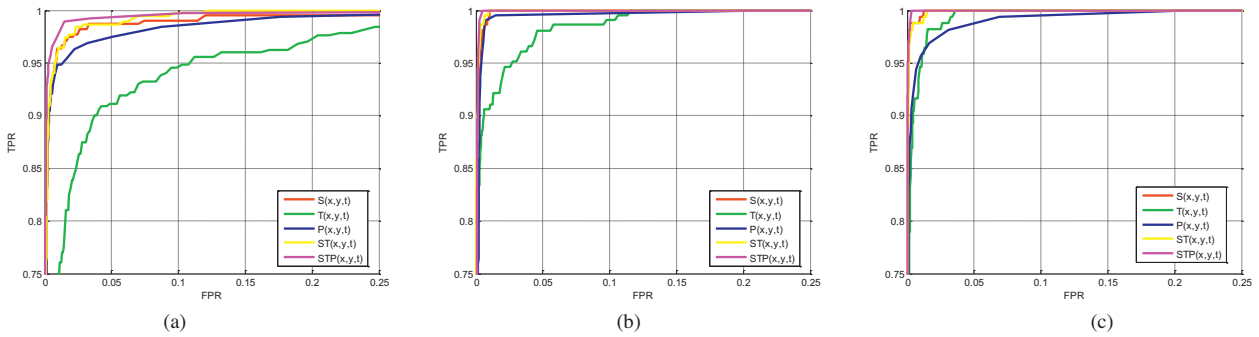

Fig. 6. The average ROC curves of the proposed approach applied to the three test sequences; (a), (b), and (c) denote the average ROC curves of the intermediate results of the proposed approach on Sequence 1, Sequence 2, and Sequence 3, respectively.

Table 3
The average evaluation scores of $ST(x, y, t)$ along with the variation of the regularized term.

		$\lambda = 10^1$	$\lambda = 10^3$	$\lambda = 10^5$	$\lambda = 10^7$	$\lambda = 10^9$
Sequence 1	$\overline{\text{SCRGain}}$	124.6	264.3	472.6	**495.9**	75.6
	$\overline{\text{BSF}}$	**4.8**	3.2	3.7	3.8	1.3
Sequence 2	$\overline{\text{SCRGain}}$	304.6	306.8	658.6	**730.2**	59.5
	$\overline{\text{BSF}}$	5.6	5.5	9.2	**10.1**	1.7
Sequence 3	$\overline{\text{SCRGain}}$	8.4	9.6	13.9	**14.21**	0.2
	$\overline{\text{BSF}}$	5.5	5.6	7.3	**7.4**	1.1

4.3. Sensitivity analysis of the regularized term

As previously discussed, the adopted local contrast operation LACRFR depends on the regularized term λ. There may be interest in the selection and sensitivity of λ under different application situations (i.e., different backgrounds) because the calculation of $S(x, y, t)$ and $T(x, y, t)$ depends on LACRFR; and λ directly influences $S(x, y, t)$ and $T(x, y, t)$ and further influences $ST(x, y, t)$, which is the combination of $S(x, y, t)$ and $T(x, y, t)$. Hence, the performance variation of $ST(x, y, t)$ along with the variation of λ, is able to reveal the sensitivity of λ.

The quantitative evaluation scores of $ST(x, y, t)$ under different λ are reported in Table 3 and Fig. 7. For Sequence 1, $\lambda = 10^1$ helped $ST(x, y, t)$ achieve the highest BSF, but $\lambda = 10^7$ helped $ST(x, y, t)$ achieve the highest SCR Gain and the best ROC Curve. Considering that the SCR Gain reflects the amplification of the target signals relative to the background after

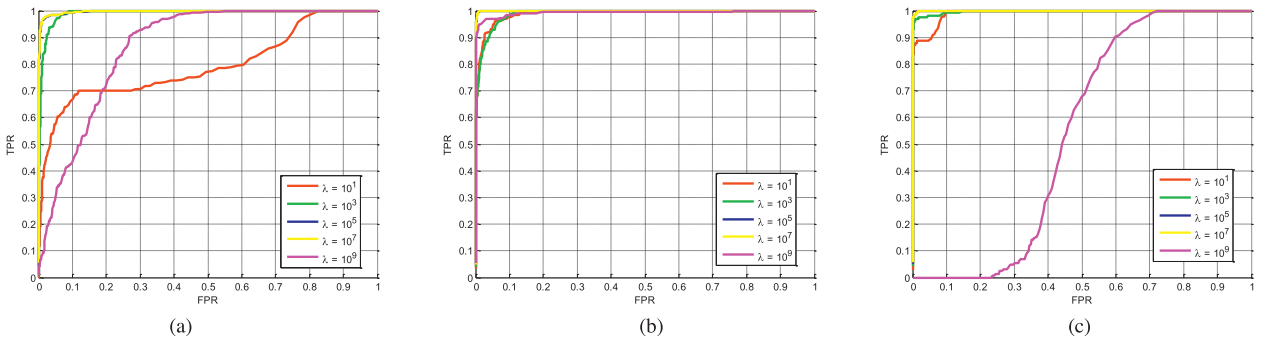

Fig. 7. The average ROC curves of the proposed approach applied to the three test sequences; (a), (b), and (c) denote the average ROC curves of $ST(x, y, t)$ under different λ on Sequence 1, Sequence 2, and Sequence 3, respectively.

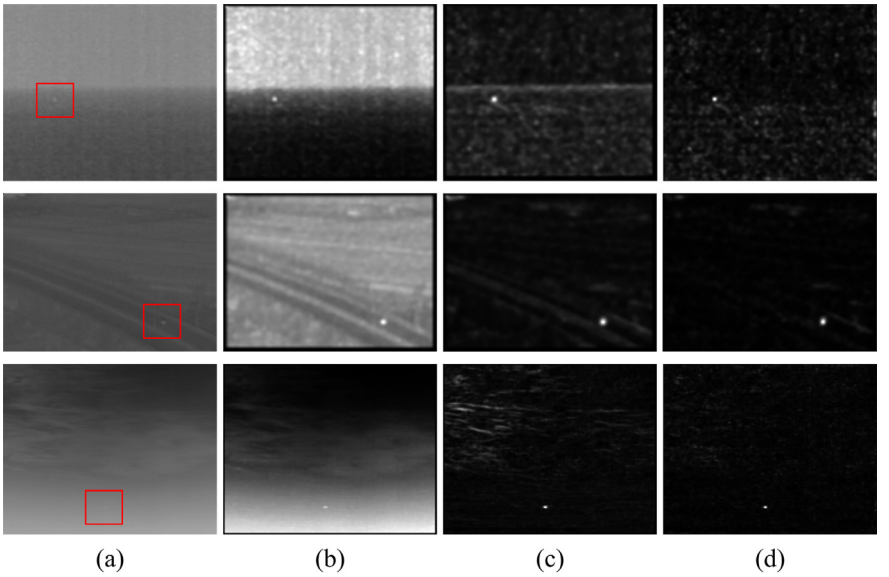

Fig. 8. The visual illustration of $ST(x, y, t)$ using different local contrast measures. The first row, the second row, and the third row visually illustrate the results of the proposed approach on Sequence 1, Sequence 2, and Sequence 3, respectively; (a) denotes the original image in which the target is labeled by a red rectangle, (b) denotes the visual results of $ST(x, y, t)$ using LCM, (c) denotes the visual results of $ST(x, y, t)$ using LCFD, and (d) denotes the visual results of $ST(x, y, t)$ using LACRFR.

and before processing and BSF only expresses the suppression level of the background without any target information, the SCR Gain is more significant compared to BSF in target detection. Hence, λ was set to 10^7 for Sequence 1. For Sequence 2 and Sequence 3, the evaluation scores including the SCR Gain, BSF, and ROC curve consistently supported that $ST(x, y, t)$ could achieve the best performance when $\lambda = 10^7$. As a whole, $\lambda = 10^7$ made $ST(x, y, t)$ achieve the best target detection performance under different backgrounds. Hence, it was concluded that the specific value of the regularized term (i.e., $\lambda = 10^7$) can help LACRFR achieve robust results under different situations.

4.4. Comparison with other local contrast measures

In this section, the superiority of the proposed LACRFR approach is verified through the experimental comparison with the aforementioned LCM [5] and LCFD in Eq. (2). Using the same configuration, LCM, LCFD, and LACRFR were taken as the LCM to implement the proposed approach, respectively. In addition, the visual results and the evaluation scores of $ST(x, y, t)$ using different local contrast measures are reported in Figs. 8, 9 and Table 4.

As depicted in Fig. 8, LCM successfully enhanced the target but lacked the ability to suppress the background clutter. On the premise of preserving the target enhancement capability, LCFR demonstrated better background suppression performance compared to LCM. However, LCFR was unable to suppress the structure clutter (i.e., the line between the sky and the sea in Sequence 1, the road in Sequence 2, and the cloud textures in Sequence 3). Fortunately, LACRFR successfully solved these problems. As depicted in Fig. 8, LACRFR not only accurately enhanced the target, but actually robustly suppressed

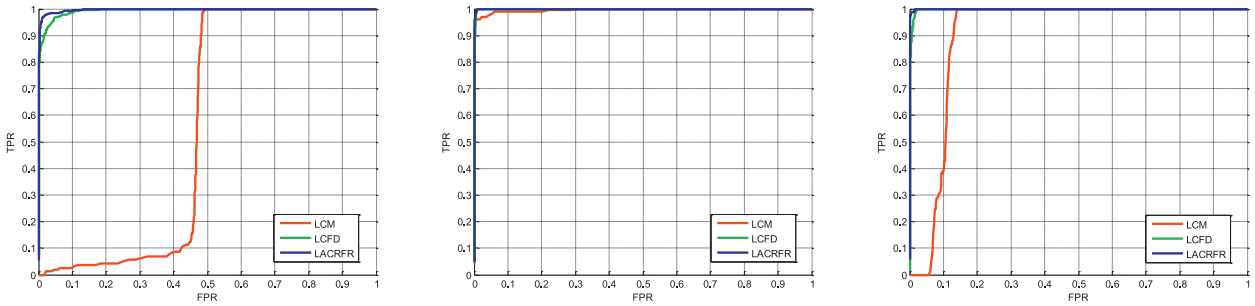

Fig. 9. The average ROC curves of the proposed approach applied to three test sequences; (a), (b), and (c) denote the average ROC curves of $ST(x, y, t)$ using different local contrast operations on Sequence 1, Sequence 2, and Sequence 3, respectively.

Table 4
The average evaluation scores of $ST(x, y, t)$ using different local contrast operations.

		LCM	LCFD	LACRFR
Sequence 1	$\overline{\text{SCRGain}}$	18.8	392.5	**495.9**
	$\overline{\text{BSF}}$	0.7	3.0	**3.8**
Sequence 2	$\overline{\text{SCRGain}}$	59.6	485.9	**730.2**
	$\overline{\text{BSF}}$	1.6	5.9	**10.1**
Sequence 3	$\overline{\text{SCRGain}}$	1.2	8.1	**14.2**
	$\overline{\text{BSF}}$	1.0	5.1	**7.4**

Table 5
The average evaluation scores of the proposed approach and existing approaches.

		New Top-Hat in [2]	LS-SVM Filter in [44]	Multi-Scale Facet in [48]	BSBR in [39]	MPPD in [23]	Our approach
Sequence 1	$\overline{\text{SCRGain}}$	1881.9	909.3	592.6	493.9	35.6	**2010.2**
	$\overline{\text{BSF}}$	18.37	14.4	15.8	5.7	2.2	**22.4**
Sequence 2	$\overline{\text{SCRGain}}$	760.6	**1267.3**	25.0	15.4	108.2	779.9
	$\overline{\text{BSF}}$	19.9	**63.3**	31.8	6.4	2.8	21.6
Sequence 3	$\overline{\text{SCRGain}}$	25.3	16.2	23.3	2.2	10.1	**39.6**
	$\overline{\text{BSF}}$	19.1	21.3	**78.7**	6.5	5.6	36.8

the complex background clutter. Compared to LCM and LCFD, LACRFR exhibited better small target detection performance, which was verified by the quantitative evaluation results shown in Table 4 and Fig. 9.

4.5. Comparison with state-of-the-art approaches

In order to show the superiority of the proposed approach, we selected three infrared small target detection approaches from the recent literature and two infrared small moving target detection approaches which utilize the image sequence to implement infrared moving target detection as the baseline. More specifically, the adopted infrared small target detection approaches, which require only one singe image to implement infrared small target detection, included the new Top-Hat based approach (New Top-Hat) in [2], the LS-SVM filter-based approach (LS-SVM Filter) in [44], and the multi-scale facet based approach (Multi-Scale Facet) in [48]. The adopted two infrared small moving target detection approaches were the background subtraction based on block reconstruction (BSBR) approach of [39] and the motion perception module based on phase discrepancy (MPPD) approach of [23]. As a whole, the comparison approaches adopted in this paper are representative of the advanced level of infrared small moving target detection today. The detection results of our proposed approach and the existing approaches are visually shown in Fig. 10. In addition, the quantitative evaluation results are provided in Table 5 and Fig. 11.

As depicted in Table 5, LS-SVM [44] achieved the highest SCR Gain and BSF in Sequence 2, but performed poorly in Sequence 1 and Sequence 3. In addition, the Multi-Scale Facet approach [48] exhibited excellent background suppression performance, but it was not applicable when the background was complex. As a whole, the success of the infrared small target detection approaches using one single image was restricted to its own specific application. Hence, exploitation of the motion cues from the image sequences was necessary to further improve the moving target detection performance. However, as depicted in Fig. 10, the existing infrared small moving target detection approaches [23,39] still performed poorly because a very challenging test dataset based on an actual application was proposed and utilized for the comparative testing in this paper. However, the proposed approach consistently performed well in all three sequences. Based on the visual comparison

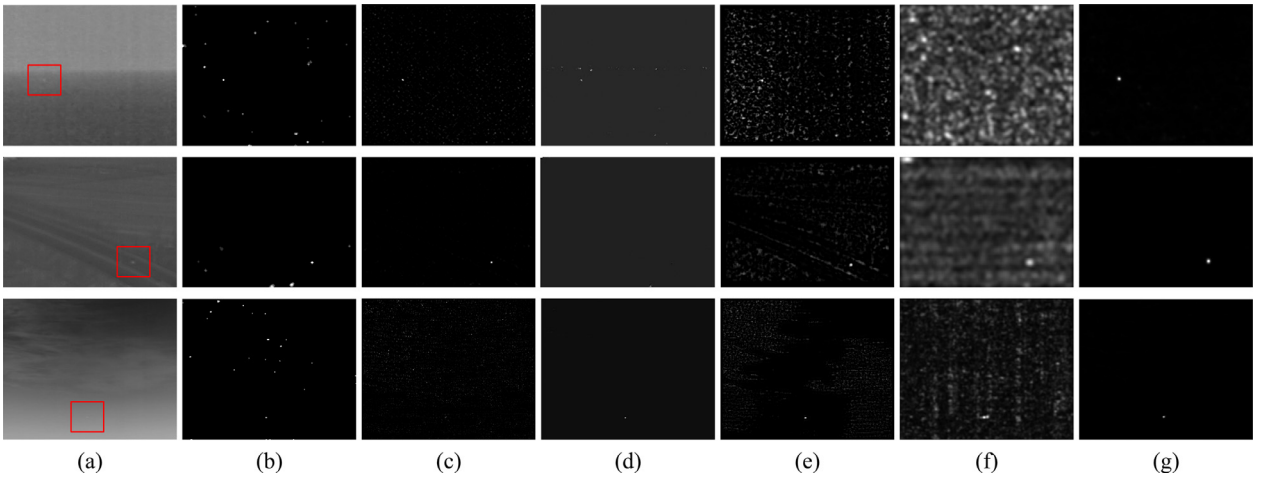

Fig. 10. The visual illustration of the results of the seven different approaches. The first row, the second row, and the third row visually illustrate the results of the proposed approach on Sequence 1, Sequence 2, and Sequence 3, respectively. (a) denotes the original image in which the target is labeled by a red rectangle, (b) denotes the detection results of New Top-Hat in [2], (c) denotes the detection results of LS-SVM filter in [44], (d) denotes the detection results of Multi-Scale Facet in [48], (e) stands for the detection results of BSBR in [39], (f) denotes the detection results of MPPD in [23], and (g) denotes the detection results of our proposed approach.

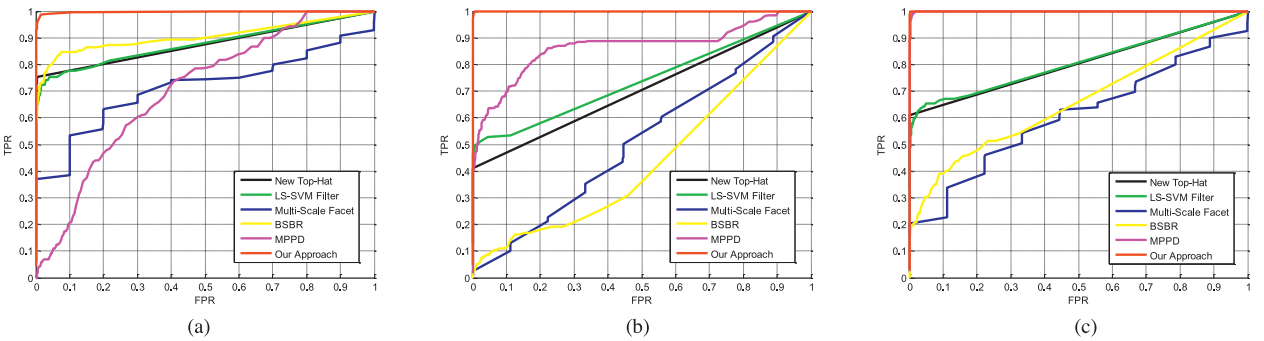

Fig. 11. The average ROC curves of the proposed approach and other existing approaches; (a), (b), and (c) denote the average ROC curves of the intermediate results of the proposed approach on Sequence 1, Sequence 2, and Sequence 3.

in Fig. 10 and the quantitative comparison in Fig. 11, it is clear that the proposed approach outperformed the existing approaches.

5. Conclusion

There is general consensus that the existing computational models for dim moving target detection continue to lag far behind the visual perceptual ability of humans. To address this problem, a novel spatio-temporal saliency approach was introduced in this paper. More specifically, based on the closed-form solution derived from regularized feature reconstruction, we proposed a local adaptive contrast operation, whereby the spatial saliency map and the temporal saliency map can be calculated on the spatial domain and the temporal domain. In order to depict the motion consistency characteristic of the moving target, a transmission operation also was proposed to generate the trajectory prediction map. The fused result of the spatial saliency map, the temporal saliency map, and the trajectory prediction map is called the spatio-temporal saliency map in this paper. Extensive experiments confirmed that the proposed spatio-temporal saliency approach achieved excellent infrared dim moving target detection performance and outperformed the state-of-the-art approaches.

In our future work, we will attempt to utilize multi-view learning methods [47] to fuse the saliency maps from the spatial domain and the temporal domain for further lifting the dim moving target detection performance. In addition, we will try to extend the proposed spatio-temporal saliency approach to more applications, such as spatio-temporal interest point detection [46], target tracking [45], action recognition [34].

Acknowledgements

This research was partially supported by the National Natural Science Foundation of China under grants 41322010, 41571434, 41371339, and 61273279, and by LIESMARS Special Research Funding. In addition, we are also grateful to the reviewers for their suggestions.

References

[1] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: application to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 28 (12) (2006) 2037–2041.
[2] X. Bai, F. Zhou, Analysis of new top-hat transformation and the application for infrared dim small target detection, Pattern Recognit. 43 (2010) 2145–2156.
[3] U. Braga-Neto, M. Choudhary, J. Goutsias, Automatic target detection and tracking in forward-looking infrared image sequences using morphological connected operators, J. Electron. Imaging 14 (3) (2004) 802–813.
[4] Z. Chen, T. Deng, L. Gao, H. Zhou, S. Luo, A novel spatial-temporal detection method of dim infrared moving small target, Infrared Phys. Technol. 66 (2014) 84–96.
[5] C. Chen, H. Li, Y. Wei, T. Xia, Y. Tang, A local contrast method for small infrared target detection, IEEE Trans. Geosci. Remote Sens. 52 (1) (2014) 574–581.
[6] M. Cheng, N. Mitra, X. Huang, P. Torr, Global contrast based salient region detection, IEEE Trans. Pattern Anal. Mach. Intell. 37 (3) (2015) 569–582.
[7] L. Deng, H. Zhu, Moving point target detection based on clutter suppression using spatiotemporal local increment coding, Electron. Lett. 51 (8) (2015) 625–626.
[8] C. Ding, J. Choi, D. Tao, L. Davis, Multi-directional multi-level dual-cross patterns for robust face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 38 (3) (2016) 518–531.
[9] C. Ding, D. Tao, A comprehensive survey on pose-invariant face recognition, Comput. Vis. Pattern Recognit. (2015) preprint, arXiv: 1502.04383.
[10] X. Dong, X. Huang, Y. Zheng, S. Bai, W. Xu, A novel infrared small moving target detection method based on tracking interest points under complicated background, Infrared Phys. Technol. 65 (2014) 36–42.
[11] D. Essen, J.H. Maunsell, Hierarchical organization and functional streams in the visual cortex, Trends Neurosci. 6 (1983) 370–375.
[12] C. Gao, D. Meng, Y. Yang, Y. Wang, X. Zhou, A. Hauptmann, Infrared patch-image model for small target detection in a single image, IEEE Trans. Image Process. 22 (12) (2013) 4996–5009.
[13] Y. Gu, C. Wang, B. Liu, Y. Zhang, A kernel-based nonparametric regression method for clutter removal in infrared small-target detection applications, IEEE Geosci. Remote Sens. Lett. 7 (3) (2014) 469–573.
[14] W. Guo, G. Chen, Human action recognition via multi-task learning base on spatial-temporal feature, Inf. Sci. 320 (2015) 418–428.
[15] J. Han, Y. Ma, B. Zhou, F. Fan, K. Liang, Y. Fang, A robust infrared small target detection algorithm based on human visual system, IEEE Geosci. Remote Sens. Lett. 11 (12) (2014) 2168–2172.
[16] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Mach. Intell. 20 (11) (1998) 1254–1259.
[17] J. Jiang, R. Hu, Z. Wang, Z. Han, Noise robust face hallucination via locality-constrained representation, IEEE Trans. Multimedia 16 (5) (2014) 1268–1281.
[18] S. Kim, High-speed incoming infrared target detection by fusion of spatial and temporal detectors, Sensors 15 (4) (2015) 7267–7293.
[19] S. Kim, J. Lee, Scale invariant small target detection by optimizing signal-to-clutter ratio in heterogeneous background for infrared search and track, Pattern Recognit. 43 (1) (2012) 393–406.
[20] Z. Li, J. Chen, Q. Hou, H. Fu, Z. Dai, G. Jin, R. Li, C. Liu, Sparse representation for infrared dim target detection via a discriminative over-complete dictionary learned online, Sensors 14 (6) (2014) 9451–9470.
[21] Z. Li, Z. Dai, H. Fu, Q. Hou, Z. Wang, L. Yang, G. Jin, C. Liu, R. Li, Dim moving target detection algorithm based on spatio-temporal classification sparse representation, Infrared Phys. Technol. 67 (2014) 273–282.
[22] Z. Li, Q. Hou, H. Fu, Z. Bai, L. Yang, G. Jin, R. Li, Infrared small moving target detection algorithm based on joint spatio-temporal sparse recovery, Infrared Phys. Technol. 69 (2015) 44–52.
[23] Y. Li, Y. Tan, H. Li, T. Li, J. Tian, Biologically inspired multilevel approach for multiple moving targets detection from airborne forward-looking infrared sequences, J. Opt. Soc. Amer. A 31 (4) (2014) 734–744.
[24] Y. Li, Y. Tan, J.-G. Yu, J. Tian, Kernel regression in mixed feature spaces for spatio-temporal saliency detection, Comput. Vis. Image Underst. 135 (2015) 126–140.
[25] Y. Li, C. Tao, Y. Tan, K. Shang, J. Tian, Unsupervised multilayer feature learning for satellite image scene classification, IEEE Geosci. Remote Sens. Lett. 13 (2) (2016) 157–161.
[26] H. Li, Y. Tan, Y. Li, J. Tian, Image layering based small infrared target detection method, Electron. Lett. 50 (2014) 42–44.
[27] D. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[28] J. Ma, J. Zhao, A. Yuille, Non-rigid point set registration by preserving global and local structures, IEEE Trans. Image Process. 25 (1) (2016) 53–64.
[29] J. Ma, H. Zhou, J. Zhao, Y. Gao, J. Jiang, J. Tian, Robust feature matching for remote sensing image registration via locally linear transforming, IEEE Trans. Geosci. Remote Sens. 53 (12) (2015) 6469–6481.
[30] W. Meng, T. Jin, X. Zhao, Adaptive method of dim small object detection with heavy clutter, Appl. Opt. 52 (2013) D64–D74.
[31] F. Perazzi, P. Krahenbuhl, Y. Pritch, A. Hornung, Saliency filters: contrast based filtering for salient region detection, in: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2012, pp. 733–740.
[32] S. Qi, J. Ma, C. Tao, C. Yang, J. Tian, A robust directional saliency-based method for infrared small-target detection under various complex backgrounds, IEEE Geosci. Remote Sens. Lett. 10 (3) (2013) 495–499.
[33] S. Qi, D. Ming, J. Ma, X. Sun, J. Tian, Robust method for infrared small-target detection based on boolean map visual theory, Appl. Opt. 53 (18) (2014) 3929–3940.
[34] K. Rapantzikos, Y. Avrithis, S. Kollias, Dense saliency-based spatiotemporal feature points for action recognition, in: IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 1454–1461.
[35] Z. Ren, S. Gao, L. Chia, D. Rajan, Regularized feature reconstruction for spatio-temporal saliency detection, IEEE Trans. Image Process. 22 (8) (2013) 3120–3132.
[36] X. Shi, N. Bruce, J. Tsotsos, Fast, recurrent, attentional modulation improves saliency representation and scene recognition, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2011, pp. 1–8.
[37] K. Simonyan, A. Zisserman, Two-stream convolutional networks for action recognition in videos, Adv. Neural Inf. Process. Syst. (2014) 568–576.
[38] B. Subudhi, S. Ghosh, S. Cho, A. Ghosh, Integration of fuzzy markov random field and local information for separation of moving objects and shadows, Inf. Sci. 331 (2016) 15–31.
[39] A. Sun, Y. Tan, J. Tian, Small target detection using min-cut and non-balanced graph, in: Proceedings of Multispectral Image Processing and Pattern Recognition, 2013, p. 89181C.
[40] V. Tom, T. Peli, M. Leung, J. Bondaryk, Morphology-based algorithm for point target detection in infrared backgrounds, in: Proceedings of Multispectral Image Processing and Pattern Recognition, 1993, pp. 2–11.
[41] G. Wang, C. Chen, X. Shen, Faced-based infrared small target detection method, Electron. Lett. 41 (2005) 1244–1246.

[42] B. Wang, S. Liu, Q. Li, R. Lei, Blind-pixel correction algorithm for an infrared focal plane array based on moving-scene analysis, Opt. Eng. 45 (3) (2006) 364–367.
[43] J. Wang, Y. Lu, L. Gu, C. Zhou, X. Chai, Moving object recognition under simulated prosthetic vision using background-subtraction-based image processing strategies, Inf. Sci. 277 (2014) 512–524.
[44] P. Wang, J. Tian, C. Gao, Infrared small target detection using directional highpass filters based on LS-SVM, Electron. Lett. 45 (2009) 156–158.
[45] F. Wang, Y. Zhen, B. Zhong, R. Ji, Robust infrared target tracking based on particle filter with embedded saliency detection, Information Sciences 301 (2015) 215–226.
[46] S. Wong, R. Cipolla, Extracting spatiotemporal interest points using global information, in: IEEE International Conference on Computer Vision, 2007, pp. 1–8.
[47] C. Xu, D. Tao, C. Xu, A survey on multi-view learning, Neural Comput. Appl. (2013) preprint, arXiv: 1304.5634.
[48] C. Yang, J. Ma, M. Zhang, S. Zheng, X. Tian, Multiscale facet model for infrared small target detection, Infrared Phys. Technol. 67 (2014) 202–209.
[49] X. Zhen, L. Shao, X. Li, Action recognition by spatio-temporal oriented energies, Inf. Sci. 281 (2016) 295–309.