Infrared Dim and Small Target Detection Based on Stable Multisubspace Learning in Heterogeneous Scene

Xiaoyang Wang, Zhenming Peng, *Member, IEEE*, Dehui Kong, and Yanmin He

Abstract—Infrared (IR) dim and small target detection in a highly complex background play an important role in many applications, and remain a challenging problem. In this paper, a novel method named stable multisubspace learning is presented to deal with this problem. The new method takes into account the inner structure of actual images so that it overcomes the shortage of the traditional method. First, by analyzing the multisubspace structure of heterogeneous background data, a corresponding image model is proposed using subspace learning strategy. This model is also stable to noise interference. Second, an efficient optimization algorithm is designed to solve the proposed IR image model. By adding the proper postprocessing procedure, we can get the detection result. Experiments on simulation scenes and real scenes show that the proposed method has superior detection ability under heterogeneous background.

Index Terms—Dim and small target detection, heterogeneous scene, infrared (IR) images, multisubspace learning.

I. INTRODUCTION

TARGET detection techniques in infrared (IR) images have been widely used in many applications, such as accurate guidance, antimissile techniques, early warning, earth observation, space debris exploration, and so on [1]–[3]. These applications desire to acquire accurate information of interested targets. Because of the long imaging distance of IR detection systems, targets always have no fixed shape or texture feature, with changing size from 2×2 to 9×9 pixels. It is because the quite small size of targets makes the detection methods sensitive to background clutter and noise, which will cause false or miss alarms. Therefore, detecting small targets from IR images with a complex background is of great practical significance, and it still remains an open problem.

Typical IR scenes in aforesaid applications include deep space background, sky-cloud background, sea background, and terrain background. Without the interference of cloud, most of the deep space backgrounds are uniform and self-correlated.

In the sky-cloud background, there exist interferences such as cirrus, banded cloud, and floccus. Different structures have different IR radiation distribution and intensity. In the sea background, waves are changing quickly, which means unstable IR radiation. In the terrain background, IR radiation and reflection are much more complicated due to the ground condition, like buildings and trees. Also, artificial objects can become severe interferences in many target detection tasks. Thus, the complex sky-cloud background, complex sea background, and terrain background contain diverse IR radiation sources, and can be considered as heterogeneous scenes. Fig. 1 shows some representative uniform backgrounds and heterogeneous backgrounds. From left to right, Fig. 1(a) shows four uniform backgrounds with sky and sea. Fig. 1(b) shows the IR images of complex sky-cloud scene, sea, mountain and man-made structure scene, cloudy sky and building scene, and a terrain scene.

Compared with the optical imagery, IR images seem to be much simpler, because of the lack of color information, texture information, and poor image quality. But in the case of small target detection, the characteristics of IR images will bring much challenge. The lack of color and texture information means there are fewer features available for distinguishing targets from background. Low image quality means the boundary between different parts of images is not clear. Besides, there are always more than one strong IR radiation sources in a heterogeneous scene. Because of the small size of targets, little interference will lead to wrong detection results. Regular detection approaches in optical images are no longer valid in IR images, especially in heterogeneous scenes. Therefore, it is necessary to develop specific IR small target detection algorithms.

Traditional IR target detection approaches, like Tophat filter [4], Maxmean, and Maxmedian filter [5], are commonly used to reduce background clutter and enhance targets. Backgrounds are estimated by opening operation, max-mean filter, and max-median filter with specific structural element, respectively. The filtering-based background estimation methods perform well in IR scenes with simple uniform background and simple point targets. Besides, the IR background can also be estimated by data regression. Considering the background scenes with no priori, nonparametric regression is introduced to estimate the changing background [6]. From another perspective, the backgrounds and targets are drawn from different subspaces. In heterogeneous scenes, the background data lie on multiple subspaces. Based on the subspace structure of IR data, a state-of-the-art approach called IR patch-image (IPI) model is proposed [7]. It concentrates on the nonlocal self-
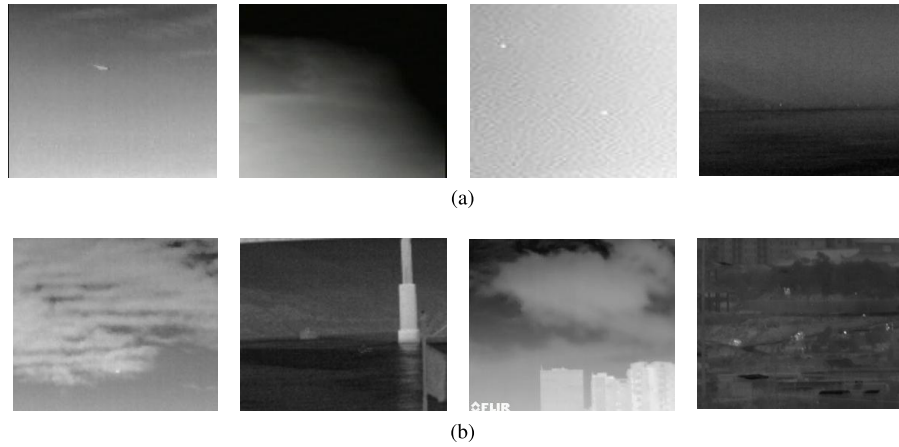
Fig. 1. Representative IR uniform scenes and heterogeneous scenes. (a) Four uniform scenes, which can be considered to be nonlocal self-correlated. (b) Four heterogeneous scenes. In these scenes, backgrounds are complex.
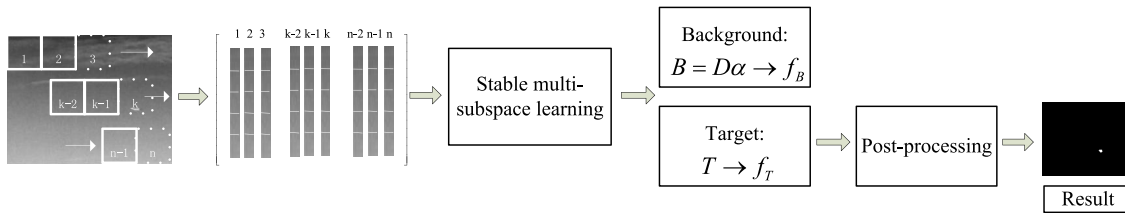

Fig. 2. IR small target detection procedure proposed in this paper using SMSL.

correlation property of IR background. The background data are modeled by a single low-rank subspace in the IPI model. This method performs well in slow-changing and uniform background, in which the background data are drawn from a single subspace. That is to say, when background data come from several different subspaces or have distinct parts, the fundamental assumption of single subspace structure in IPI model will no longer hold. This will cause the decrease of background estimation performance, thereby increasing false alarm rate in the detection result.

Our intention is to propose a novel background measure by the multisubspace property, and further develop a small target detection approach. The motivation is mainly based on the analysis that the data in IR images come from diverse radiation sources, which can be modeled by multisubspace structure. For complex background, multisubspace model is a reasonable measure. For uniform and slow-changing background, the multisubspace measure will also be suitable. Compared with single subspace model, it gives a more precise description of background, which will improve the detection accuracy.

To solve the IR small target detection problem, especially for separating small target from highly heterogeneous scenes, we propose a novel method termed stable multisubspace learning (SMSL). Using the multisubspace recovery theory and a subspace learning strategy, we construct an optimization model containing group sparsity under orthonormal subspaces and the ℓ_1 norm. The main contributions of this paper are summarized as follows.

1) By analyzing the structure of background data, a stable multisubspace model is proposed to separate IR small target and complex background, as well as dealing with unavoidable noise. We promote the state-of-art method based on single subspace hypothesis to multisubspace

data using a subspace learning strategy. This promotion shows a good performance when dealing with highly heterogeneous background.

2) An optimization solver based on accelerated proximal gradient (APG) is developed to solve the proposed model, with a shrinking procedure of group sparsity. This solver is of high efficiency.

3) Our method is autonomous, i.e., no preset dictionary is required. When choosing the proper parameter given in this paper, the target and background will be separated automatically. Noise is removed in this procedure as well.

The remainder of this paper is organized as follows. In Section II, we review the related work about IR small target detection and subspace methods. In Section III, we give an overview of the detection procedure in IR images using the proposed method. In Section IV, an algorithm is proposed to solve the model. In Section V, several experiments on simulation scenes and real scenes are presented to demonstrate the efficiency of the proposed method. The discussion is also given. The conclusions and future work are shown in Section VI.

II. RELATED WORK

In general, the existing detection methods can be divided into two types: detection before track (DBT) method and track before detection (TBD) method. Typical DBT methods include single image detection method based on image filtering and pattern recognition. Image filtering-based methods contain Tophat filtering method [4], Maxmedian and Maxmean method [5], 2-D least mean square method [8], transform domain method [9], [10], and so on. Note that the transform domain methods are usually implemented in spectral domain,

while other image filtering-based methods are implemented in spatial domain. For spectral domain methods, the background is considered to be of low frequency, while the target is considered to be of high frequency. The difference of frequency component between the background and target can be used to detect target from smooth and uniform background. Pattern recognition-based method treats dim and small target detection as an classification problem, dividing pixels into the background class and the target class [11]. In recent years, visual saliency-based detection methods have also become a research trend. A typical saliency method is called local contrast method, which computes the local contrast map of the IR image to enhance the target [12]. After that, multiple saliency methods are proposed, including improved local contrast measure [13], weighted local difference measure [14], revised visual attention model [15], and so on. Note that these methods have one thing in common: they rely on the algorithm of detecting target from a single image frame. In other words, some DBT methods value the spatial information over temporal information. In some situation, the interframe temporal information should also be used, which leads to the TBD methods. Typical TBD methods include 3-D matched filtering [16], [17], dynamic programming method [18]–[20], and maximum likelihood method [21]. Also, there exist types of methods that treat small target detection problem as a 1-D signal process problems [22]. Actually, temporal hypothesis testing method is an extent of this idea, which is good at detecting point targets [23], [24].

Except the traditional DBT and TBD methods, new methods concentrating on the data structure of IR images have attracted increasingly interests. The data structure is usually described by two ways: one is to find the low-dimensional subspace structure and the other is using a preset over-complete dictionary to reveal the data structure. The typical approach of the former way is the IPI model [7], which seeks the low-rank subspace structure of background data and the sparse structure of foreground data. A similar approach can also been done in compressive domain using compressive measure and rank estimate [3]. As for the dictionary representation method, a method called low-rank and sparse representation uses an over-complete dictionary to find the target [25]. Also, a spatiotemporal over-complete dictionary is designed to classify image pixels to target and background [26]. Detection method based on robust dictionary learning is presented using structure sparse coding [27].

Subspace methods are commonly used approaches to reveal the underlying structure of data. Some methods as principal component analysis (PCA) and robust PCA (RPCA) [28] aim at learning a single subspace to represent the data structure while removing the noise or random corruptions. RPCA is also extended to a more stable version called stable principal component pursuit (SPCP), which has high stability in noise conditions [29]. These methods are widely used in visual data analysis and remote sensing applications such as motion detection [30], image fusion [31]–[34], image classification [35], and so on. The aforesaid IPI model also uses SPCP method to detect IR small targets, where the uniform background data can be seen as lying in a single low-rank subspace.

However, the single subspace structure does not match the reality in some cases. A more reasonable assumption is that real data lie near a union of linear subspaces. A representative multisubspace methods are low-rank representation (LRR) [36] and sparse subspace clustering (SSC) [37]. LRR finds the LRR of data through a preset dictionary, while SSC finds the sparse representation. The dictionary is constructed by the input data itself, which can be seen as a self-expression process. It has been shown that these methods can achieve more accurate multisubspace estimation than traditional single subspace methods. Thus, the multisubspace assumption is widely used in image classification [38], face clustering [39], motion segmentation [40], and so on.

III. MODEL CONSTRUCTION AND DETECTION PROCEDURE

In general, the IR image data can be described by an additive formulation as follows, no matter how the data are collected [6], [7]:

$$f(x, y) = f_T(x, y) + f_B(x, y) + f_N(x, y) \qquad (1)$$

where $f(x, y)$ is the gray level (gray scale) of a pixel (x, y) in the original image, $f_T(x, y)$ and $f_B(x, y)$ are the gray levels of *target* and *background*, respectively, and $f_N(x, y)$ stands for the intensity of *noise*.

Fig. 2 shows the whole method of detecting IR dim and small target proposed in this paper. Based on the above additive model, we first slide a window from left and top to right and down in the original image. A series of image blocks is acquired. Then we vectorize each block to gain the patch model I, where $I = T + B + N$, corresponding to (1). We perform the SMSL procedure on model I, and we get the target data T and background data $B = D\alpha$, where D stands for the multisubspace dictionary, which is acquired from data I itself during the procedure. α is the coefficients. Then we reverse transform the data T and B to real image domain. Because of obtaining the target is our purpose in this application, we threshold the target data, and get the detection result.

When constructing the model I, we set the vertical and horizontal sliding step to be equal to the height and width of window, respectively. In other words, there is no overlap between each image block. The generated model I has two advantages compared with the original image. The first is that in most cases, pixels in one window contain only one type of scene due to the small window size. The nonoverlap patch model will enhance the multisubspace property of heterogeneous background. The second is that pixels in model I can be considered as a rearrangement of pixels in the original IR image $f(x, y)$, which means this model does not cause extra computation complexity.

IV. TARGET DETECTION USING STABLE MULTISUBSPACE LEARNING

In this section, we introduce the proposed SMSL method and corresponding dim and small target detection procedure. This procedure applies the multisubspace-cluster assumption of background region, making it suitable for dealing with highly heterogeneous scenes.

A. Multisubspace Structure of Infrared Images

The process of detecting dim and small target is to separate background cluster and target region, as well as eliminating random noise from original data, simultaneously. Each part contains some types of data structures that enable intelligent separation.

As analyzed in Section I, background region has the majority of pixels and energy in IR images. Due to atmospheric environment and current image acquisition technology, some background regions in IR images have the nonlocal self-similarity property [7], [41]. Thus, f_B is considered as low-rank matrices. For the uniform background like blurred sea-sky background, cloudy background, and so on, a single low-rank subspace is generally enough to model the background situation, which means the pixel distribution structures in background region are of some similarity. However, when dealing with highly heterogeneous scene [see Fig. 1(b)], there are different types of IR radiation sources. It is more reasonable to assume that the background data are lying near several low-rank subspaces rather than a single one.

B. Formulation of Stable Multisubspace Learning Method

From the above discussion, background data f_B come from a mixture of low-rank subspaces. Assuming that $I \in R^{m \times n}$, to better handle the mixed background data B, a general rank minimization procedure in [36] can be used

$$\min_{\alpha} \ \text{rank}(\alpha), \quad \text{s.t.} \ B = D\alpha \tag{2}$$

where $D = [D_1, D_2, \ldots, D_k] \in R^{m \times k}$ is a dictionary that spans the background data space, $\alpha = [\alpha_1; \alpha_2; \ldots; \alpha_k] \in R^{k \times n}$ is the coefficients, and k is the subspace dimension. This can be regarded as a subspace learning problem. By choosing an appropriate dictionary D, the solution to (2) could have good performance in distinguishing the data drawn from a union of multiple low-rank subspaces with theoretical guarantees [36].

The rank function (2) has a discrete nature, which makes it difficult to solve. We should find a proper surrogate to replace this rank measure. The rank of α should not be larger than its number of nonzero rows, resulting in $\text{rank}(B) \leq \text{rank}(\alpha) \leq \|\alpha\|_{\text{row}-0}$. Also, like the relationship between ℓ_0-norm and ℓ_1-norm, the row-1 norm, which is defined as $\|\alpha\|_{\text{row}-1} = \sum_{i=1}^{k} \|\alpha_i\|_2$, is a good heuristic for row-0 norm. Then the rank function (2) could be replaced by the following minimization problem:

$$\min_{\alpha} \ \|\alpha\|_{\text{row}-1}, \quad \text{s.t.} \ B = D\alpha. \tag{3}$$

Motivated by [42], instead of choosing a preset dictionary to reveal the multisubspace structure of input data, we can seek an adaptive compact subspace from the data itself to represent low-rank matrix. We hope that the subspace bases and coefficients can best fit the constraint $B = D\alpha$. In this occasion, there should be more restrictions on D

$$\min_{D,\alpha} \ \|\alpha\|_{\text{row}-1}, \quad \text{s.t.} \ B = D\alpha, \ D^T D = I_k \ \forall i \tag{4}$$

where the columns of D span an orthonormal space. Under the above assumption, the correlation of columns of D is low, which enables $\|\alpha\|_{\text{row}-1}$ to measure the data from multiple

Fig. 3. IR target and local background area.

low-rank subspaces effectively. Also, this subspace learning procedure avoids iterations of singular value decomposition (SVD), which will lead to high computational efficiency.

As to the target image f_T, it is less relevant to background data, which can be considered as corruption. Also, it is small with respect to the whole image, so its corresponding data model T has sparse supports, which can be solved by ℓ_1 norm minimization in practice. Also, original IR images contain some random noises f_N, which can be assumed as independent identically distributed. Considering these factors, in order to separate B and T from I, a straightforward approach is to minimize the following function:

$$\min_{D,\alpha,T} \ \|\alpha\|_{\text{row}-1} + \lambda \|T\|_1$$
$$\text{s.t.} \ \|I - D\alpha - T\|_F \leq \delta, \quad D^T D = I_k \ \forall i \tag{5}$$

where $\lambda > 0$ is a regularization parameter, $\delta > 0$ is a parameter related to noise, and $\| \cdot \|_F$ is the Frobenius norm [i.e., $\|X\|_F = (\sum_{ij} X_{ij}^2)^{1/2}$]. This model is stable to small noise perturbation. By solving the above convex optimization problem, we can automatically obtain an accurate multisubspace model of complex IR background and a sparse matrix, which represents the small target.

C. Optimization Algorithm

We presented an efficient solver to the problem in (5). Note that the noise part N is constrained by an inequality in (5), and it can be translated to

$$\min_{D,\alpha,T} \|\alpha\|_{\text{row}-1} + \lambda \|T\|_1 + \frac{1}{2\mu} \|I - D\alpha - T\|_F^2,$$
$$\text{s.t.} \ D^T D = I_k \ \forall i \tag{6}$$

where $\mu(\delta)$ is a regularization parameter. It is proved that (6) is equivalent to (5) for some value of μ, and the recovery accuracy of $D\alpha$ and T is inversely proportional to noise level δ [43]. In practice, f_N has small amplitude compared with valid data, and thus the recovery error is small.

The first two terms in (6) are convex, and thus a fast proximal gradient algorithm called APG in [44] can be applied. The algorithm is precisely stated in Algorithm 1, where k is the initialization of subspace dimension and $S_\varepsilon[X] = \max\{\text{abs}(X) - \varepsilon, 0\} \cdot \text{sign}(X)$ is a widely used soft-thresholding operator in compressive sensing [45].

In Step 5, we apply the BCD method [46], [47] to solve each pair of (D_i, α_i). Also, the constraint $D^T D = I_k$ is guaranteed during the procedure. The details of BCD algorithm are given in Algorithm 2.

The key step is to shrink the subspace coefficient and receive the LRR under multisubspace bases. Fortunately, we find that

---

Algorithm 1 APG/BCD Solver to the Proposed Method

Input:

    IR image model $I \in R^{m \times n}$, k, λ;

Output:

    $D = D_j$, $\alpha = \alpha_j$, $T = T_j$;

1: Initialize: $D^0 = D^{-1} = zeros(m, k)$, $\alpha^0 = \alpha^{-1} = rand(k, n)$, $T^0 = T^{-1} = zeros(m, n)$;
    $t^0 = t^{-1} = 1$; $\mu_0 > 0$, $\bar{\mu} > 0$; $\eta < 1$;

2: **while** not converged **do**

3:    $Y_j^{B=D\alpha} = B_j + \frac{t_{j-1}-1}{t_j}[B_j - B_{j-1}], Y_j^T = T_j + \frac{t_{j-1}-1}{t_j}(T_j - T_{j-1})$;

4:    $G_j^B = Y_j^B - \frac{1}{2}(Y_j^B + Y_j^T - I)$;

5:    Using block coordinate descent (BCD) scheme and amplitude shrinking to find D_{j+1} and α_{j+1};

6:    $B_{j+1} = D_{j+1}\alpha_{j+1}$;

7:    Update subspace dimension k to the number of columns of D_{j+1};

8:    $G_j^T = Y_j^T - \frac{1}{2}(Y_j^B + Y_j^T - I)$;

9:    $T_{j+1} = S_{\frac{\lambda \mu_j}{2}}[G_j^T]$;

10:   $t_{j+1} = \frac{1+\sqrt{4t_j^2+1}}{2}$;

11:   $\mu_{j+1} = \max(\eta \mu_j, \bar{\mu})$;

12:   $j = j + 1$;

13: **end while**

Algorithm 2 BCD

Input:

    G_j^B, μ_j, k;

Output:

    D_{j+1}, α_{j+1};

1: **for** $i = 1 : k$ **do**

2:    $R_j^i = G_j^B - \sum_{p<i} D_{j+1}^p \alpha_{j+1}^p - \sum_{p>i} D_j^p \alpha_j^p$, $R_j^i = R_j^i - \sum_{p=1}^{i-1} D_{j+1}^p (D_{j+1}^p)^T R_j^i$;

3:    $D_{j+1}^i = R_j^i \alpha_j^{i\,T}$;

4:    $D_{j+1}^i = D_{j+1}^i / (\left\| D_{j+1}^i \right\|_2)$;

5:    $\alpha_{j+1}^i = S_{\frac{\mu_j}{2}}[\left\| D_{j+1}^{i\,T} R_j^i \right\|_2]$;

6: **end for**

7: Delete zero rows in α_{j+1} and corresponding columns in D_{j+1}

---

the row-1 norm optimization problem in the above procedure has a closed-form expression, as shown in Step 5. Once we received α_k and D_k from intermediate variable G_k^B, α_{k+1} can be given by soft-thresholding the $\ell - 2$ norm of rows in α_k. This achieves the goal of minimizing $\|\alpha\|_{\text{row}-1}$, or reducing the rank of α. Computational efficiency can be improved by deleting zero rows in α_{k+1} and corresponding columns in D_{k+1}; for this amplitude, thresholding cannot be recovered in the next steps. The experimental results have proved that this procedure can reach good performance. To the best of our knowledge, there is no similar method combined APG with row-1 norm soft-thresholding, which makes our solver unique and effective.

It is vital to choose an appropriate μ in the solving process. From (6), we can see that with the increase in μ, $1/2\,\mu$ will decrease, which leads to an increase in $\|I - D\alpha - T\|_F^2$ in the optimal result. Thus, μ should be large enough to get rid of the noise, but not too large to overshrink $D\alpha$ and T. From Algorithm 1, we can see that μ varies during the solving procedure, with an initial value μ_0 and a lower bound $\bar{\mu}$.

While our analysis shares some similar features with previous work in low-rank matrices recovery [42], it is considerably more challenging due to the fact that real IR images contain various degrees of noise. The existence of noise requires for a modification of matrix recovery theory. Without this modification, we will get a model deviating from reality.

D. Computational Complexity

We make a brief discussion of the computational complexity of the proposed method. As shown in Fig. 2, the dominant computation in the algorithm is the SMSL process. The time consumption of the SMSL algorithm (please see Algorithm 1) is mainly determined by two steps: 1) BCD (Step 5) and 2) the soft-thresholding operation (Step 9).

For the BCD algorithm, detailed steps are shown in Algorithm 2, whose complexity is determined by the subspace dimension and the matrix size, i.e., $O(kmn)$, where m
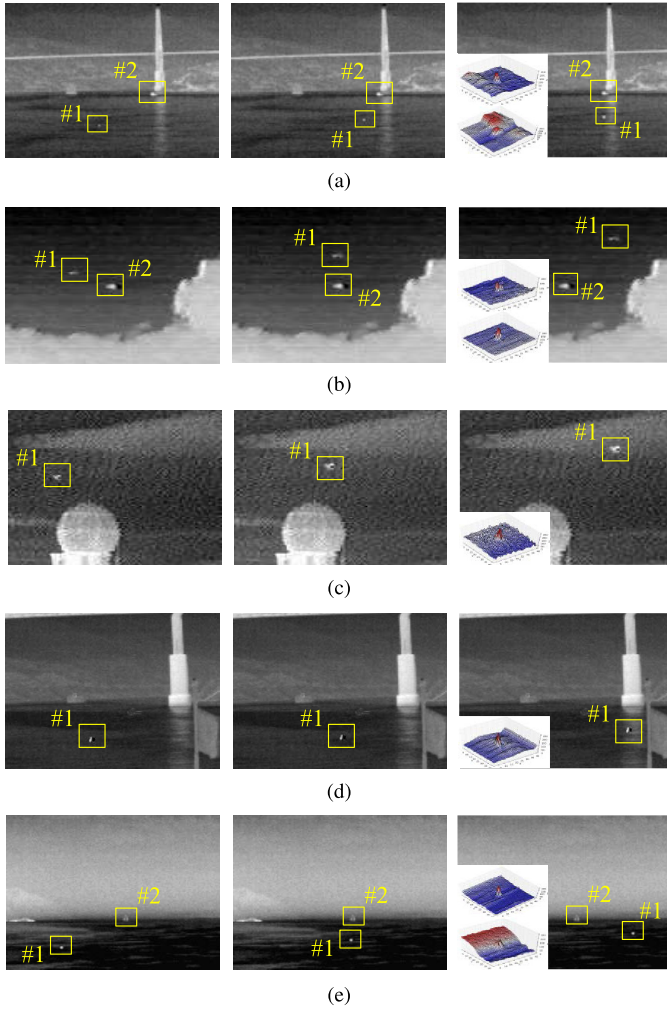
Fig. 4. (a)–(e) Simulation scenes 1–5 and 3-D projections of IR small targets in each scene. Rectangles with numbers mark the position of IR small targets. For scenes with more than one target, 3-D projections are shown in the order of target number.

and n represent the size of patch model and k is the subspace dimension.

For the soft-thresholding operation, it has a complexity of $O(mn)$. Assuming that the iteration number of Algorithm 1 is N, then the entire computational complexity of the proposed method is around $O(Nkmn)$.

V. EXPERIMENTS AND ANALYSIS

We test our algorithm on both simulation scenes and real scenes. First, we introduce the evaluation metrics for our experiments. Then we introduce the test scenes and the corresponding results. Finally, the time-consuming comparison of each method is given.

A. Evaluation Metrics

Here we introduce several frequently used metrics to evaluate the performance of target detection algorithms in IR images.

The signal-to-clutter ratio (SCR) is a measurement of target saliency. It can also be used to measure the detection difficulty. SCR is calculated in the neighborhood area around the target

(see Fig. 3), which is defined as

$$\text{SCR} = \frac{|\mu_t - \mu_b|}{\sigma_b} \tag{7}$$

where μ_t and μ_b are the average pixel value of the target region and a neighborhood area around the target, separately; σ_b is the corresponding standard deviation of neighborhood area. Usually, the higher the SCR is, the easier the detection will be. In our experiment, we set $d = 20$. Then we introduce SCR gain, which is defined as

$$G_{\text{SCR}} = \frac{\text{SCR}_{\text{out}}}{\text{SCR}_{\text{in}}} \tag{8}$$

where SCR_{in} and SCR_{out} stand for the SCR of original IR image and the image processed by the detection algorithms, respectively. Higher G_{SCR} means the detection algorithm has better target enhancement ability. Another evaluation indicator is the background suppression factor (BSF), which shows the background suppression ability of detection algorithms

$$\text{BSF} = \frac{C_{\text{in}}}{C_{\text{out}}} \tag{9}$$

where C_{in} and C_{out} are the standard deviation of background region in the original image and the processed image, respectively.

Despite the target enhancement and background suppression ability measurement, we also introduce the probability of detection P_d and the false alarm points ratio $F_a\text{PR}$. They are defined as follows:

$$P_d = \frac{\text{number of true detections}}{\text{number of actual targets}} \tag{10}$$

$$F_a\text{PR} = \frac{\text{number of pixels in false detections}}{\text{number of pixels in all test images}}. \tag{11}$$

The numerical range of the above two indicators is [0, 1].

B. Simulation Scenes

We simulate five image sequences to evaluate the detection ability and stability of the proposed method under diverse backgrounds. The synthetic data are created using five real background data and simulated target data. Given the trajectories of targets, generated targets are put in the corresponding locations of each frame using Poisson fusion strategy [48]. Fig. 4 shows some selected frames of each image scene, and the 3-D projections of target areas marked by numbered rectangles. Some of them contain more than one target. Synthetic targets are studied to reveal the detective ability of the proposed method. Every row of Fig. 4 represents an image sequence. Table I shows the detailed information. The background scenes contain three sea backgrounds and two sky backgrounds, which makes the simulated data suitable for testing the adaptability of proposed method. Also, there are three scenes with noise, which can be used to test the stability against random noise.

1) Parameters: There are several important parameters in our method. By testing the synthetic data, we have chosen the appropriate values for each parameter. Suppose that the original IR image $f(x, y) \in R^{p \times q}$ and the patch model $I \in R^{m \times n}$, we choose the patch size of vertical and horizontal direction to be $\sqrt{p}$ and $\sqrt{q}$, which can adaptively change with the size

TABLE I
DETAILED INFORMATION OF FIVE SIMULATION DATA SETS

No.	Frame number	Background description	Target description (Size unit: pixels)	Noise level
1	73	Sea and mountain scene with an obvious artificial structure.	#1: Synthetic infrared small target with Gaussian distribution. Its size changes from 2×2 to 5×6. #2: Constant infrared target locating under the artificial structure with size 9×6.	40dB
2	31	Sky-cloud background.	#1: Synthetic moving target with size 6×3. #2: Slow-moving target with size 6×4.	40dB
3	31	Sky background with large and bright man-made structure.	#1: Synthetic infrared small target fused to the background. Its size changes from 2×3 to 6×6.	40dB
4	67	Similar to the background of simulated scene 1 with a brighter man-made structure. Constant clutter and water wave exist in this scene.	#1: Synthetic infrared small target moves from the left-bottom to the right of the scene.	0
5	67	Sea-sky scene with bright mountain background and a large ship, which can be considered as clutters.	#1: Synthetic target generated by Gaussian model with some blur. Its size changes from 3×3 to 4×4. #2: A constant infrared dim target with size 8×7.	0

TABLE II
DETAILED INFORMATION OF EACH SIMULATION IMAGE GROUP

	Group 1	Group 2	Group 3	Group 4	Group 5
Target amount	37	70	56	29	69
SCR range	$0 < SCR \leq 2$	$2 < SCR \leq 3$	$3 < SCR \leq 4$	$4 < SCR \leq 5$	$5 < SCR$
$\overline{SCR}$	1.55	2.43	3.46	4.51	6.76
$\overline{size}$	16.41	21.29	16.79	17.76	29.52

of original IR images. When solving the optimization problem in (6), we set $\lambda = (1/(\min(m,n))^{1/2}) * 3$, which is proved by experiments that it can best handle the SMSL problem. As to parameter μ, it changes from μ_0 to $\bar{\mu}$ in the solving procedure The initial value μ_0 partly decides the rank of coefficient α, as we can see in the soft-thresholding procedure (Step 1 in Algorithm 2). Thus, μ_0 could not be too large in order to avoid over-shrinking $rank(B)$. $\bar{\mu}$ decides the denoise effect of the algorithm. In practice, we set $\mu_0 = 0.5 * s_4$ and $\bar{\mu} = 0.05 * s_5$, where s is the singular values of I, arranged from the largest to the smallest. The choice of μ_0 and $\bar{\mu}$ refers to the parameter in [7], with proper modification based on the condition of our method.

For a specific image, these parameters may not lead to the very best performance of the proposed method; but in general, this set of parameters has excellent performance for most cases.

2) Experiments on Simulation Scenes With Different SCRs: As we can see, due to different background complexities, target properties, and noise levels, the difficulty of detecting target from each frame is different. Here we divide all synthetic targets into five groups according to their SCR. Table II shows the detailed information of each group, where $\overline{SCR}$ and $\overline{size}$ are the average SCR and average size of targets in each group, respectively.

We can see that from group 1 to group 5, the SCR of targets is increasing, which means the detection becomes easier, and we expect to acquire better detection results. Also, in a group whose targets are with same SCR, the method with high P_d has better performance.

We compare the proposed method with three baseline methods: Tophat, Maxmedian, Maxmean and two state-of-the-art methods: the kernel-based nonparametric regression method (called KR-CFAR) [6] and the IPI model method [7]. The filter sizes used in Maxmean and Maxmedian methods

are both 15×15. For KR-CFAR, we use the classic Gaussian kernel and the parameters are set according to [6]. The parameters of IPI method are the same as those of [7]. Table III shows the comparison result. In order to make the detection result of each method comparable, the F_aPR in each image group is fixed to 0.05%, with a very small error.

From Table III, we can see that in groups 2 and 3, the proposed method has higher P_d than other methods. In groups 4 and 5, the proposed method reaches 100% detection rate. The result means that using multisubspaces property of background data, the proposed method has superior detection capability in highly complex scenes. As to group 1, we found that the Maxmedian method receives the best performance. This is because that targets with low SCR always have very small size, which means these targets are point targets. The Maxmedian method and Maxmean method are designed to deal with point targets. If the target does not have a point shape, the detection performance will decrease despite the improvement of imaging quality. It can be easily seen from the test result of groups 2–4. With the increasing of SCR, the target no longer remains standard point shape. The targets in groups 2–4 have much higher SCR than group 1, but the detection performance of Maxmedian method is decreasing. Apart from Maxmedian method, the proposed method gains the best performance in group 1, with $P_d = 0.89$, which is way higher than Tophat and IPI model methods. In group 4, the proposed method and the KR-CFAR have reached $P_d = 1$ first. In group 5, all the five methods have good performance, as the $\overline{SCR} = 6.76$. Note that each group contains data from different image sequences, so the detection result can demonstrate the robustness of the proposed method for diverse background scene, target type, and noise level.

3) Experiments on Simulation Scenes: We use two evaluation metrics: the $\overline{G_{SCR}}$ and $\overline{BSF}$ to give more intuitive display of the detection performance in five simulation IR image

TABLE III
PROBABILITY OF DETECTION BY SIX METHODS OF EACH SIMULATION IMAGE GROUP

Metrics	Method	Group 1	Group 2	Group 3	Group 4	Group 5
	Tophat	0.38	0.73	0.80	0.9655	1
	Maxmedian	**0.9730**	0.9143	0.8393	0.9655	1
P_d	Maxmean	0.7027	0.9143	0.8393	0.9655	1
	KR-CFAR	0.7297	0.8571	0.8929	**1**	1
	IPI Model	0.32	0.71	0.84	0.9655	1
	Proposed method	**0.89**	**0.9429**	**0.9286**	**1**	**1**

TABLE IV
$\overline{G_{\mathrm{SCR}}}$ AND $\overline{\mathrm{BSF}}$ COMPARISON OF SIMULATION SCENES

Detective Methods	Evaluation indicators	Scene 1	Scene 2	Scene 3	Scene 4	Scene 5
Tophat	$\overline{G_{SCR}}$	1.8479	**2.2657**	1.3867	1.9707	2.0304
	$\overline{BSF}$	1.5812	5.7575	2.1180	1.8154	5.8112
Maxmedian	$\overline{G_{SCR}}$	1.8496	1.9154	1.4437	1.2777	2.7432
	$\overline{BSF}$	5.0416	**10.0850**	3.2341	**7.8426**	12.0482
Maxmean	$\overline{G_{SCR}}$	0.8399	0.8774	1.2065	0.0841	1.1520
	$\overline{BSF}$	4.0698	6.3121	2.7417	4.4126	9.7684
KR-CFAR	$\overline{G_{SCR}}$	0.8499	0.6273	1.3020	0.7338	0.6243
	$\overline{BSF}$	**5.7113**	7.0642	**3.4102**	6.9248	10.6042
IPI model	$\overline{G_{SCR}}$	**4.5713**	2.2129	**1.5272**	**2.8411**	**3.2223**
	$\overline{BSF}$	4.3523	2.8980	1.8251	2.7870	**12.1406**
Proposed method	$\overline{G_{SCR}}$	**8.6459**	**2.3061**	**1.9987**	**3.6704**	**7.9317**
	$\overline{BSF}$	**8.8486**	**17.6229**	**5.5921**	**12.8227**	**26.2327**

sequences. The results are shown in Table IV. The highest value of each evaluation indicator in each column is marked in red, while the second highest value is marked in blue. For simulated scenes with more than one target, the targets are detected simultaneously.

In Table IV, we can see that the proposed method achieves the best performance in all the test image sequences, especially with remarkable background suppression performance. Using multisubspace structure to measure the background, the complex backgrounds in five simulation scenes are well described, which leads to the highest SCR and BSF. Despite the proposed method, the IPI method has good target enhancement ability in four of five scenes. But for the background suppression performance, the Maxmedian method and the KR-CFAR method share similar performance. These two methods concentrate on estimating the background component. But they do not have the ability to gain good target enhancement performance. The drawback of low G_{SCR} limits their application in the small target detection task. On the contrary, the IPI method can enhance target in some cases. But since the single subspace assumption does not match the reality in most of the complex IR scenes, the background suppression performance of IPI model is not outstanding. The low BSF of IPI model in turn limits the performance of target enhancement. Only the proposed method can achieve remarkable performance in both the target enhancement and background suppression. This can prove the invalidity of single subspace measure in complex background.
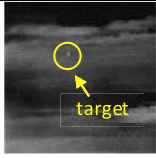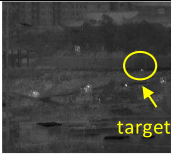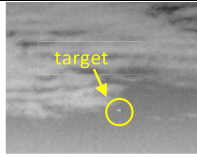
C. Real Scenes

We test the proposed method in three real image scenes. Table V shows the representative images and detailed information of each image scene. These image scenes cover different types of background including sky background and ground background. In scene 2, there are three bright targets: one is an intensity-changing IR source marked by the circle in Table V and other two are handheld IR sources, located in the left and right of marked target, respectively. These two handheld sources are very unstable, and thus we choose the marked target as a representative in scene 2 to do the following numerical analysis.

We compare the proposed method with Tophat, Maxmedian, Maxmean, KR-CFAR method, and the IPI model method. Using threshold as the postprocessing approach to segment the detection results, we can locate the potential targets and false alarms. Every row of Fig. 5 represents a test scene and the corresponding detection results of six detective methods. The methods, from the left to the right, are Tophat, Maxmedian, Maxmean, KR-CFAR, IPI model, and the proposed method. The real target in each image is marked by a yellow circle, while the false alarms are marked by blue circles. Note that for each detective method, the detection results are normalized to [0, 255]. The threshold values are the same.

In real scenes 1 and 3, the Tophat filter results and IPI model results have large size of false alarms, which indicates the drawback of simple morphology filter method and the single subspace method. In real scene 3, the false alarm level of the KR-CFAR method is also high. The Maxmedian and Maxmean method can better handle the banded cloud and floccus background, with little false detections. On the contrast, the proposed method can get good detection result without any false alarm under the given threshold value. In real scene 2, the most completed ground scene, all the six detective methods have false alarms under the given threshold value. But we can still see that the proposed method and Tophat method have minimal false alarms. And the false alarms in the proposed method have the smallest size, which means the best detection performance. The Maxmedian, Maxmean, KR-CFAR, and the IPI model have more false detections under the given threshold value in real scene 2.

TABLE V
DETAILED INFORMATION OF THREE REAL SCENES

|	Seq 1	Seq 2	Seq 3
Representative image			
Size	128×128	227×260	200×256
Target description	An airplane with small size	quite small size; changing slightly in the sequence	An airplane with changing size
Background description	sky scene with changing banded cloud	complex ground scene with buildings and trees	heavily cloudy scene with banded cloud and floccus; changing in the sequence

(a)

(b)

(c)
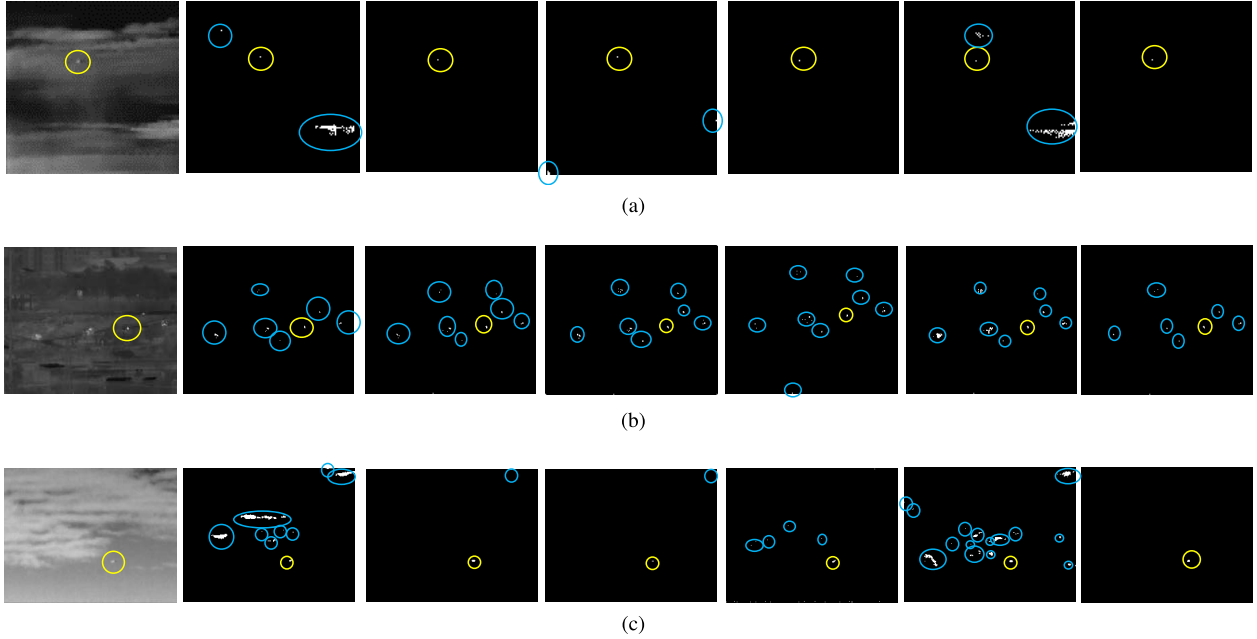
Fig. 5. Detection results of six detective methods on three real scenes. (a)–(c) Detection results of three real scenes, respectively. The segmentation threshold of each result in one scene is the same. The first column is the representative frames of three real scenes. (Left to right) Detective methods are Tophat, Maxmedian, Maxmean, KR-CFAR, IPI model, and the proposed method. Yellow circles mark the real targets, while blue circles mark the false alarms.

TABLE VI
$\overline{G_{\mathrm{SCR}}}$ AND $\overline{\mathrm{BSF}}$ COMPARISON OF REAL SCENES

Detective Methods	Evaluation indicators	Scene 1	Scene 2	Scene 3
Tophat	$\overline{G_{SCR}}$	1.0359	1.2351	2.5249
	$\overline{BSF}$	0.8987	1.2792	0.6553
Maxmedian	$\overline{G_{SCR}}$	**1.6470**	1.4545	3.0885
	$\overline{BSF}$	3.3242	2.2390	3.7496
Maxmean	$\overline{G_{SCR}}$	0.8213	0.8658	2.2295
	$\overline{BSF}$	3.0772	2.0432	3.5005
KR-CFAR	$\overline{G_{SCR}}$	1.0630	0.6740	1.5262
	$\overline{BSF}$	**4.0602**	**3.4783**	**8.2183**
IPI model	$\overline{G_{SCR}}$	0.7280	**1.8567**	**5.1038**
	$\overline{BSF}$	2.3057	1.8345	4.1004
Proposed method	$\overline{G_{SCR}}$	**2.9474**	**3.9212**	**10.7777**
	$\overline{BSF}$	**7.1976**	**4.4946**	**20.1860**

Table VI shows the average G_{SCR} and the average BSF of six methods in three real image sequences. It can be seen that the proposed method has gain the highest $\overline{G_{\mathrm{SCR}}}$ and $\overline{\mathrm{BSF}}$ in three scenes. The result confirms the reasonability

of multisubspace assumption on complex background. The effective measure of background can not only cause good background suppression performance but also gain good target enhancement performance. In addition, the IPI method has
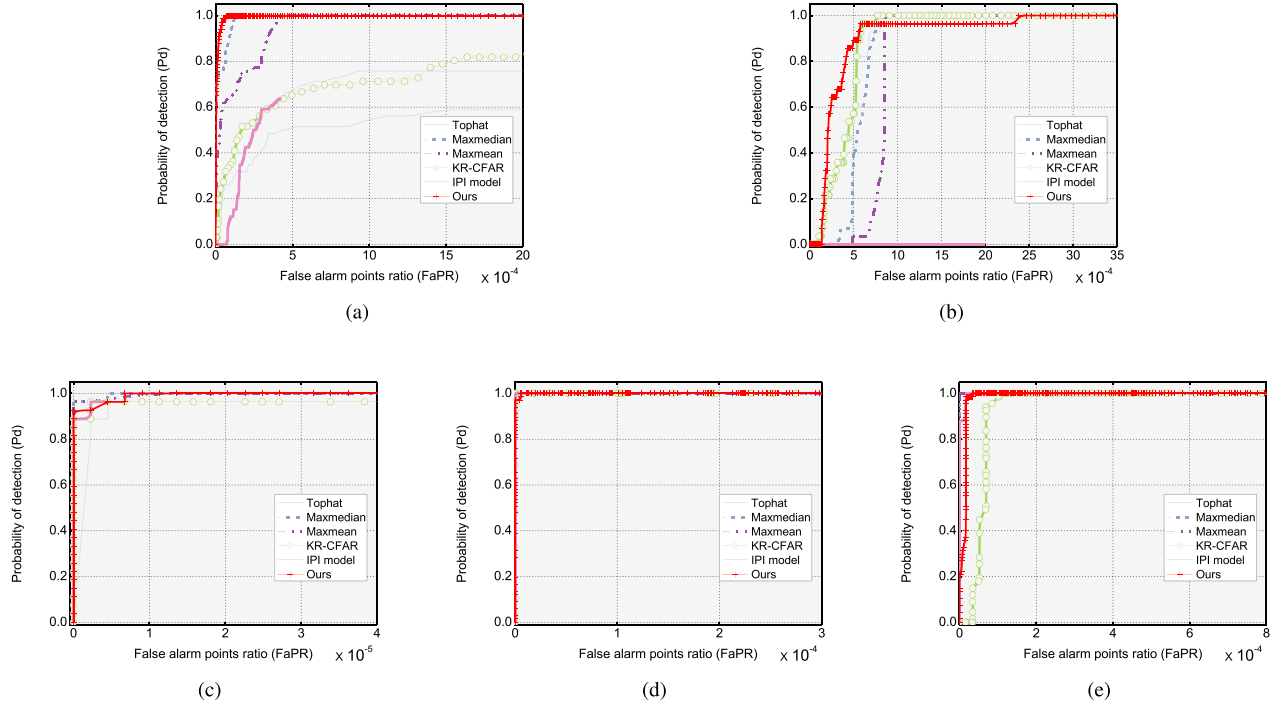
Fig. 6.   ROC curves of simulation scenes. (a)–(e) ROC curves of simulation scenes 1–5, respectively.
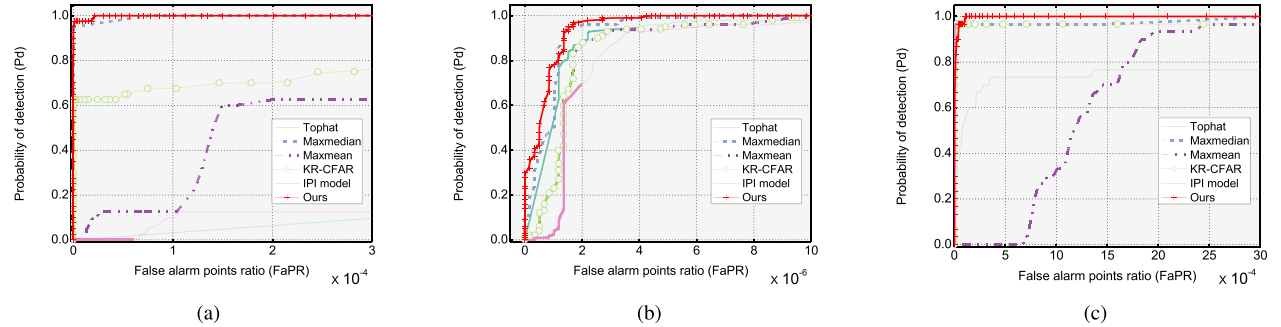

Fig. 7.   ROC curves and AUC values of real scenes. (a)–(c) ROC curves of real scenes 1–3, respectively.

the second best target enhancement performance, while the KR-CFAR method has the second best background suppression performance. The performances of other three methods are not outstanding. This result indicates that the baseline method can achieve a good performance only in particular images. The proposed method, however, considers the commonly existing multisubspace property in real cases. This leads to an accurate estimate of background region and target region.

D. ROC Curve Evaluation

To evaluate the robustness of the proposed method under different segmentation thresholds, we give the receiver operating characteristic (ROC) curves of all the test scenes. The horizontal axis of ROC curve is the F_aPR of this sequence; the vertical axis is the P_d of this sequence. Also, we calculate the area under curve (AUC) of each method. A larger AUC value means a better target detection performance in the ROC curve evaluation system. The ROC curves of simulation

scenes and real scenes are shown in Figs. 6 and 7. The AUC values are shown in Tables VII and VIII, respectively.

From Fig. 6, we can see that in simulated scene 1, the proposed method clearly has better performance than other methods. Maxmedian and Maxmean methods perform better than KR-CFAR, IPI model, and the Tophat method. In simulated scene 2, the proposed method shows a strong ability to detect more than 80% of the targets when F_aPR is very low, which is a prominent advantage. Also, the AUC value of the proposed method is the largest. The performance of IPI model is not very good in simulated scene 2, which means the single subspace assumption is not suitable in this scene. Note that these two simulation scenes contain noise (see Table I), which means the proposed method has the stability against noise. In simulated scenes 3 and 4, six detective methods all have good performance in the ROC evaluation system. In simulated scene 5, the performance of three filtering-based methods is impressive. It is because that the target in simulated scene 5

TABLE VII

AUC VALUES OF SIMULATION SCENES $(\times 10^{-3})$

	Tophat	Maxmedian	Maxmean	KR-CFAR	IPI model	The Proposed Method
Simulated Scene 1	994.0147	999.9755	999.8902	999.2132	997.8854	999.9904
Simulated Scene 2	999.4608	999.4268	999.1925	999.6011	986.1873	999.6589
Simulated Scene 3	999.9975	999.9998	999.9998	999.9836	999.9996	999.9996
Simulated Scene 4	1000.0000	1000.0000	1000.0000	1000.0000	1000.0000	1000.0000
Simulated Scene 5	1000.0000	1000.0000	1000.0000	999.9404	999.9981	999.9875

TABLE VIII

AUC VALUES OF REAL SCENES $(\times 10^{-3})$

	Tophat	Maxmedian	Maxmean	KR-CFAR	IPI model	The Proposed Method
Real Scene 1	995.6596	999.9981	999.5033	999.7880	993.1990	999.9995
Real Scene 2	999.9982	999.9991	999.9983	999.9875	999.9977	999.9994
Real Scene 3	996.8214	999.9210	998.6675	997.6527	999.9939	999.9910

TABLE IX

TIME CONSUMING COMPARISON (s)

	Tophat	Maxmedian	Maxmean	KR-CFAR	IPI model	The Proposed Method
Real Scene 1	0.03226	0.30701	0.36842	0.1987	8.08770	0.35780
Real Scene 2	0.03400	1.10212	1.30563	0.6158	51.58740	0.70730
Real Scene 3	0.03320	0.95656	1.12623	1.0272	33.30190	0.69500

is a point-like target, which is very suitable for the Tophat, Maxmedian, and Maxmean method. Also, we should note that the performance of the proposed method is the most stable in five simulation scenes, despite different backgrounds, target types, and noise levels.

In Fig. 7(a), the proposed method has the best detection performance. When the F_aPR is very low, the proposed method has the highest P_d. And the proposed method reaches $P_d = 1$ first. Compared with the KR-CFAR and the IPI model, we can see that the proposed method has greatly improved the detection performance under complex background, due to the use of multisubspace property. For real scene 2, the proposed method has better performance than other five methods, with the highest AUC value. For real scene 3, the proposed method and the IPI model both have good performance in the ROC evaluation system. But the proposed method reaches $P_d = 1$ faster than the other methods. The P_d of KR-CFAR and Maxmedian method are also high when the F_aPR is low. The performances of the Tophat and Maxmean method are not very good.

The experimental results show that the proposed method has outstanding performance compared with other five detective methods. Also, the proposed method does not rely on specific background scene, target type, or noise level. In other words, the proposed method is stable in different scenes. Thus, it is suitable for detecting small target from complex IR with theoretical guarantee, especially for heterogeneous scenes.

E. Time-Consuming Evaluation

In this section, we calculate the average time consuming in a single frame of six detective methods in three real scenes. The results are shown in Table IX. All experiments are implemented by MATLAB software on a personal computer with 8-GB memory and 4-GHz Intel Pentium processor.

From Table IX, we see that the Tophat method is the most efficient among six methods. But its detection performance is not satisfying, which can be seen from the former analysis. The

IPI model consumes too much time and does not have good performance in heterogeneous scenes. The time consuming of Maxmedian, Maxmean, KR-CFAR, and the proposed method is comparable. In scene 1, the time cost of the proposed method is between the time cost of the Maxmedian method and the Maxmean method. In scene 2, the size of image is larger than scene 1 (see Table V). The proposed method runs faster than the Maxmedian method and the Maxmean method, but a little slower than KR-CFAR. In scene 3, the proposed method runs faster than the Maxmedian, Maxmean, and KR-CFAR methods. The efficiency of the proposed method is not sensitive to the image size.

Compared with the state-of-the-art IPI model, which uses the single subspace assumption on the background data, the proposed method has a distinct advantage on computational efficiency. One reason is that the proposed method avoids the iteration of SVD; another reason is that the model used in this paper can be considered as a rearrangement of pixels in the original image, which adds no extra computational complexity. Instead, the IPI model method generates a rather large patch model to enhance the consistency of the background data.

Although all the experiments are done by applying the proposed method to single frames of each image sequence, the background recovery results of the previous frames can be used to give a background estimation of next frame. In this way, all we need to do in the next frame is subtracting the background recovery result of the previous frame. After a few frames, we use the proposed method again to get new background estimation. With an accurate estimation of background region, it will further improve the efficiency of detection algorithm.

VI. CONCLUSION

A new IR dim and small target detection method called SMSL is proposed in this paper. We focus on solving the problem of detecting small IR targets from highly heterogeneous scenes. Considering the imaging mechanism of IR images and

its application scene, we find that heterogeneous scenes are very common. It is difficult for traditional filter-based methods and single-subspace-based method to enhance targets in complex scenes. Using the multisubspace property to describe the data distribution in heterogeneous scene, a novel model is present to solve the IR small target detection problem. This model contains an orthonormal subspace learning strategy. An optimization solver combining APG and BCD strategy is proposed to solve this new model. Experiments on diverse scenes and different SCRs show that the proposed SMSL algorithm provides improved target enhancement ability as well as outstanding background suppression performance. The detection rate is found to be high with low false alarm rate, compared with four conventional methods. However, the ability to detect small point target of the SMSL method can be improved in the future. Also, the subspace learning results of the SMSL method can be used to do background classification. Application scenario-specific detection algorithms will also be researched, which can further improve the detection ability of IR systems.

ACKNOWLEDGMENT

REFERENCES

[1] K. Luo, "Space-based infrared sensor scheduling with high uncertainty: Issues and challenges," *Syst. Eng.*, vol. 18, no. 1, pp. 102–113, Jan. 2015.

[2] J. A. Dawson and C. T. Bankston, "Space debris characterization using thermal imaging systems," in *Proc. Adv. Maui Opt. Space Surveill. Technol. Conf.*, Wailea, HI, USA, Sep. 2010, pp. 1–11, paper E40.

[3] L. Li, H. Li, T. Li, and F. Gao, "Infrared small target detection in compressive domain," *Electron. Lett.*, vol. 50, no. 7, pp. 510–512, Mar. 2014.

[4] V. T. Tom, T. Peli, M. Leung, and J. E. Bondaryk, "Morphology-based algorithm for point target detection in infrared backgrounds," *Proc. SPIE*, vol. 1954, pp. 2–11, Oct. 1993.

[5] S. D. Deshpande, M. H. Er, R. Venkateswarlu, and P. Chan, "Max-mean and max-median filters for detection of small targets," *Proc. SPIE*, vol. 3809, pp. 74–83, Oct. 1999.

[6] Y. Gu, C. Wang, B. Liu, and Y. Zhang, "A kernel-based nonparametric regression method for clutter removal in infrared small-target detection applications," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 3, pp. 469–473, Jul. 2010.

[7] C. Q. Gao, D. Meng, Y. Yang, Y. Wang, X. Zhou, and A. G. Hauptmann, "Infrared patch-image model for small target detection in a single image," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4996–5009, Dec. 2013.

[8] M. M. Hadhoud and D. W. Thomas, "The two-dimensional adaptive LMS (TDLMS) algorithm," *IEEE Trans. Circuits Syst.*, vol. 35, no. 5, pp. 485–494, May 1988.

[9] G. Davidson and H. D. Griffiths, "Wavelet detection scheme for small targets in sea clutter," *Electron. Lett.*, vol. 38, no. 19, pp. 1128–1130, Sep. 2002.

[10] Z. Peng, Q. Zhang, J. Wang, and Q. P. Zhang, "Dim target detection based on nonlinear multifeature fusion by Karhunen–Loeve transform," *Opt. Eng.*, vol. 43, no. 12, pp. 2954–2958, Dec. 2004.

[11] Y. Cao, R. M. Liu, and J. Yang, "Infrared small target detection using PPCA," *Int. J. Infr. Millim. Waves*, vol. 29, no. 4, pp. 385–395, Apr. 2008.

[12] C. L. P. Chen, H. Li, Y. Wei, T. Xia, and Y. Y. Tang, "A local contrast method for small infrared target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 574–581, Jan. 2014.

[13] J. Han, Y. Ma, B. Zhou, F. Fan, K. Liang, and Y. Fang, "A robust infrared small target detection algorithm based on human visual system," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 12, pp. 2168–2172, Dec. 2014.

[14] H. Deng, X. Sun, M. Liu, C. Ye, and X. Zhou, "Small infrared target detection based on weighted local difference measure," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 4204–4214, Jul. 2016.

[15] L. Dong, B. Wang, M. Zhao, and W. Xu, "Robust infrared maritime target detection based on visual attention and spatiotemporal filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 3037–3050, May 2017.

[16] I. S. Reed, R. M. Gagliardi, and L. B. Stotts, "Optical moving target detection with 3-D matched filtering," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 327–336, Jul. 1988.

[17] R. W. Fries, "Three dimensional matched filtering," *Proc. SPIE*, vol. 1050, pp. 19–27, Jun. 1989.

[18] Y. Barniv, "Dynamic programming solution for detecting dim moving targets," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-21, no. 1, pp. 144–156, Jan. 1985.

[19] L. A. Johnston and V. Krishnamurthy, "Performance analysis of a dynamic programming track before detect algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 1, pp. 228–242, Jan. 2002.

[20] Z. Peng, Q. Zhang, and A. Guan, "Extended target tracking using projection curves and matching pel count," *Opt. Eng.*, vol. 46, no. 6, pp. 066401-1–066401-6, Jun. 2007.

[21] S. M. Tonissen and Y. Bar-Shalom, "Maximum likelihood track-before-detect with fluctuating target amplitude," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 3, pp. 796–809, Jul. 1998.

[22] J. Silverman, J. M. Mooney, and C. E. Caefer, "Tracking point targets in cloud clutter," *Proc. SPIE*, vol. 3110, pp. 496–507, Sep. 1997.

[23] S. D. Blostein and T. S. Huang, "Detection of small moving objects in image sequences using multistage hypothesis testing," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, New York, NY, USA, Apr. 1988, pp. 1068–1071.

[24] A. P. Tzannes and D. H. Brooks, "Point target detection in IR image sequences: A hypothesis-testing approach based on target and clutter temporal profile modeling," *Opt. Eng.*, vol. 39, no. 8, pp. 2270–2278, Aug. 2000.

[25] Y. He, M. Li, J. Zhang, and Q. An, "Small infrared target detection based on low-rank and sparse representation," *Infr. Phys. Technol.*, vol. 68, pp. 98–109, Jan. 2015.

[26] Z. Li *et al.*, "Dim moving target detection algorithm based on spatio-temporal classification sparse representation," *Infr. Phys. Technol.*, vol. 67, pp. 273–282, Nov. 2014.

[27] C. Yang, H. Liu, S. Liao, and S. Wang, "Small target detection in infrared video sequence using robust dictionary learning," *Infr. Phys. Technol.*, vol. 68, pp. 1–9, Jan. 2015.

[28] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, p. 11, May 2009.

[29] Z. H. Zhou, X. Li, J. Wright, E. Candès, and Y. Ma, "Stable principal component pursuit," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, USA, Jun. 2010, pp. 1518–1522.

[30] Z. Gao, L.-F. Cheong, and Y.-X. Wang, "Block-sparse RPCA for salient motion detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 10, pp. 1975–1987, Oct. 2014.

[31] R. Ibrahim, J. Alirezaie, and P. Babyn, "Pixel level jointed sparse representation with RPCA image fusion algorithm," in *Proc. 38th Int. Conf. Telecommun. Signal Process. (TSP)*, Jul. 2015, pp. 592–595.

[32] Z. Fu, X. Wang, J. Xu, N. Zhou, and Y. Zhao, "Infrared and visible images fusion based on RPCA and NSCT," *Infr. Phys. Technol.*, vol. 77, pp. 114–123, Jul. 2016.

[33] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J. Y. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3658–3668, Jul. 2015.

[34] R. Hang, Q. Liu, H. Song, and Y. Sun, "Matrix-based discriminant subspace ensemble for hyperspectral image spatial–spectral feature fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 2, pp. 783–794, Feb. 2016.

[35] X. Zhang, S. Hao, C. Xu, X. Qian, M. Wang, and J. Jiang, "Image classification based on low-rank matrix recovery and Naive Bayes collaborative representation," *Neurocomputing*, vol. 169, pp. 110–118, Dec. 2015.

[36] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.

[37] S. Tierney, J. Gao, and Y. Guo, "Subspace clustering for sequential data," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Columbus, OH, USA, Jun. 2014, pp. 1019–1026.

[38] F. de Morsier, M. Borgeaud, V. Gass, J.-P. Thiran, and D. Tuia, "Kernel low-rank and sparse graph for unsupervised and semi-supervised classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3410–3420, Jun. 2016.

[39] S. Xiao, M. Tan, and D. Xu, "Weighted block-sparse low rank representation for face clustering in videos," in *Proc. Eur. Conf. Comput. Vis.*, Zürich, Switzerland, 2014, pp. 123–138.

[40] P. Ji, Y. Zhong, H. Li, and M. Salzmann, "Null space clustering with applications to motion segmentation and face clustering," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Paris, France, Oct. 2014, pp. 283–287.

[41] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. San Diego, CA, USA, Jun. 2005, pp. 60–65.

[42] X. Shu, F. Porikli, and N. Ahuja, "Robust orthonormal subspace learning: Efficient recovery of corrupted low-rank matrices," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Columbus, OH, USA, Jun. 2014, pp. 3874–3881.

[43] E. J. Candès and Y. Plan, "Matrix completion with noise," *Proc. IEEE*, vol. 98, no. 6, pp. 925–936, Jun. 2010.

[44] A. Ganesh, Z. Lin, J. Wright, L. Wu, M. Chen, and Y. Ma, "Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix," in *Proc. Comput. Adv. Multi-Sensor Adapt. Process.*, vol. 61. Aruba, Dutch Antilles, Dec. 2009, pp. 213–216.

[45] Z. Lin, M. Chen, and Y. Ma. (2010). "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices." [Online]. Available: https://arxiv.org/abs/1009.5055

[46] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM J. Imag. Sci.*, vol. 6, no. 3, pp. 1758–1789, May 2013.

[47] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *J. Optim. Theory Appl.*, vol. 109, no. 3, pp. 475–494, Jun. 2001.

[48] P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," in *Proc. ACM SIGGRAPH Papers*, San Diego, CA, USA, Jul. 2003, pp. 313–318.

Zhenming Peng (M'06) received the Ph.D. degree in geodetection and information technology from the Chengdu University of Technology, Chengdu, China, in 2001.

From 2001 to 2003, he was a Post-Doctoral Researcher with the Institute of Optics and Electronics, Chinese Academy of Sciences, Beijing, China. He is currently a Professor with the University of Electronic Science and Technology of China, Chengdu. His research interests include image processing, signal processing, and target recognition and tracking.

Prof. Peng is a member of Optical Society of America, the Optical Engineering Society, and the Aerospace Society of China.

Dehui Kong received the bachelor's degree in electronic science and technology from the University of Electronic Science and Technology of China, Chengdu, China, in 2011, where he is currently pursuing the Ph.D. degree in signal and information processing.

His research interests include compressed sensing, sparse representation, and inverse problems.

Xiaoyang Wang received the B.Sc. degree in electronic science and technology from the University of Electronic Science and Technology of China, Chengdu, China, in 2013, where she is currently pursuing the Ph.D. degree in signal and information processing.

She has been a Visiting Student with the Department of Computer Science, University of Bristol, Bristol, U.K., since 2017. Her research interests include computer vision and machine learning.

Yanmin He received the Ph.D. degree from the School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu, China, in 2011.

She is currently an Assistant Professor with the School of Optoelectronic Information, University of Electronic Science and Technology of China. She has authored more than 20 peer-reviewed papers, and is the Inventor or Co-Inventor of ten patents. Her research interests include signal processing and pattern recognition.