

Infrared Patch-Image Model for Small Target Detection in a Single Image

Chenqiang Gao, Deyu Meng, Yi Yang, Yongtao Wang, Xiaofang Zhou, *Senior Member, IEEE*, and Alexander G. Hauptmann

Abstract—The robust detection of small targets is one of the key techniques in infrared search and tracking applications. A novel small target detection method in a single infrared image is proposed in this paper. Initially, the traditional infrared image model is generalized to a new infrared patch-image model using local patch construction. Then, because of the non-local self-correlation property of the infrared background image, based on the new model small target detection is formulated as an optimization problem of recovering low-rank and sparse matrices, which is effectively solved using stable principle component pursuit. Finally, a simple adaptive segmentation method is used to segment the target image and the segmentation result can be refined by post-processing. Extensive synthetic and real data experiments show that under different clutter backgrounds the proposed method not only works more stably for different target sizes and signal-to-clutter ratio values, but also has better detection performance compared with conventional baseline methods.

Index Terms—Infrared image, small target detection, low-rank matrix recovery.

I. INTRODUCTION

INFRARED small target detection is one of the key techniques in infrared search and track (IRST) systems. The performance of the whole IRST system depends on the accuracy of detection results. When there exists heavy noise and clutter such as cloud clutter and sea clutter, small targets are usually buried in a complex background with low signal-to-clutter ratio (SCR). Moreover, small targets have no concrete shape and texture because of the long imaging

distance. Therefore, small target detection in complex infrared background is considered a difficult and challenging problem. Although many research efforts have been focused on this area in past decades, it still remains an open problem.

For static background or consistent targets in adjacent frames, given some prior knowledge on targets, sequential detection methods can perform well, such as 3D matched filtering [1], improved three dimensional filter [2], multiscale adaptive matched filter [3], and so on, since spatial-temporal information of targets can be used. However, on one hand, in practical applications where the IRST systems need to search and track the targets, the imaging backgrounds usually change fast since many targets, such as anti-ship missiles, high speed airplanes and so on, have very high velocities. On the other hand, the applications with fast moving sensor platforms such as airplane based and missile based IRST systems also lead to fast changing backgrounds even though the interest targets may remain static most of the time. In addition, these applications usually make motion traces of the targets inconsistent. As a result, the performance of sequential detection methods could degrade rapidly. Therefore, detecting small targets in a single image is often necessary in such applications, and this problem is attracting much attention in recent years [4]–[6].

Generally, the infrared image model can be formulated as follows [6]:

$$f_D(x, y) = f_T(x, y) + f_B(x, y) + f_N(x, y), \quad (1)$$

where f_D , f_T , f_B , f_N , and (x, y) are the original infrared image, the target image, the background image, the random noise image and the pixel location, respectively. Depending on whether the focus is on the target image f_T , the background image f_B or both of them leads to different methods.

Reference [7] pointed out that small target shapes were roughly circular, and there were no anisotropy and prevailing orientations. Some methods [8], [9] modeled a small target by using a 2D Gaussian function:

$$s(x, y) = \gamma e^{-\frac{1}{2}\left(\left(\frac{x}{\sigma_x}\right)^2 + \left(\frac{y}{\sigma_y}\right)^2\right)}, \quad (2)$$

where the target is completely defined by its peak height γ , horizontal and vertical extent parameters σ_x and σ_y , respectively. Actually, the target may have a “flat top” shape like Fig. 1(d) or sometimes its shape is different from a single Gaussian shape like Fig. 1(b). In addition, the target size may vary from 2×2 to more than 10×10 pixels and its brightness also varies from dim to bright due to different types of targets, imaging distances, environments and so on (Please see Fig. 1).

Manuscript received March 18, 2013; revised August 2, 2013; accepted August 28, 2013. Date of publication September 11, 2013; date of current version October 2, 2013. This work was supported in part by the National Natural Science Foundation of China under Grants 61102131, 61373114, and 11131006, in part by the National Science Foundation Project of Chongqing under Grant CSTC 2010BB2411, and in part by the Chongqing Distinguished Youth Foundation under Grant CSTC 2011jjj40002. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Xin Li.

C. Gao is with the Chongqing Key Laboratory of Signal and Information Processing, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: gaochenqiang@gmail.com).

D. Meng is with the Institute for Information and System Sciences and the Ministry of Education Key Laboratory for Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: dymeng@mail.xjtu.edu.cn).

Y. Yang and X. Zhou are with the School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane QLD 4072, Australia (e-mail: yi.yang@itee.uq.edu.au; zxf@itee.uq.edu.au).

Y. Wang is with the Institute of Computer Science and Technology, Peking University, Beijing 100871, China (e-mail: wyt@pku.edu.cn).

A. Hauptmann is with the School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213 USA (e-mail: alex@cs.cmu.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2013.2281420

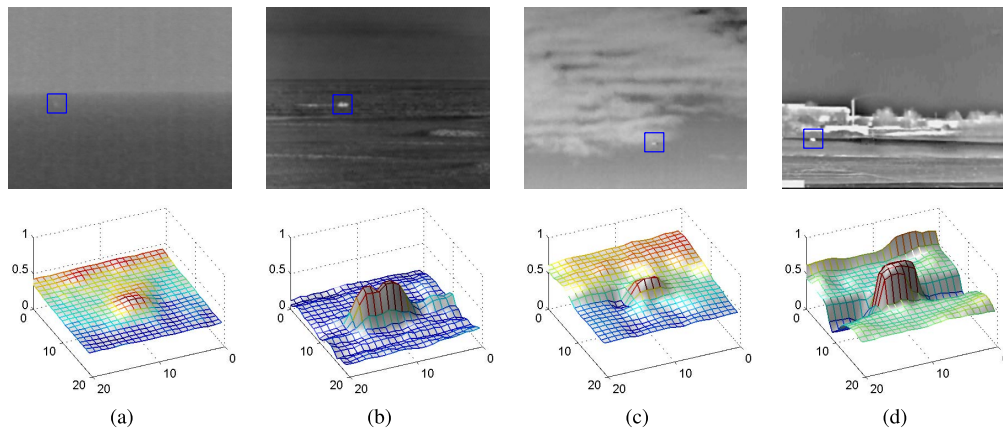


Fig. 1. Representative targets (upper) and the corresponding 3D surfaces (lower) in different backgrounds (normalized). (a) A dim small ship target in sea-sky background. (b) A bright ship target in sea-sky background. (c) A dim aeroplane target in sky cloud background. (d) A bright vehicle target in sky-ground background.

Thus sometimes it may be invalid to directly model the small target using a 2D Gaussian function.

Some conventional small target detection methods, such as TopHat filtering [10], [11] and so on, focused on background image f_B and used filters to predict the background image in order to suppress the background clutter. These methods can usually achieve good performance in applications. However, they are sensitive to noise and usually could not work stably when the target size varies within a somewhat large range [4].

Generally speaking, small target detection methods are usually based on some assumptions on the target, the background or both of them. The suitability of the assumption would determine the robustness of the corresponding method in applications.

In this paper, a novel small target detection technique in a single infrared image is presented. We generalize the traditional infrared image model (1) to a new infrared patch-image (IPI) model (4) by using local patch construction. All images in (1) are transformed into corresponding patch-images in (4) and the traditional model corresponds to a specific case of the new model. In the new model, we assume that the target patch-image is a sparse matrix and the background patch-image would be a low-rank matrix. The small target detection task is then transformed into an optimization problem of recovering the low-rank and sparse matrices. Our motivation is mainly based on the observation that the local patches in distant regions in an infrared background image are usually approximately linearly correlated with each other. Thus the background patch-image constructed from the local patches tends to be a low-rank matrix.

Our method has two advantages. The first is that our assumption fits the reality well. The assumption for the target patch-image is valid for almost all cases because the target size is very small with respect to the original image size. The assumption for the background patch-image would be commonly valid as mentioned above, which will be further discussed in detail in Section IV-B. The second advantage is that the proposed IPI model can be solved effectively by the low-rank matrix recovery techniques [12]–[15]. These recently proposed techniques can guarantee to exactly recover the

target patch-image and the background patch-image without information about the rank of the background patch-image and the support of the target patch-image. We use synthetic and real data to test the proposed method and extensive experimental results show that under different clutter backgrounds our method can not only work more stably for different target sizes and SCR values, but also has better detection performance compared to conventional baseline methods.

The remainder of this paper is organized as follows. In Section II we review related work. In Section III, we present construction and reconstruction of a patch-image and then propose our new IPI model. In Section IV, we explain the small target detection based on IPI model in detail. In Section V, we give extensive synthetic and real data experimental results and discussions. Conclusions and perspectives are given in Section VI.

II. RELATED WORK

There exists a number of approaches for detecting small targets in infrared images. These methods can be generally categorized into two groups: the sequential detection methods and the single-frame detection methods. The sequential detection methods usually process a number of frames to estimate targets. One of early classical method is the 3D matched (directional) filtering [1] which is a so-called track-before-detection (TBD) technique. This method can detect moving constant-velocity targets, where the knowledge of the target shape and velocity is required. Parallel banks of directional filters using 3D fast Fourier transform was suggested in [16]. In order to further improve the ability of detecting weaker targets, a new 3D double directional filter was presented in [2]. Compared to the conventional 3D matched filter, this method can further increase the target energy accumulation ability. Recently, an improved 3D directional filter method was proposed in [17]. This method firstly employed a dual-diffusion partial differential equation (DFPDE) to pre-whiten each image and then used a wide-to-exact search technique to improve the speed of filtering. The wide-to-exact search technique was also used in [18]. While the 3D directional filter method can be considered as a path-based statistic, a

pixel-based statistic for the analysis of multiframe target detection was proposed in [19]. This method needs to assume a known knowledge of the maximum target velocity.

Instead of considering the small target detection problem in a 3D spatial-temporal space as the above sequential detection methods, the idea of the temporal profile method proposed in [20] considered this problem in a 1D signal space. This method can reduce the computational complexity dramatically. Based on this line of research, many improved methods were prompted. Temporal hypothesis testing was used to distinguish target temporal profile from background temporal profile in [21]. The variance of the temporal profile was used to discriminate the target and background in [22]. The connecting line of the stagnation points (CLSP) of the temporal profile was utilized well to design the small moving target detection algorithm in [23]. The temporal profile was combined with spatial bilateral filter (BF) to detect small targets in [24].

Furthermore, some other sequential detection methods firstly focused on suppressing the background clutter or enhancing the small targets in a single image and then used multiframe accumulation, autocorrelation or other techniques to suppress random noise or delete false alarms. The performance of these methods greatly relies on the processed result of each frame. A method of small infrared target fusion detection based on support vector machines (SVM) in the wavelet domain was presented in [25]. This method applied SVM to extract feature images in wavelet domain for each frame and then used a fusion strategy for consecutive frames to enhance small targets. An adaptive anisotropic filter based on a modified partial differential equation (AFMPDE) is proposed in [18]. The key step of this method is background clutter suppression and target enhancement for each frame.

The sequential methods usually depends on some certain assumptions of consistent information of targets and background between frames, as well as the prior knowledge of target shape and velocity. These preset assumptions and prior knowledge could hardly be attained in applications. Thus, single-frame based small target detection has been attracting much attention [4]–[6], [26]–[28].

The typical single-frame detection methods are listed as follows. The method presented in [27] investigated the performance of two dimensional least mean square (TDLMS) adaptive filters as whitening filters for the small target detection. Along this line, some new related methods have been presented, such as two-dimensional block diagonal LMS adaptive filtering [29], improved 2D adaptive lattice algorithm (2D AL) [30], TDLMS filter based on neighborhood analysis [31] and so on.

Based on the local difference between the background and the small target, some single-frame detection methods have been presented. Probabilistic PCA (PPCA) was used to detect small targets in [31] and this method is similar to the face detection scheme using PCA. A new technique based on empirical mode decomposition (EMD) and modified local entropy was proposed for small target detection in [32]. The kernel-based nonparametric regression technique was applied to background prediction and clutter removal in [6]. Motivated

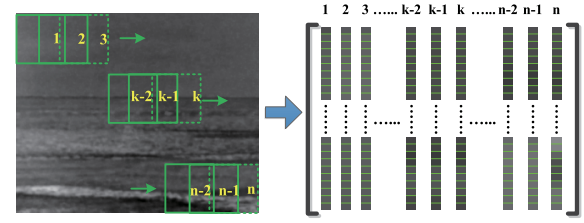


Fig. 2. Constructing the patch-image from a original image by using image patches. (left) An original image. (right) The patch-image.

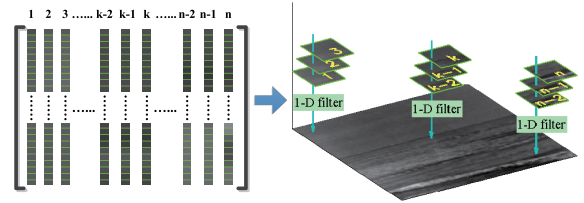


Fig. 3. Reconstructing the image from a patch-image. (left) The patch-image. (right) The reconstructed image.

by the robust properties of the human visual system (HVS), a scale invariant small target detection method for incoming target detection was presented in [4]. A sparse ring representation model was proposed to describe small targets in [33].

More commonly used single-frame detection methods including TopHat filtering [10], [11] and so on, focused on background clutter suppression. Based on TopHat filtering, some related methods have been presented. In [34], neural network and genetic algorithm were used to develop optimal TopHat filtering parameters. According to the property of the target region, a modified Top-hat transformation based on contour structuring element was proposed to enhance and detect the infrared small target in [35]. The relationships of different modified top-hat transformations were analyzed in [5].

III. INFRARED PATCH-IMAGE MODEL

A. Construction and Reconstruction of a Patch-Image

Fig. 2 shows the construction of a patch-image from an image. Firstly, a series of local image patches can be obtained by using a sliding window from left and top to right and down in an image. Then we vectorize each patch as a column of a new image (matrix). In this way, we can convert an image to a new image which is called a *patch-image* in this paper. Obviously the size of the patch-image depends not only on the size of the original image, but also on both the size of the sliding window and the horizontal and vertical sliding steps. For some special cases, such as when the size of the sliding window is the same as the original image or the sliding window equals the column of the original image and the horizontal sliding step is 1, the patch-image is the same as the original image. How to choose the sliding steps and the sliding window, namely the patch size, depends on the specific applications which will be discussed in Section V-B.2.

Fig. 3 shows the reconstruction of an image from a patch-image. Because the local patches usually overlap each other,

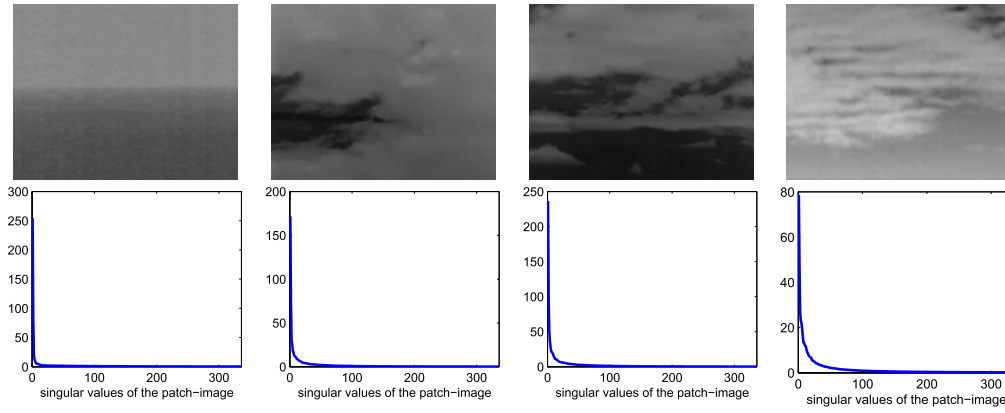


Fig. 4. The low-rank property of the background patch-images. The first row are four representative background images and the second one are the singular values of the corresponding background patch-images.

a pixel location in reconstructed image would correspond to several values from different patches. Thus, we should define a 1D filter function to determine the pixel value v :

$$v = f(\mathbf{x}), \quad (3)$$

where $v \in \mathbf{R}$, $\mathbf{x} \in \mathbf{R}^p$ is a vector containing the corresponding values from p patches. The definition of the 1D filter function f depends on the practical applications and we can define it as $v = \text{median}(\mathbf{x})$, $v = \text{mean}(\mathbf{x})$, $v = \text{max}(\mathbf{x})$, or $v = \text{min}(\mathbf{x})$ and so on.

B. Infrared Patch-Image Model

Given the original image f_D , the background image f_B , the target image f_T and the noise image f_N in (1), we can construct the corresponding patch-images D , B , T and N , respectively. Then the traditional infrared image model (1) is converted to a new one called *Infrared Patch-image (IPI)* model:

$$D = B + T + N. \quad (4)$$

Obviously, the traditional infrared image model (1) is a specific case of our new IPI model (4) under certain construction parameter configurations as mentioned before. In Section IV, we will see that our new IPI model can make use of the current research achievements of the low-rank matrix recovery that are well suited to the problem of small target detection.

IV. SMALL TARGET DETECTION BASED ON IPI MODEL

In this section, we first discuss the properties of the proposed IPI model in detail. Then we formulate the small target detection task as an optimization problem of recovering the low-rank and sparse matrices, and the corresponding algorithm is explained in detail. Furthermore, we present the whole method of small target detection. Finally, we analyze the computational complexity of the proposed method.

A. Target Patch-Image T

In practical applications, a small target usually keeps changing all the time. The brightness may vary from dim to bright

(as shown in Fig. 1) and its size may vary from 2×2 to more than 10×10 (in pixel). However, it is small with respect to the whole image, thus the target image f_T can be considered as a sparse matrix, which makes the corresponding target patch-image T still be a sparse matrix. That is:

$$\|T\|_0 < k, \quad (5)$$

where $\|\cdot\|_0$ denotes the ℓ_0 -norm which counts the number of nonzero entries, and k is determined by the number of small targets and their sizes (support), and apparently $k \ll m \times n$ ($m \times n$ is the size of T), which means that most of the entries of the matrix T are zero. Besides this sparseness property, we do not make any additional assumption on the target image.

B. Background Patch-Image B

Due to atmospheric refraction, dispersion, optical defocusing, lens aberration, diffraction, deformation of mirror, and detector tilt, the original infrared background image f_B tends to be slightly blurred [7], [36] and many local patches are approximately linearly correlated with each other even though the pixel distance between two patches may be large in an image. This property of non-local self-correlation exists commonly in infrared background images.

The first row of Fig. 4 shows four representative infrared background images with the same size of 256×200 . We construct four corresponding patch-images, where the patch size is 50×50 and the horizontal and vertical sliding steps are 10. As a result, the sizes of all corresponding patch-images are 2500×336 . The singular values of corresponding patch-images are shown in the second row of Fig. 4, from which we can see that the singular values of all four patch-images rapidly decrease to zero though their original background images are different.

Based on the above discussion, we can consider the background patch-image B as a low-rank matrix. That is:

$$\text{rank}(B) \leq r, \quad (6)$$

where r is a constant. Intrinsically, r constrains the complexity of the background image and the value of r is larger for the complex background than for the uniform background.

Similar to non-local self-correlation, the property of non-local self-similarity also exists commonly in natural images, which motivates many well-known non-local methods, such as the non-local method for texture synthesis [37], the non-local algorithm for image denoising [38]–[40], the non-local sparse models for image restoration [41], and so on.

It should be noted that we can also employ a recently proposed more general low-rank assumption that all background patches come from a mixture of low-rank subspace clusters while not only one [42]. When the small target is located in highly heterogeneous background, this multi-subspace-cluster assumption might be possibly more appropriate. In our context, however, the background is generally with a non-local self-correlation configuration, as easily seen in Fig. 4. In such cases, one low-rank subspace cluster is generally enough to model the background situations. We thus only employ one low-rank subspace assumption in our model to make it with a more concise form.

C. Noise Patch-Image N

In this paper, we just assume that the random noise is i.i.d. and $\|N\|_F \leq \delta$ for some $\delta > 0$. Thus we have:

$$\|D - B - T\|_F \leq \delta, \quad (7)$$

where $\|\cdot\|_F$ is the Frobenius norm (i.e. $\|X\|_F = \sqrt{\sum_{ij} X_{ij}^2}$).

Although the parameters k , r and δ in (5), (6) and (7) vary for different infrared images, it is not necessary to compute their values directly, which will be discussed in the following Section.

D. Formulation of Small Target Detection and Solution

In order to detect small targets in a single image f_D , we need to obtain the target image f_T . Because f_T can be reconstructed from the target patch-image T , it is a key step to effectively estimate the target patch-image T .

Firstly, we assume that an infrared image contains no noise. Thus (4) is changed into $D = B + T$. As afore discussed, the background patch-image B is a low-rank matrix and the target patch-image T is a sparse matrix. Thus the small target detection task is intrinsically a typical problem of recovering a low-rank component and a sparse component from a data matrix. This problem can be effectively solved via Principal Component Pursuit (PCP) to solve the following convex optimization problem [12]:

$$\min_{B,T} \|B\|_* + \lambda \|T\|_1 \quad \text{s.t.} \quad D = B + T, \quad (8)$$

where $\|\cdot\|_*$ is the nuclear norm of a matrix (i.e. the sum of singular values), $\|\cdot\|_1$ is the ℓ_1 -norm (i.e. $\|X\|_1 = \sum_{ij} |X_{ij}|$), and λ is a positive weight constant. Here, $\|B\|_*$, $\|T\|_1$ replace $\text{rank}(B)$ in (6) and $\|T\|_0$ in (5), respectively, for tractable computation.

It has been shown in [12] that under rather broad conditions (some entries of T may be *arbitrarily large* and no other information about the rank of B and/or the support of T is given), PCP (8) can accurately and efficiently recovers B and T . This will bring our method at least two advantages:

(i) The parameters k , r in (5), (6) do not need to be estimated in advance according to different images, which is not a trivial problem; (ii) The robustness of small target detection would be guaranteed since the above convex optimization can be suitable not only for variable sizes (support of T) and variable brightness (entries of T) of small targets, but also for different clutter backgrounds (rank of B).

Now we return to the IPI model (4) considering the situation of noise in an image. It has been shown in [13] that under the *same conditions* as PCP (8), B and T can be stably estimated when $\|N\|_F \leq \delta$ for some $\delta > 0$ and the PCP is relaxed to the following version called Stable Principal Component Pursuit:

$$\min_{B,T} \|B\|_* + \lambda \|T\|_1 \quad \text{s.t.} \quad \|D - B - T\|_F \leq \delta. \quad (9)$$

Solving the above optimization (9) can estimate the target patch-image T and the background patch-image B , simultaneously, from the noisy image. Although the main objective is usually to estimate the target image, for some applications it is still important to estimate the background image which can be reconstructed from the background patch-image B . For example, we can use the complexity of the background image to automatically evaluate the reliability of detection results. In addition, a good background estimation can be used to image registration for the motion imaging system.

Instead of directly solving (9), the following dual problem is solved:

$$\min_{B,T} \|B\|_* + \lambda \|T\|_1 + \frac{1}{2\mu} \|D - B - T\|_F^2, \quad (10)$$

where μ is a positive weight parameter and it is well established that (9) is equivalent to (10) for some value $\mu(\delta)$ [13]. However, μ is varied during solving the above optimization in this paper. This continuation technique [43] can greatly reduce computation and the experiments show it works well.

The optimization problem (10) is convex and can be solved by applying the *Accelerated Proximal Gradient (APG)* approach proposed in [43]. The solution via APG is described in Algorithm 1 (for details please see [43]), where V^T is the transpose of the matrix V and $\mathcal{S}_\varepsilon[\cdot]$ is a soft-thresholding operator [44]:

$$\mathcal{S}_\varepsilon[x] = \begin{cases} x - \varepsilon, & \text{if } x > \varepsilon, \\ x + \varepsilon, & \text{if } x < -\varepsilon, \\ 0, & \text{otherwise,} \end{cases} \quad (11)$$

where $x \in \mathbb{R}$ and $\varepsilon > 0$. For the cases of matrices in Algorithm 1, this operator is extended by applying it element-wise.

E. The Whole Method of Small Target Detection

Fig. 5 shows the whole method of small target detection proposed in this paper. First, the patch-image D is constructed from the original infrared image f_D which is obtained from an image sequence. Second, Algorithm 1 is applied to the patch-image D to estimate simultaneously the low-rank background patch-image B and the sparse target patch-image T . Third, we reconstruct the background image f_B and target image f_T from the patch-images B and T , respectively.

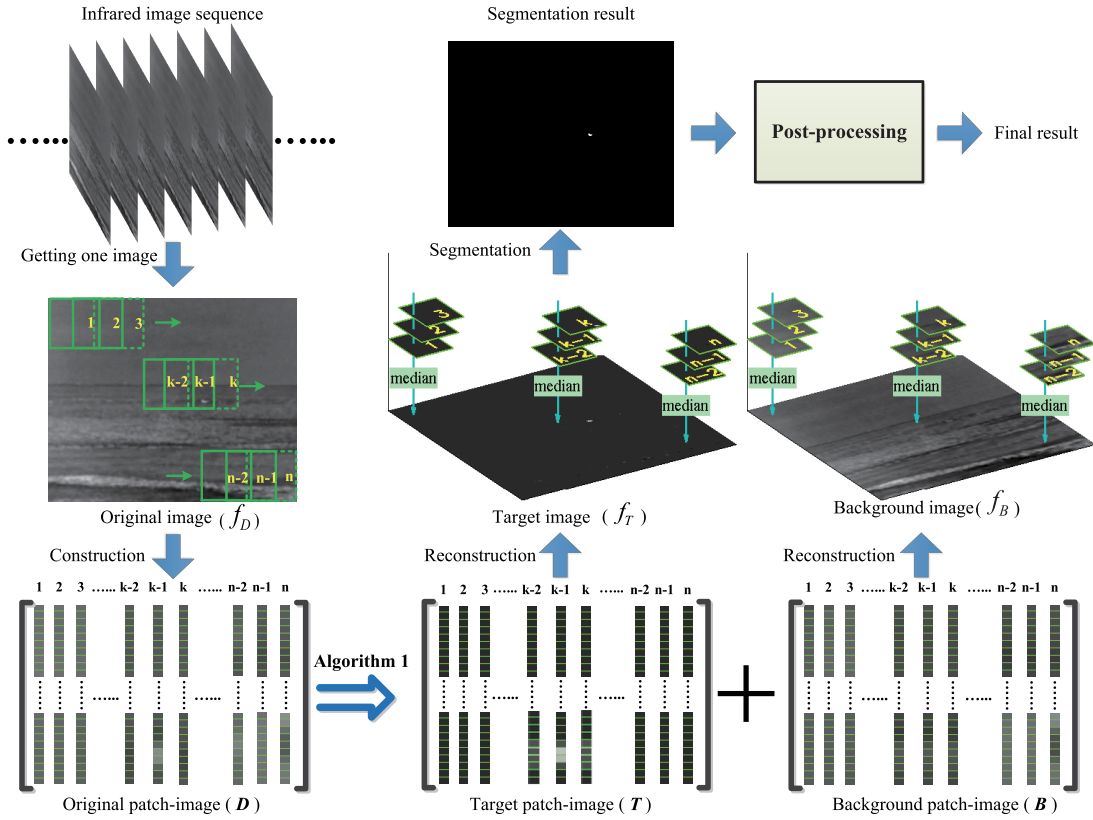


Fig. 5. The overview of the proposed method in this paper.

Algorithm 1 Solution via Accelerated Proximal Gradient

Input: Infrared patch-image matrix $D \in R^{m \times n}$, λ .

1: $B_0 = B_{-1} = 0; T_0 = T_{-1} = 0; a_0 = a_{-1} = 1; \mu_0 > 0; \bar{\mu} > 0; \eta < 1$.

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

3: $Y_k^B = B_k + \frac{a_{k-1}-1}{a_k}(B_k - B_{k-1}), Y_k^T = T_k + \frac{a_{k-1}-1}{a_k}(T_k - T_{k-1})$.

4: $G_k^B = Y_k^B - \frac{1}{2}(Y_k^B + Y_k^T - D)$.

5: $(U, S, V) = \text{svd}(G_k^B), B_{k+1} = US_{\frac{\mu_k}{2}}[S]V^t$.

6: $G_k^T = Y_k^T - \frac{1}{2}(Y_k^B + Y_k^T - D)$.

7: $T_{k+1} = S_{\frac{\lambda\mu_k}{2}}[G_k^T]$.

8: $a_{k+1} = \frac{1 + \sqrt{4a_k^2 + 1}}{2}; \mu_{k+1} = \max(\eta\mu_k, \bar{\mu})$.

9: $k = k + 1$.

10: **end while**

Output: $B = B_k, T = T_k$.

Fourth, we use a simple segmentation method to segment adaptively the target image f_T since it contains some errors with small values. Finally, by post-processing, we refine the segmentation result to obtain the final detection result.

In the second step, for Algorithm 1, we choose $\lambda = 1/\sqrt{\max(m, n)}$, $\eta = 0.99$ and $\mu_0 = s_2$, $\bar{\mu} = 0.05s_4$, where s_2, s_4 are the second and the fourth largest singular values of D in our implementation, respectively.

In the third step, we choose the *1D median* filter function, namely $v = \text{median}(\mathbf{x})$, to implement the mapping (3): $\mathbf{R}^p \rightarrow \mathbf{R}$. We also test the *1D mean* filter function and

the experimental results show that the *1D median* filter function is more robust than the *1D mean* filter function. If the background image is not used in the applications, we can just reconstruct the target image, which can reduce the computation.

In the fourth step, the adaptive threshold t_{up} is determined by:

$$t_{up} = \max(v_{min}, \mu + k\sigma), \quad (12)$$

where μ and σ are the mean value and standard deviation of the target image f_T , respectively, k and v_{min} are constants determined experientially. v_{min} is to delete false targets. In this paper, we focus on so-called “bright” targets. Therefore, we can segment a pixel at (x, y) as the target pixel if $f_T(x, y) > t_{up}$, otherwise it is a background pixel.

If “bright” and “dark” targets need to be detected simultaneously, we can use double thresholds (t_{up}, t_{down}) to segment the target image f_T . The other threshold t_{down} can be obtained by modifying (12) as following:

$$t_{down} = \min(v_{max}, \mu - k\sigma), \quad (13)$$

where v_{max} is also a constant determined experientially, which can delete false targets like v_{min} in (12). Then we can judge a pixel at (x, y) as the target pixel if $f_T(x, y) > t_{up}$ or $f_T(x, y) < t_{down}$, otherwise it is a background pixel.

In the final step, some post-processing techniques can be used to refine the segmentation results. For example, we can use region analysis techniques to delete false detections or use

morphological techniques to refine the target regions. In addition, for any detected target we can use statistical techniques to estimate the complexity of its corresponding local region in the reconstructed background image, and then the estimated result can be used to evaluate its reliability. For example, if the corresponding local background of a target is simple, we could believe that the detection target is more reliable and vice versa. Generally speaking, the post-processing is still an important step to achieve good detection results. However, for fair comparison with the baseline methods, we do not perform post-processing and directly use the segmentation results to evaluate the performance in our experiments in Section V.

F. Computational Complexity

We briefly discuss the complexity of the proposed method for small target detection. As shown in Fig. 5, it is easy to see that the computational time of our method mainly consists of three parts: the PCP computation (Algorithm 1), the median operation for reconstruction and the target segmentation.

The computational complexity of the PCP computation (please see Algorithm 1) is essentially determined by the truncated SVD computation on G_k^B (Step 5) and the soft thresholding operators on G_k^T (Step 7). By utilizing the current fast SVD technique (e.g., QUIC-SVD [45]), the truncated SVD in step 5 can be implemented in $O(kmn)$ time, where k is the number of nonzero singular values (rank) of G_k^T . By employing the well-known heap sort algorithm [46], the soft thresholding operation in step 7 can be computed in $O(mn \log(mn))$ time. So the entire computational complexity of Algorithm 1 is around $O(Nkmn \log(mn))$, where N is the iteration number of the algorithm.

For reconstruction, we need to implement the median operations across all pixels. For each pixel, the median operation needs $O(p)$ cost [46], where p the overlapping pixel number during the transformation from the target/background patch-image to the reconstruction image. Thus the total computational complexity of this step is around $O(rcp)$, where r and c are the row and column numbers of the original image, respectively.

For target segmentation, an adaptive threshold (or double thresholds) is used. For each pixel, only a simple comparison operator is implemented. The entire cost of this step is thus around $O(rc)$.

Based on the aforementioned analysis, the entire computational complexity of the proposed method is thus around $O(Nkmn \log(mn) + rc(p + 1))$.

V. EXPERIMENTS

In this section, we firstly introduce the evaluation metrics and the baseline methods for comparison in this paper. Then we perform simulation experiments to evaluate the effects of parameters of the proposed method and its performances with respect to target sizes, SCR values and target numbers. Finally, we use the real image sequences to test the proposed method.

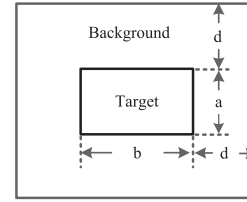


Fig. 6. The external rectangle and neighboring background rectangle of a small target.

A. Metrics and Baseline Methods

The most important metrics of evaluating the detection performance are the probability of detection P_d and false-alarm rate F_a which are defined as following [7]:

$$P_d = \frac{\# \text{ number of true detections}}{\# \text{ number of actual targets}}, \quad (14)$$

$$F_a = \frac{\# \text{ number of false detections}}{\# \text{ number of images}}. \quad (15)$$

The detected result is considered correct if it simultaneously meets two requirements: (i) The result and a ground truth have overlap pixels; (ii) The pixel distance between centers of the ground truth and the result is within a threshold (4 pixels).

Signal-to-clutter ratio (SCR) can also be used to describe the difficulty of small target detection. Generally speaking, the higher the SCR of a small target is, the easier it can be detected. Since the target size is small, the local SCR is used in this paper which is defined as follows [33], [47]:

$$SCR = \frac{|\mu_t - \mu_b|}{\sigma_b}, \quad (16)$$

where μ_t is the average pixel value of the target, μ_b and σ_b are the average pixel value and the standard deviation of the pixel values in neighboring area around the target, respectively. Because the small target size varies within a range, we use a variable neighboring area with respect to the target. As shown in Fig. 6, if the size of a small target is $a \times b$, the size of its background rectangle is $(a + 2d) \times (b + 2d)$, where d is a constant which equals to 20 pixels in this paper.

Based on the Equation (16), the average SCR value (denoted as \overline{SCR}) of targets is defined as follows:

$$\overline{SCR} = \frac{1}{N} \sum_{i=1}^N SCR_i, \quad (17)$$

where N is the number of targets and SCR_i is the SCR of the i^{th} target. In this paper, we use the \overline{SCR}_d and \overline{SCR}_l to denote \overline{SCR} s of detected true targets and lost true targets, respectively. The smaller the \overline{SCR}_d and \overline{SCR}_l are, the better the performance of one method is to some extent when fixing the F_a .

The same as the recent work [4], we choose the *TopHat* filtering method [10], [11] as one baseline method. Moreover, *MaxMean* and *MaxMedian* filtering methods [48] are also chosen as the baseline methods in this paper since these methods are well studied and are usually used as baseline methods for assessing new methods [33], [28].

B. Simulation

In this section, we perform simulation experiments to evaluate the effects of the parameters, including the sliding step and the patch size, and the performance of the proposed method compared to the baseline methods.

1) *Data Setup*: Test images are synthesized by using 100 real infrared background images and 80 targets. The background images are chosen from several real image sequences with different clutters and the targets are obtained by using the bicubic interpolation method to resize four real targets from Fig. 1, where the original sizes of four targets are 5×4 , 8×4 , 5×3 and 6×4 , respectively. Suppose that the original size of a target is $m \times n$, we resize the target to $am \times an$, where $a \in (\frac{2}{\min(m,n)}, \frac{3}{\min(m,n)}, \dots, \frac{i}{\min(m,n)}, \dots, \frac{21}{\min(m,n)})$. In this way we can obtain 20 targets with different sizes for each original target. For example, the second target with size of 8×4 can produce 20 targets with sizes of 4×2 , 6×3 , 8×4 , 10×5 , \dots , 38×19 , 40×20 and 42×21 , respectively, where the target with size of 8×4 is the same as the original target. For convenience of synthesizing images below, the pixel values of all target images are normalized to (0, 1).

A synthesized image f_D can be achieved by embedding a target image f_T with size of $m \times n$ into a background image f_B by the following way:

$$f_D(x, y) = \begin{cases} \max \left(r f_T(x - x_0, y - y_0), \right. \\ \quad \left. f_B(x, y) \right) & x \in (1 + x_0, n + x_0), \\ & y \in (1 + y_0, m + y_0) \\ f_B(x, y) & \text{Otherwise,} \end{cases} \quad (18)$$

where (x_0, y_0) is a randomly produced pixel location which the left upper corner of the image f_T corresponds to in the image f_B , r is also produced randomly within the range of $[h, 255]$, here h is the maximum pixel value of the background image patch covered by the image f_T (note that the target image f_T is the same size as the images f_D and f_B except for Equation (18) in this paper). Finally we use a Gaussian filter to blur the synthesized target to make it close to a real one. This way, targets in the synthesized image are usually smaller than their original sizes. Actually, the same target usually have different sizes and SCR values in different locations even in the same synthesized image.

In the above simulation method, four image sequences are synthesized and all images in the same sequence have the same number of embedded targets which are 1, 4, 7 and 10, respectively. We choose 1676 targets from 2200 targets which meet the requirements of the area and SCR value (please see Table I). Here, the area of a target is the actual number of pixels. According to the areas of targets, we divide all targets into 10 groups and every group contains more than 90 targets. This division strategy can evaluate objectively the performance of the proposed method in term of different target sizes. The details of targets are listed in Table I. In total, the target areas range from 4 (about 2×2) to 130 pixels (about 11.4×11.4); The average area of all 1676 targets is 37.5 pixels and their average size is 6.1×6.1 ; The SCR

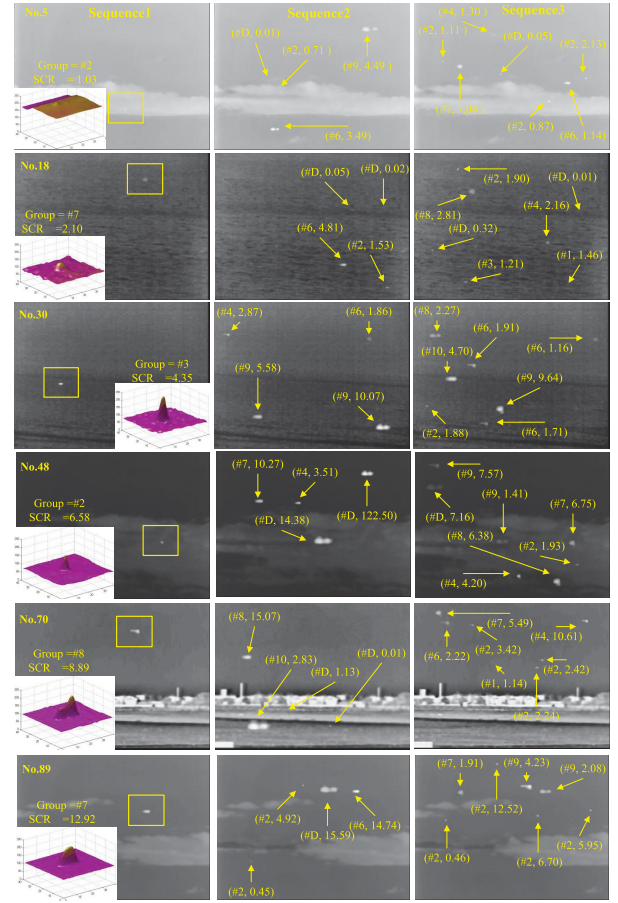


Fig. 7. The representative synthetic images for three sequences. Images in each column are from the same sequence and images in each row have the same background.

values range from 0.4 to 18.2 and their average is 3.2 (please see Table I). Fig. 7 shows the representative images from three image sequences (the fourth image sequence is not given because of the spacial limitation). The first column images are chosen from the first image sequence (denoted as Sequence 1) and are arrayed according to SCR values in ascending order, where the frame numbers, target regions (rectangles), group numbers, SCR values and their 3D surfaces are overlaid. The other two column images are chosen from the second and third image sequences (denoted as Sequence 2 and Sequence 3) according to the frame numbers in Sequence 1. In the second and third column images all target locations are denoted by arrows, and the corresponding group numbers and SCR values are denoted by the form of (# group number, SCR value). The group numbers denoted as “D” mean that the corresponding targets do not meet the requirements of the area or SCR value as mentioned above.

From the Fig. 7, we can see that the targets in synthetic images vary from the small dim ones to the large bright ones. In addition, the different numbers of targets are contained in a single image. Thus it is reasonable to use these synthetic images to test the proposed method.

2) *Effects of Parameters*: In this section, we perform two experiments to discuss the effects of the patch size and the

TABLE I
THE DETAILS FOR TARGETS IN SYNTHETIC IMAGES. THE AREA OF A TARGET IS THE
NUMBER OF ITS PIXELS AND THE N_{target} IS THE NUMBER OF TARGETS

	Group1	Group2	Group3	Group4	Group5	Group6	Group7	Group8	Group9	Group10	The whole
area	4~6	7~16	17~20	21~27	28~36	37~46	47~57	58~70	71~95	96~130	4~130
area	4.9	11.1	18.3	23.6	31.9	41.6	52.4	63.9	80.6	108.0	37.5
size	2.2×2.2	3.3×3.3	4.3×4.3	4.9×4.9	5.6×5.6	6.5×6.5	7.2×7.2	8.0×8.0	9.0×9.0	10.4×10.4	6.1×6.1
N_{target}	122	379	139	161	165	164	159	145	151	91	1676
SCR	0.4~6.2	0.4~14.2	0.4~15.0	0.4~14.2	0.4~18.2	0.5~15.2	0.4~14.9	0.6~15.1	0.5~14.9	0.9~15.1	0.4~18.2
SCR	1.5	2.3	2.5	2.7	3.6	3.4	3.9	4.6	4.4	5.2	3.2

TABLE II
THE DETECTION PERFORMANCE OF THE PROPOSED METHOD WITH RESPECT TO DIFFERENT SLIDING STEPS AND HERE $step_m$ MEANS
THAT THE HORIZONTAL AND VERTICAL SLIDING STEPS ARE m PIXELS ($F_a = 2.5/IMAGE$ AND THE PATCH SIZE IS 50×50)

Metrics	Step	Group1	Group2	Group3	Group4	Group5	Group6	Group7	Group8	Group9	Group10	The whole
P_d	$step_6$	0.67	0.84	0.84	0.88	0.85	0.79	0.75	0.68	0.60	0.52	0.81
	$step_8$	0.61	0.83	0.91	0.92	0.88	0.82	0.79	0.79	0.72	0.67	0.82
	$step_{10}$	0.51	0.79	0.88	0.85	0.88	0.78	0.80	0.85	0.74	0.77	0.79
	$step_{12}$	0.60	0.82	0.91	0.94	0.92	0.85	0.83	0.87	0.87	0.81	0.81
	$step_{14}$	0.56	0.82	0.88	0.94	0.92	0.89	0.84	0.91	0.86	0.90	0.79
\overline{SCR}_d	$step_6$	1.79	2.49	2.77	2.98	4.00	3.92	4.67	5.31	5.29	6.29	3.60
	$step_8$	1.84	2.49	2.63	2.88	3.95	3.87	4.46	4.96	5.04	6.43	3.59
	$step_{10}$	1.89	2.54	2.71	3.01	3.94	3.93	4.47	5.05	5.02	6.17	3.69
	$step_{12}$	1.81	2.50	2.66	2.84	3.82	3.80	4.35	4.98	4.71	6.06	3.64
	$step_{14}$	1.80	2.50	2.70	2.83	3.83	3.71	4.30	4.86	4.75	5.71	3.69
\overline{SCR}_l	$step_6$	0.87	1.05	1.01	0.92	1.25	1.64	1.70	2.95	3.13	4.15	1.77
	$step_8$	0.94	1.14	1.02	1.00	1.14	1.55	1.86	3.11	2.64	2.75	1.61
	$step_{10}$	1.07	1.20	0.96	1.06	1.14	1.74	1.73	1.87	2.73	2.69	1.59
	$step_{12}$	1.01	1.19	0.84	0.99	1.23	1.38	1.83	1.83	2.57	2.45	1.56
	$step_{14}$	1.09	1.17	0.87	0.92	1.17	1.34	1.88	1.61	2.45	2.47	1.55

sliding step, respectively. For quantitative analysis, both experiments have fixed false-alarm rates F_a (2.5/image) by changing the segmentation thresholds for each group. In addition, we give the receiver operating characteristic (ROC) curves of the whole for different parameters in Fig. 8, where the whole means that targets includes from Group 1 to Group 10.

In the first experiment, we fix the patch size as 50×50 and set the sliding step as 6, 8, 10, 12, 14 and then test the proposed method, respectively. Here, we keep the column and row steps the same. The evaluation results are shown in Table II. Here, the images used to compute the false-alarm rate F_a of group $\#i$ are those which contain at least one true target of group $\#i$.

From Table II, we can see that the performance of small sliding steps is better than that of big ones when the targets are small. In our method (please see the Fig. 5), every pixel value in the reconstructed image is obtained by using the 1D median filter for a serial of values from overlapped image patches. The smaller the sliding step is, the more the values are used for the 1D median filter. This would make the proposed method more robust. As the targets become bigger, the larger sliding steps have better performance. However, from the last column in Table II and the ROC curves in Fig. 8(a), we can see that the performance for different steps is close to each other when considering all targets together, although the performance with sliding step 8 is slightly better than other sliding steps.

In the second experiment, we use the square patch size $m \times m$ and set it as 30×30 , 40×40 , 50×50 , 60×60 , 70×70 , 80×80 , 81×81 , 82×82 , 83×83 , respectively,

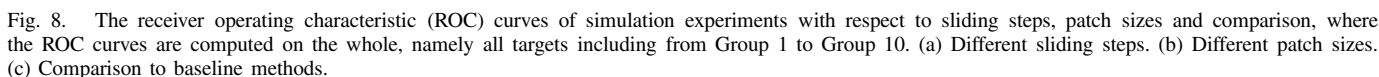
and the corresponding horizontal and vertical sliding steps (0.17m) are 5, 7, 9, 10, 12, 14, 14, 14 and 14. The evaluation results of the proposed method are shown in Table III and in Fig. 8(b).

It can be seen from Table III that up to the patch size of 80×80 , the larger the patch size is, the better the proposed method is likely to perform. This is because the target patch-image tends to be sparser when using a bigger patch size. However, when the patch size is bigger than 80×80 , the performance begins to degrade. This indicates that there is an appropriate patch size in a reasonable range for specific test data and this appropriate patch size can be obtained by empirical experiments. Through the ROC curves in Fig. 8(b), we can see that for the fixed probability of detection P_d , the false-alarm rate F_a of patch size 80×80 is smallest and when the patch sizes are bigger or smaller than 80×80 , the corresponding false-alarm rates F_a inclines to decrease. Comparing Fig. 8(a) to Fig. 8(b), it can be seen that the parameter of patch size has more influence on the performance than the sliding step.

3) *Comparison to Baseline Methods:* Table IV and Fig. 8(c) give comparisons between the proposed method with two parameter configurations (denoted as “Our method 1” and “Our method 2” and the corresponding patch sizes and sliding steps are configurations of the $(50 \times 50, 9)$ and $(80 \times 80, 14)$, respectively) and the baseline methods *TopHat*, *MaxMean* and *MaxMedian* whose filter sizes are 15×15 , and in Table IV all four methods have the same false-alarm rate

Metrics	Size	Group1	Group2	Group3	Group4	Group5	Group6	Group7	Group8	Group9	Group10	The whole
P_d	size ₃₀	0.61	0.77	0.85	0.78	0.71	0.69	0.60	0.48	0.32	0.22	0.65
	size ₄₀	0.61	0.82	0.89	0.87	0.85	0.78	0.75	0.67	0.57	0.44	0.77
	size ₅₀	0.61	0.83	0.91	0.91	0.87	0.84	0.81	0.82	0.74	0.70	0.82
	size ₆₀	0.57	0.80	0.88	0.89	0.87	0.81	0.84	0.86	0.78	0.74	0.81
	size ₇₀	0.70	0.86	0.94	0.95	0.92	0.91	0.91	0.92	0.88	0.86	0.88
	size ₈₀	0.80	0.90	0.94	0.95	0.93	0.94	0.92	0.96	0.90	0.86	0.91
	size ₈₁	0.80	0.89	0.94	0.95	0.93	0.93	0.92	0.95	0.90	0.86	0.90
	size ₈₂	0.71	0.85	0.93	0.95	0.91	0.88	0.91	0.94	0.90	0.85	0.89
	size ₈₃	0.25	0.35	0.63	0.40	0.35	0.44	0.63	0.55	0.41	0.33	0.60
\overline{SCR}_d	size ₃₀	1.83	2.59	2.77	3.17	4.44	4.18	5.22	5.63	5.50	5.94	3.64
	size ₄₀	1.84	2.50	2.68	2.98	4.04	3.94	4.59	5.36	5.17	6.16	3.60
	size ₅₀	1.83	2.50	2.65	2.89	3.96	3.83	4.40	5.14	4.98	6.27	3.60
	size ₆₀	1.84	2.54	2.70	2.93	3.98	3.84	4.34	5.00	4.91	6.24	3.62
	size ₇₀	1.77	2.46	2.59	2.83	3.83	3.66	4.11	4.83	4.66	5.67	3.49
	size ₈₀	1.67	2.41	2.60	2.83	3.82	3.60	4.15	4.69	4.68	5.66	3.44
	size ₈₁	1.67	2.43	2.60	2.83	3.82	3.60	4.17	4.72	4.68	5.66	3.46
	size ₈₂	1.78	2.47	2.62	2.83	3.87	3.72	4.15	4.74	4.68	5.69	3.47
	size ₈₃	1.94	3.48	3.34	4.09	5.35	4.88	4.97	6.01	5.55	6.54	4.28
\overline{SCR}_l	size ₃₀	0.96	1.16	0.94	1.19	1.58	1.82	1.99	3.57	3.84	5.02	2.45
	size ₄₀	0.95	1.15	0.96	1.03	1.18	1.70	1.94	2.95	3.33	4.48	1.96
	size ₅₀	0.95	1.14	0.99	1.03	1.17	1.49	1.93	1.91	2.64	2.73	1.59
	size ₆₀	1.03	1.12	0.99	1.01	1.15	1.77	1.79	1.84	2.45	2.37	1.49
	size ₇₀	0.81	1.06	0.93	0.74	0.98	1.32	1.97	1.62	2.30	2.54	1.35
	size ₈₀	0.76	0.93	0.75	0.80	0.87	1.14	1.06	1.66	1.59	2.54	1.17
	size ₈₁	0.76	0.98	0.75	0.80	0.84	1.38	1.07	1.55	1.59	2.54	1.19
	size ₈₂	0.75	1.06	0.85	0.68	0.97	1.37	1.60	1.55	1.61	2.63	1.19
	size ₈₃	1.34	1.60	1.07	1.83	2.66	2.32	2.15	2.79	3.56	4.57	1.67

Metrics	Methods	Group1	Group2	Group3	Group4	Group5	Group6	Group7	Group8	Group9	Group10	The whole
P_d	TopHat	0.36	0.58	0.67	0.64	0.70	0.70	0.75	0.81	0.84	0.85	0.65
	MaxMean	0.49	0.67	0.72	0.67	0.68	0.60	0.66	0.57	0.48	0.34	0.62
	MaxMedian	0.60	0.84	0.91	0.89	0.84	0.75	0.67	0.63	0.36	0.25	0.71
	Our method 1	0.61	0.83	0.91	0.91	0.87	0.84	0.81	0.82	0.74	0.70	0.82
	Our method 2	0.80	0.90	0.94	0.95	0.93	0.94	0.92	0.96	0.90	0.86	0.91
\overline{SCR}_d	TopHat	1.97	2.92	3.21	3.34	4.39	4.16	4.62	4.97	4.74	5.53	4.04
	MaxMean	1.95	2.81	3.05	3.25	4.44	4.47	4.90	5.47	5.16	6.28	3.91
	MaxMedian	1.91	2.49	2.66	2.87	4.04	3.88	4.54	5.20	5.25	5.03	3.46
	Our method 1	1.83	2.50	2.65	2.89	3.96	3.83	4.40	5.14	4.98	6.27	3.60
	Our method 2	1.67	2.41	2.60	2.83	3.82	3.60	4.15	4.69	4.68	5.66	3.44
\overline{SCR}_t	TopHat	1.22	1.36	1.05	1.63	1.81	1.77	1.84	2.87	2.37	3.16	1.67
	MaxMean	1.04	1.16	1.06	1.66	1.84	1.93	2.01	3.36	3.64	4.67	2.09
	MaxMedian	0.86	1.06	0.90	1.59	1.39	2.16	2.69	3.46	3.88	5.28	2.64
	Our method 1	0.95	1.14	0.99	1.03	1.17	1.49	1.93	1.91	2.64	2.73	1.59
	Our method 2	0.76	0.93	0.75	0.80	0.87	1.14	1.06	1.66	1.59	2.54	1.17



to the baseline methods *TopHat*, *MaxMean* and *MaxMedian*. Table IV also shows that the *TopHat* method has superior detection capability for large targets than small ones, and

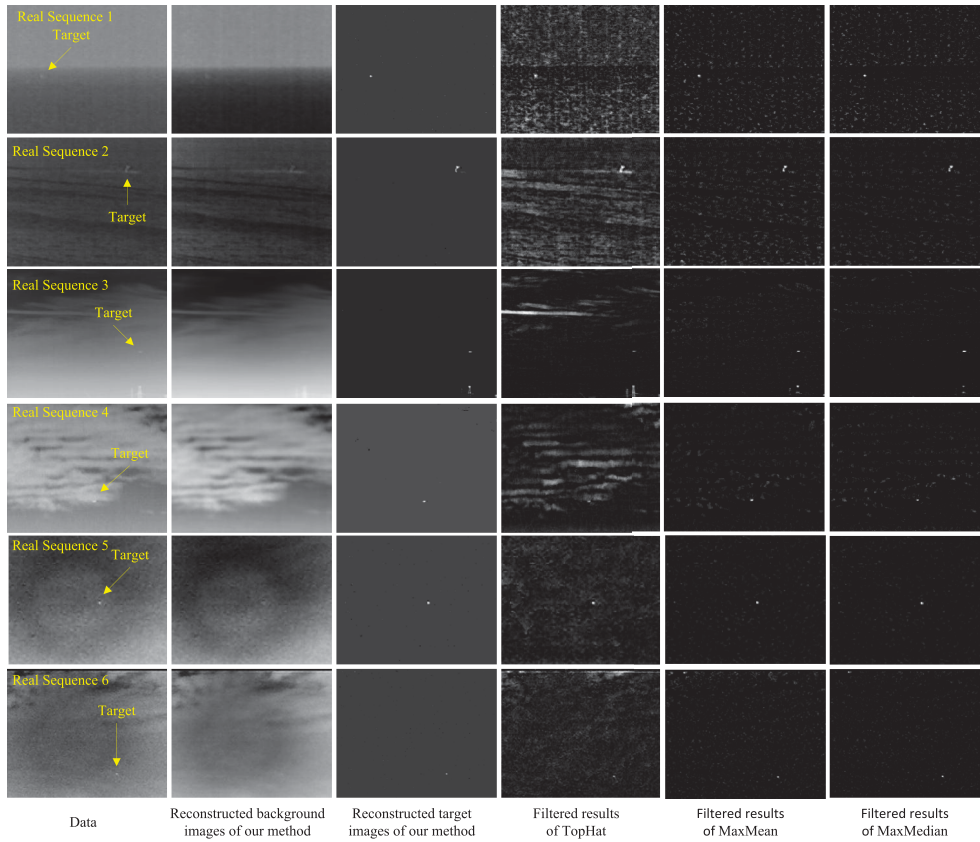


Fig. 9. The representative images of the six real image sequences and the corresponding processed results of different methods.

TABLE V
THE DETAILS OF THE SIX REAL SEQUENCES

Sequence	Target type	Target details	Background details
Real Sequence 1	A small ship	<ul style="list-style-type: none"> • A long imaging distance. • Low SCR value. • A small size with a little change. 	<ul style="list-style-type: none"> • Blurred sea-sky backgrounds. • Almost keeping the same.
Real Sequence 2	A small ship	<ul style="list-style-type: none"> • A short imaging distance. • A changing size within a big range. • Low SCR value. 	<ul style="list-style-type: none"> • Heavy sea clutter backgrounds. • Changing backgrounds.
Real Sequence 3	An airplane	<ul style="list-style-type: none"> • A long imaging distance. • Keeping horizontal motion. • A small size. 	<ul style="list-style-type: none"> • Somewhat heavily cloudy clutter background. • Changing backgrounds. • There are “bright” man-made buildings.
Real Sequence 4	An airplane	<ul style="list-style-type: none"> • A long imaging distance. • A changing size within a big range. 	<ul style="list-style-type: none"> • Uniform backgrounds to heavy cloudy backgrounds.
Real Sequence 5	A helicopter	<ul style="list-style-type: none"> • A single target to two targets. • A small size. 	<ul style="list-style-type: none"> • Almost keeping the same. • There are heavy noises.
Real Sequence 6	A helicopter	<ul style="list-style-type: none"> • A changing size within a small range. 	<ul style="list-style-type: none"> • Almost keeping the same. • Heavy noise.

the *MaxMean* and *MaxMedian* methods are the opposite. By contrast, the proposed method can work stably from small targets to large targets (from Group1 to Group10). Especially, our method with the patch size of 80×80 and the sliding step of 14 improves the probability of detection P_d by more than 0.15 (15%) for Group 1, 2, 6, 7, 8, 9 and 10 compared to both *MaxMean* and *MaxMedian* methods. From the last column in Table IV, we can see that when considering all targets as the whole, our method for two parameter configurations improves the probability of detection P_d by more than 0.1 (10%) compared to all three baseline methods. Through the ROC curves in Fig. 8(c), we can see that when the

false-alarm is bigger than 1.5, the patch size of 50×50 has better probability of detection P_d compared to all three baseline methods. When the patch size is 80×80 , the probability of detection P_d of the proposed method is the best for different false-alarm rates and the false alarm rates are the smallest for fixed the probabilities of detection P_d .

C. Real Images

We use six real image sequences to test the proposed methods. The first column in Fig. 9 are representative images of six sequences denoted as Real Sequence 1 to 6, respectively

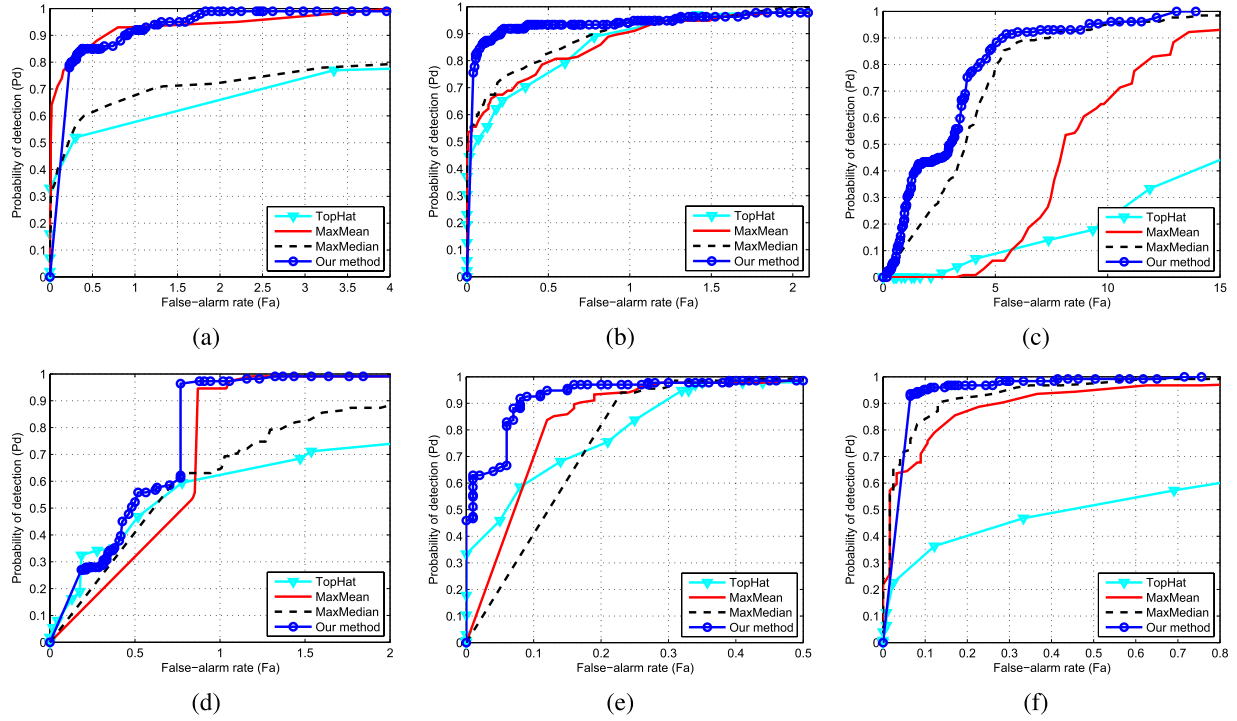


Fig. 10. The receiver operating characteristic (ROC) curves of four methods for the six real image sequences. (a) Real Sequence 1. (b) Real Sequence 2. (c) Real Sequence 3. (d) Real Sequence 4. (e) Real Sequence 5. (f) Real Sequence 6.

and their details are listed in Table V. Although the proposed method with size 80×80 in simulation experiments has better performance than that with size 50×50 (please see Table IV), we use the latter parameter configuration based on the consideration of computational cost. Because the target varies from a small size to a large one, the filter sizes of baseline methods are still set as 15×15 which is the same as in our simulation experiments.

The columns from the third to the sixth in Fig. 9 show processed results of four methods before segmentation. It can be seen that our method has less clutter and noise residual for different clutter backgrounds compared to baseline methods, which is the key to keep lower false-alarm rates (F_a) under the same probability of detection (P_d). From the second column we can see that our method could recover the background images effectively. This is very useful for some applications. For example, we can use the recovered background to estimate the reliability of detection results as discussed before.

Fig. 10 shows the ROC curves of the four methods for six real image sequences. It can be seen that the proposed method has better performance than baseline methods. Especially for Real Sequence 2 to 6, our method has higher probabilities of detection (P_d) compared to baseline methods while the false-alarm rates (F_a) are low. For the Real Sequence 1, the *MaxMean* method has a little better performance than our method when $F_a \leq 1.5$, but our method can reach 1 (100%) faster than the *MaxMean* method when $F_a > 1.5$. Fig. 10 also shows that the *MaxMean* method has better performance than the other baseline methods for Real Sequence 1, 4 and 5. For Real Sequence 2, 3 and 6, the *MaxMedian* has better performance than other baseline methods. However,

our method can obtain the best performance for all six real sequences, which means that our method can work more stably for different clutter and noisy backgrounds and target types.

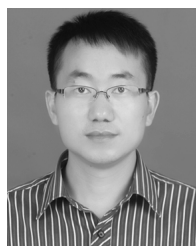
VI. CONCLUSION

A new infrared image model called IPI model for small target detection is presented based on the non-local self-correlation property of the infrared image in this paper. Then the small target detection task is transformed into an optimization problem of recovering low-rank and sparse matrices, which can be effectively solved by using Stable Principal Component Pursuit. Extensive synthetic and real data experiments show that under different backgrounds the proposed method can not only work more stably for different target sizes and SCR values, but also has better detection performance compared to conventional baseline methods. In the future, we will investigate faster version of the current algorithm. In addition, we will further generalize our 2D patch model into 3D or more dimensions and investigate applications of the N -D patch model. We will also try recent multi-subspace cluster strategies [42] to further improve the flexibility of our method in highly variant background cases in our future investigation.

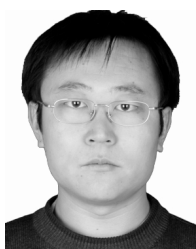
REFERENCES

- [1] I. S. Reed, R. M. Gagliardi, and L. B. Stotts, "Optical moving target detection with 3D matched filtering," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 327–336, Jan. 1988.
- [2] M. Li, T. Zhang, W. Yang, and X. Sun, "Moving weak point target detection and estimation with three-dimensional double directional filter in IR cluttered background," *Opt. Eng.*, vol. 44, pp. 107007-1–107007-4, Oct. 2005.

- [3] K. Melendez and J. Modestino, "Spatiotemporal multiscale adaptive matched filtering," *Proc. SPIE*, vol. 2561, pp. 51–65, Sep. 1995.
- [4] S. Kim and J. Lee, "Scale invariant small target detection by optimizing signal-to-clutter ratio in heterogeneous background for infrared search and track," *Pattern Recognit.*, vol. 45, pp. 393–406, Jan. 2012.
- [5] X. Bai and F. Zhou, "Analysis of different modified top-hat transformations based on structuring element construction," *Signal Process.*, vol. 90, pp. 2999–3003, Nov. 2010.
- [6] Y. F. Gu, C. Wang, B. X. 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] J. F. Rivest and R. Fortin, "Detection of dim targets in digital infrared imagery by morphological image processing," *Opt. Eng.*, vol. 35, pp. 1886–1893, Jul. 1996.
- [8] K. L. Anderson and R. A. Itlis, "A tracking algorithm for infrared images based on reduced sufficient statistics," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, no. 2, pp. 464–472, Apr. 1997.
- [9] X. Yan, P. Jia-Xiong, D. Ming-yue, and X. Dong-Hui, "An extended track-before-detect algorithm for infrared target detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, no. 3, pp. 1087–1092, Jul. 1997.
- [10] V. Tom, T. Peli, M. Leung, and J. Bondaryk, "Morphology-based algorithm for point target detection in infrared backgrounds," *Proc. SPIE*, vol. 1954, pp. 25–32, Oct. 1993.
- [11] A. Toet and T. Wu, "Small maritime target detection through false color fusion," *Proc. SPIE*, vol. 6945, pp. 69450V–69453V, Apr. 2008.
- [12] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, pp. 1–37, May 2011.
- [13] Z. Zihan, L. Xiaodong, J. Wright, E. Cande, and M. Yi, "Stable principal component pursuit," in *Proc. IEEE ISIT*, Jun. 2010, pp. 1518–1522.
- [14] D. Meng and F. De la Torre, "Robust matrix factorization with unknown noise," in *Proc. ICCV*, 2013.
- [15] D. Meng, Z. Xu, L. Zhang, and J. Zhao, "A cyclic weighted median method for 11 low-rank matrix factorization with missing entries," in *Proc. AAAI*, 2013.
- [16] B. Porat and B. Friedlander, "A frequency domain algorithm for multiframe detection and estimation of dim targets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 4, pp. 398–401, Apr. 1990.
- [17] X. Liu and Z. Zuo, *A Dim Small Infrared Moving Target Detection Algorithm Based on Improved Three-Dimensional Directional Filtering Communications in Computer and Information Science*, vol. 363, New York, NY, USA: Springer-Verlag, 2013, pp. 102–108, ch. 13.
- [18] B. Zhang, T. Zhang, Z. Cao, and K. Zhang, "Fast new small-target detection algorithm based on a modified partial differential equation in infrared clutter," *Opt. Eng.*, vol. 46, no. 10, pp. 106401-1–106401-4, 2007.
- [19] P. Wei, B. Zeidler, and W. Ku, "Analysis of multiframe target detection using pixel statistics," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 1, pp. 238–247, Jan. 1995.
- [20] J. Silverman, J. M. Mooney, and C. E. Cafer, "Tracking point targets in cloud clutter," *Proc. SPIE*, vol. 3061, pp. 496–507, Aug. 1997.
- [21] A. P. Tzannes and D. H. Brooks, "Detecting small moving objects using temporal hypothesis testing," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 2, pp. 570–586, Apr. 2002.
- [22] E. T. Lim, L. Shue, and R. Venkateswarlu, "Adaptive mean and variance filter for detection of dim point-like targets," *Proc. SPIE*, vol. 4728, pp. 492–502, Aug. 2002.
- [23] D. Liu, J. Zhang, and W. Dong, "Temporal profile based small moving target detection algorithm in infrared image sequences," *Int. J. Infr. Millimeter Waves*, vol. 28, no. 5, pp. 373–381, 2007.
- [24] T.-W. Bae, "Small target detection using bilateral filter and temporal cross product in infrared images," *Infr. Phys. Technol.*, vol. 54, pp. 403–411, Sep. 2011.
- [25] Z. Wang, J. Tian, J. Liu, and S. Zheng, "Small infrared target fusion-detection based on support vector machines in the wavelet domain," *Opt. Eng.*, vol. 45, no. 7, pp. 076401-1–076401-3, 2006.
- [26] D. Chan, D. Langan, and D. Staver, "Spatial-processing techniques for the detection of small targets in IR clutter," *Proc. SPIE*, vol. 1305, pp. 53–62, Oct. 1990.
- [27] T. Soni, J. R. Zeidler, and W. H. Ku, "Performance evaluation of 2D adaptive prediction filters for detection of small objects in image data," *IEEE Trans. Image Process.*, vol. 2, no. 3, pp. 327–340, Jul. 1993.
- [28] X. P. Wang and T. X. Zhang, "Clutter-adaptive infrared small target detection in infrared maritime scenarios," *Opt. Eng.*, vol. 50, no. 6, pp. 067001-1–067001-12, Jun. 2011.
- [29] M. R. Azimi-Sadjadi and P. Hongye, "Two-dimensional block diagonal LMS adaptive filtering," *IEEE Trans. Signal Process.*, vol. 42, no. 9, pp. 2420–2429, Sep. 1994.
- [30] P. A. Ffrench, J. R. Zeidler, and W. H. Ku, "Enhanced detectability of small objects in correlated clutter using an improved 2D adaptive lattice algorithm," *IEEE Trans. Image Process.*, vol. 6, no. 3, pp. 383–397, Mar. 1997.
- [31] Y. Cao, R. Liu, and J. Yang, "Small target detection using two-dimensional least mean square (TDLMS) filter based on neighborhood analysis," *Int. J. Infr. Millimeter Waves*, vol. 29, no. 2, pp. 188–200, 2008.
- [32] H. Deng, J. G. Liu, and Z. Chen, "Infrared small target detection based on modified local entropy and EMD," *Chin. Opt. Lett.*, vol. 8, pp. 24–28, Jan. 2010.
- [33] C. Q. Gao, T. Q. Zhang, and Q. Li, "Small infrared target detection using sparse ring representation," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 27, no. 3, pp. 21–30, Mar. 2012.
- [34] M. Zeng, J. Li, and Z. Peng, "The design of top-hat morphological filter and application to infrared target detection," *Infr. Phys. Technol.*, vol. 48, pp. 67–76, Apr. 2006.
- [35] X. Bai, F. Zhou, Y. Xie, and T. Jin, "Modified top-hat transformation based on contour structuring element to detect infrared small target," in *Proc. 3rd IEEE Conf. ICIEA*, Jun. 2008, pp. 575–579.
- [36] J. Ardouin, "Point source detection based on point spread function symmetry," *Opt. Eng.*, vol. 32, pp. 2156–2164, Sep. 1993.
- [37] A. A. Efros and T. K. Leung, "Texture synthesis by non-parametric sampling," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, vol. 2, Sep. 1999, pp. 1033–1038.
- [38] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. CVPR. Soc. Conf.*, vol. 2, Jun. 2005, pp. 60–65.
- [39] C. Kervrann and J. Boulanger, "Optimal spatial adaptation for patch-based image denoising," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 2866–2878, Oct. 2006.
- [40] A. Buades, B. Coll, and J.-M. Morel, "Nonlocal image and movie denoising," *Int. J. Comput. Vis.*, vol. 76, no. 2, pp. 123–139, 2008.
- [41] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 2272–2279.
- [42] L. Guangcan, L. Zhouchen, Y. Shuicheng, S. Ju, Y. Yong, and M. Yi, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.
- [43] Z. Lin, A. Ganesh, J. Wright, L. Wu, M. Chen, and Y. Ma, "Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix," in *Proc. CAMSAP*, 2009, pp. 1–18.
- [44] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-Rank matrices," UIUC, Champaign, IL, USA, Tech. Rep. UILU-ENG-09-2215, 2009.
- [45] M. P. Holmes, J. Isbell, C. Lee, and A. G. Gray, "QUIC-SVD: Fast SVD using cosine trees," in *Proc. Adv. Neural Inf. Process. Syst.*, 2008, pp. 673–680.
- [46] D. Knuth, *The Art of Computer Programming*, 3rd ed. Reading, MA, USA: Addison-Wesley, 1997.
- [47] M. Li, T. X. Zhang, Z. R. Zuo, X. C. Sun, and W. D. Yang, "Novel dim target detection and estimation algorithm based on double threshold partial differential equation," *Opt. Eng.*, vol. 45, no. 9, pp. 090502-1–090502-3, 2006.
- [48] S. Deshpande, M. Er, V. Ronda, and P. Chan, "Max-mean and max-median filters for detection of small-targets," *Proc. SPIE*, vol. 3809, pp. 74–83, Oct. 1999.



Chenqiang Gao received the B.S. degree in computer science from the China University of Geosciences, Wuhan, China, in 2004 and the Ph.D. degree in pattern recognition and intelligence systems from the Huazhong University of Science and Technology, Wuhan, China, in 2009. He is currently an Associate Professor with Chongqing University of Posts and Telecommunications, Chongqing, China. His current research interests include image processing, infrared target detection, and event detection.



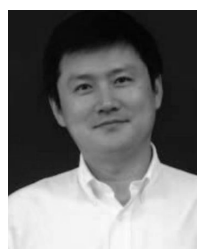
Deyu Meng received the B.Sc., M.Sc., and Ph.D. degrees in 2001, 2004, and 2008, respectively, from Xi'an Jiaotong University, Xi'an, China. He is currently an Associate Professor with the Institute for Information and System Sciences, Faculty of Science, Xi'an Jiaotong University, Xi'an. His current research interests include principal component analysis, nonlinear dimensionality reduction, feature extraction and selection, compressed sensing, and sparse machine learning methods.



Yi Yang is a DECRA Fellow with the School of Information Technology and Electrical Engineering, The University of Queensland, Queensland, Australia. Prior to that, he was a Post-Doctoral Fellow with the School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA. He received the Ph.D. degree in computer science from Zhejiang University, Zhejiang, China, in 2010. His current research interests include machine learning and its applications to multimedia content analysis and indexing.



Yongtao Wang received the Ph.D. degree in pattern recognition and intelligent system in 2009 from the Huazhong University of Science and Technology, Hubei, China. In 2010, he was a Research Scientist with the Temasek Laboratories, Nanyang Technological University, Singapore. He is currently an Assistant Professor with the Institute of Computer Science and Technology, Peking University, Beijing, China. His current research interests include wide baseline matching, motion segmentation, object recognition, and document image understanding.



Xiaofang Zhou (SM'06) received the B.S. and M.S. degrees in computer science from Nanjing University, Nanjing, China, in 1984 and 1987, respectively, and the Ph.D. degree in computer science from The University of Queensland, Queensland, Australia, in 1994. He is a Professor of computer science with The University of Queensland. He is the Head of the Data and Knowledge Engineering Research Division, School of Information Technology and Electrical Engineering, Kharagpur, India. He is the Director of the Australian Research Council (ARC)

Research Network in Enterprise Information Infrastructure and a Chief Investigator of the ARC Centre of Excellence in Bioinformatics. He is also an Adjunct Professor with Renmin University of China, Beijing, China, appointed under the Chinese National Qianren Scheme. From 1994 to 1999, he was a Senior Research Scientist and Project Leader in the Commonwealth Scientific and Industrial Research Organisation. His research was focused on completing effective and efficient solutions to managing, integrating, and analyzing very large amounts of complex data for business and scientific applications. His current research interests include spatial and multimedia databases, high performance query processing, web information systems, data mining, bioinformatics, and e-research.



Alexander G. Hauptmann received the B.A. and M.A. degrees in psychology from Johns Hopkins University, Baltimore, MD, USA, the degree in computer science from the Technischen Universität Berlin, Berlin, Germany, in 1984, and the Ph.D. degree in computer science from Carnegie Mellon University (CMU), Pittsburgh, PA, USA, in 1991. He is currently with the faculty of the Department of Computer Science and the Language Technologies Institute, CMU. His current research interests include man-machine communication, natural language processing, speech understanding and synthesis, video analysis, and machine learning.

From 1984 to 1994, he worked on speech and machine translation, when he joined the Informedia project for digital video analysis and retrieval and led the development and evaluation of news-on-demand applications.