# Robust Small Target Co-Detection from Airborne Infrared Image Sequences

**Jingli Gao [1,3], Chenglin Wen [2,*] and Meiqin Liu [1]** iD

[1]   College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China; gjl991@zju.edu.cn (J.G.);
     liumeiqin@zju.edu.cn (M.L.)
[2]   School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China
[3]   College of Software Engineering, Pingdingshan University, Pingdingshan 467000, China
[*]   Correspondence: wencl@hdu.edu.cn; Tel.: +86-138-1946-1626

**Abstract:** In this paper, a novel infrared target co-detection model combining the self-correlation features of backgrounds and the commonality features of targets in the spatio-temporal domain is proposed to detect small targets in a sequence of infrared images with complex backgrounds. Firstly, a dense target extraction model based on nonlinear weights is proposed, which can better suppress background of images and enhance small targets than weights of singular values. Secondly, a sparse target extraction model based on entry-wise weighted robust principal component analysis is proposed. The entry-wise weight adaptively incorporates structural prior in terms of local weighted entropy, thus, it can extract real targets accurately and suppress background clutters efficiently. Finally, the commonality of targets in the spatio-temporal domain are used to construct target refinement model for false alarms suppression and target confirmation. Since real targets could appear in both of the dense and sparse reconstruction maps of a single frame, and form trajectories after tracklet association of consecutive frames, the location correlation of the dense and sparse reconstruction maps for a single frame and tracklet association of the location correlation maps for successive frames have strong ability to discriminate between small targets and background clutters. Experimental results demonstrate that the proposed small target co-detection method can not only suppress background clutters effectively, but also detect targets accurately even if with target-like interference.

**Keywords:** infrared background; target extraction; target refinement; small target co-detection

## 1. Introduction

Infrared small target detection has been widely used in the airborne early warning, infrared guidance, surveillance and tracking and other fields [1–4]. In these applications, the infrared small targets have the following characteristics: (1) often immersed in strong noises or complex background (cloud clutter, plants and buildings, etc.), (2) with less texture and shape Information, (3) non-cooperative and without fixed law of movement. These characteristics make it very difficult to detect infrared small targets, and it has always been the hot and difficult issue of infrared detection field.

Because of the movement (jitter) of the infrared observation platform or the change of the imaging background, it is difficult to obtain the accurate infrared background by sequential detection methods [5–7], because the infrared small targets are easily mistaken for background and vice versa. In this case, the single frame detection methods have received a great attention recently, and are valid for infrared small target detection with static or changing backgrounds [8–10]. However, it is difficult to suppress clutters (cloud boundary, targe-like artifacts), which are very similar to real targets from the view of high intensities, because of the limited target information available in a single frame.

Fortunately, the commonality of targets in the spatio-temporal domain can be used to build better target detection models and suppress suspected clutters and noise.

To the best of our knowledge, tracklets information are rarely used in existing infrared target detection methods. Note that tracklets information are widely used in tracking problem, in which the target position in the first frame is given in advance [11,12]. However, there is no such prior target information in detection problem in which either a small target exists in a frame or not is still ambiguous.

As discussed above, upon encountering suspected targets or clutters, using the commonality of targets in the spatio-temporal domain is necessary for better detection performance. The commonality features of targets in the spatial domain can be utilized by combining two one-dimensional dense and sparse reconstruction models [13,14]. Different from the one-dimensional dense and sparse reconstruction models [13,14], in this paper we consider the two-dimensional form of dense and sparse reconstruction no longer transforming a matrix to a vector. A two-dimensional dense reconstruction model is proposed based on the global singular value decomposition (SVD) [15], which sets the first few singular values equal to zero and preserves the remaining singular values unchanged. However, this method does not give a general method to select the scope of singular values, and the center-bias mechanism will suppress small targets located at the edges of the image while suppressing clutters or noise. To address this limitation of the global SVD-based reconstruction method [15], we use differences of adjacent singular values to select the proper singular value scope for target extraction, and meanwhile use a sigmoid function to regularize the singular values in order to suppress the background components. The intuition is that each singular value indicates the ability of the corresponding sub-image to approximate the original image. In [8,16], the authors give one-dimensional sparse reconstruction models based on the patch-image model. However, these methods have the following limitations: (1) The detection performance depends largely on the patch size (it was set to $50 \times 50$ in [8] or $51 \times 51$ in [16]), and the patch vectorization and the pixel reconstruction from overlapped patches could also increase the running time of the algorithm. Moreover, in the patch-image model, one target may appear in different locations of several aligned patches, and after vectorization the intrinsic structure and correlations in the image could be broken, which could influence the separation of target and background later; (2) The algorithm uses L1-norm to measure the sparsity of small targets, but L1-norm treats each pixel independently in terms of intensity, thus the pixels with higher intensities (cloud border, artifacts), are easily mistaken for target pixels, and difficult to be removed through a global threshold [8]. Due to our observation, in an infrared background image, columns (rows) also have non-local self-correlation property and columns (rows) in distant locations are approximately linearly correlated with each other. Hence, to address the first limitation of the patch-image model, we directly consider each column (row) of an image as a column (row) of the observation matrix instead of dividing the image into patches and forming a patch vectorization matrix. Thus, we refer the proposed sparse reconstruction model as a global sparse reconstruction model. Moreover, we exploit entry-wise prior in the sparse reconstruction model to better separate targets from complex backgrounds. The intuition behind the entry-wise prior is that, each pixel in a target should be weighted differently according to its local weighted entropy which measures the local difference between the target and neighboring background. Thus, both the local target features and the global background features are incorporated into the proposed sparse reconstruction model.

For each frame, to increase the confidence level that candidates are real targets, correspondence between suspected targets obtained by dense and sparse reconstructions is conducted to suppress clutters and false alarms further. As we know, the target region in an infrared image has striking discontinuity with the surrounding background. However, due to our observation, the pixels with higher intensities (cloud border, artifacts) as a whole also have this property. Because of the limited target information available in a single frame, these targe-like false alarms could also be detected as real targets. In order to suppress false alarms further, especially the highly suspected targets, in this paper we adopt multiple frame target refinement by tracklets association, based on the facts

that real targets and false alarms have different movement characteristics, and false alarms should not be temporally continuous between successive frames like real targets. Due to that the spatio-temporal target commonality is used to refine the rough detection result of each frame in this paper, thus we refer to the propose method as a target co-detection model.

In this paper, we propose a novel infrared target co-detection model that combines the self-correlation features of backgrounds and the commonality features of targets in the spatio-temporal domain to detect infrared small targets in a sequence of images with complex backgrounds. In the first step, the dense reconstruction model is proposed to extract a coarse target map with benefit of regularization of singular values. In the second step, we design a sparse reconstruction model to extract a sparse target map. In the third step, the correspondence between suspected targets of two types of target maps are conducted to suppress clutters and noise. In the fourth step, the tracklets are associated to suppress false alarms and form trajectories which are used to confirm targets for each frame.

The contributions of this paper are summarized as: (1) A dense target extraction method based on regularization of singular values is proposed. Due to the introduction of a sigmoid function, the background components in the target map can be inhibited further. It should be noticed that we do not minimize the nuclear norm but only use the singular value information; (2) A sparse target extraction method based on entry-wise weighted robust principal component analysis is presented. The entry-wise weight uses the structure prior based on the local difference between the target and neighboring background existing in a natural scene from viewpoint of human recognition, which can promote the complex background suppression effect and keep the small target, and (3) we propose a false alarm suppression and target refinement method based on location correlation of the dense and sparse reconstruction maps for a single frame and tracklet association of the location correlation maps for successive frames. Based on the spatio-temporal commonality features of targets, this method can effectively detect small targets and suppress false alarms as much as possible.

The remainder of this paper is organized as follows. Section 2 reviews the related work from the view of processing units in the target detection. Section 3 presents our detection approach comprising of single frame target extraction and multiple frame target refinement. The evaluation on real infrared data set and comparisons are presented in Section 4. Conclusions are given in Section 5.

## 2. Related Work

In recent years, many infrared small target detection methods have been proposed for different applications. According to the processing units in the detection process, we categorize these approaches into pixel-wise, patch-image, and whole-image groups. As discussed later, each group has its own characteristic. The pixel-wise detection methods usually estimate one pixel at a time based on its local neighborhood or its temporal profile, so they could make better use of local difference between the current pixel and its neighborhoods, but are not suitable for the cases when the background scene in a sequence changes fast and weak dim targets are contained in a single image. Besides, the patch-image detection methods calculate each patch in light of the patch set which consists of low-rank and sparse parts, could suffer from more running time caused by vectorization and is also not suitable for detection of weak dim targets with complex background, and the non-local background patches could help separating targets from the patch set. Also according to low-rank matrix approximation, the whole-image detection methods could estimate a whole image from a sequence or a single image, thus the global background feature is used in separating targets from the background, but the whole-image model could also suffer from the problems of vectorization, fast changing background and low signal-to-clutter ratio.

### 2.1. Pixel-Wise Detection Methods

Besides some classical pixel-wise detection methods, such as the topHat method [17], the maxMean and maxMedian methods [18], more pixel-wise approaches have been proposed recently. In [9], the authors proposed an effective small target detection approach according to the contrast mechanism of human vision system and derived kernel model, therein the local contrast measure was defined

to compute the dissimilarity between the current location and its neighborhood. In [19], the authors presented local mutation weighted information entropy to suppress background and enhance the gray value of targets. In [10], inspired by the concept of local difference, the authors proposed a weighted local difference measure for the detection of infrared small targets. In addition, in [20], the authors developed a multiscale facet model to enhance targets and then used the multiresolution representation to reduce the false alarm rate. A fast-saliency method based on the facet model was presented for real-time infrared small target detection, and therein the facet kernel operator was designed and used in separating small targets from the background [21]. A biologically inspired method named multiscale patch-based contrast measure was proposed for small infrared target detection, which could increase the contrast between target and background [22]. Furthermore, other pixel-wise detection methods are also proposed using temporal information. Suggested by the singular value decomposition, a temporal filter was developed for dim target detection in evolving cloud clutters [23]. A nonlinear adaptive filter was proposed to detect infrared moving dim targets, and has high performance in removing large fluctuations on temporal profiles that are caused by evolving clutters [5]. By combining spatial and temporal information together, a target detection method was introduced using spatial bilateral filter and temporal cross product, which are respectively used to extract the spatial target information and the features of temporal profiles [6]. Subsequently, a spatial-temporal bilateral filter was presented to detect target trajectories, by extracting spatial and temporal target information simultaneously [7]. As discussed above, pixel-wise detection methods use a local region or a temporal profile to extract target information under the assumption that the target location has conspicuous discontinuity with the nearby background. However, when the the imaging background changes fast or the background has many types of clutters, jamming objects and noises are still the key factor to influence the detection performance.

### 2.2. Patch-Image Detection Methods

In [8], the authors proposed an infrared patch-image (IPI) model for target detection in a single frame. In the IPI model, a frame is divided into small patches and the patches are stacked as columns of a new matrix for robust principal component analysis (RPCA). The intuition behind the IPI model is that the local patches in distant regions in an infrared background image could be approximately linearly correlated with each other. Subsequently, the IPI model inspires much related work [16,24]. In [24], the authors generated an image patch set according to multi-scale transform and patch transform, and every patch was given an individual regulation weight which was computed by combining the information of patch size, patch entropy and target saliency level. In [16], the authors generated an image patch set according to the same scheme in [8], and also stacked all the patches as columns of a matrix, and each patch is given an individual regulation weight based on the steering kernel. However, one target may appear in different locations of several aligned patches, so adding the steering kernel at the central position is not always applicable. As a whole, the patch-image models describe the sparsity of small targets with L1-norm, and the cloud borders or target-like artifacts which have similar intensities with targets are easily mistaken for target pixels. In addition, the performance of IPI models depends on the patch size (it was set to $50 \times 50$ in [8] or $51 \times 51$ in [16]), and when the patch has a higher dimension, the vectorization in patch-image model needs more computation time.

### 2.3. Whole-Image Detection Methods

In [15], the authors proposed a visible image saliency detection approach based on SVD. The intuition behind this approach is that the large singular values mainly indicate the non-salient background information and slight saliency information, while the intermediate singular values indicate most or even all of the saliency information, and the small singular values contain little or even none of the saliency information. However, this approach does not give a general selection scheme for the scope of singular values which indicates the salient components, and the used center-bias mechanism could suppress small targets at the edges of an infrared image. In [25,26], the authors

considered the low-level vision problems where a priori target rank information is available in advance, and minimized the partial sum of singular values instead of minimizing the nuclear norm [27]. In [25–27], the individual frames are stacked as columns of a matrix before performing RPCA and each frame is seen as an independent entity. RPCA-based background modeling for static video sequence assumes that these background variations are low-rank and the foreground activity is sparse due to spatially localized. When the background changes fast, it is difficult to obtain accurate backgrounds and then the target regions.

## 3. The Proposed Method

In this section, we aim to design an infrared small target detection system which consists of two parts as shown in Figure 1. The first part aims to extract highly suspected targets in each frame, and the second part is to confirm true targets from highly suspected targets. The first step of the first part is to suppress complex backgrounds (such as clouds, plants, and strong noises), and detect the suspected targets from a single frame using dense and sparse computation models separately. The second step of the first part is to associate the dense and sparse reconstruction maps obtained by the two computation models in the first step, and suppress false alarms in each frame. Repeating the first step and the second step for consecutive frames, many single-frame detection results can be obtained accordingly. So the second part is to refine the single-frame detection results of different frames based on tracklet association using target location and appearance features.
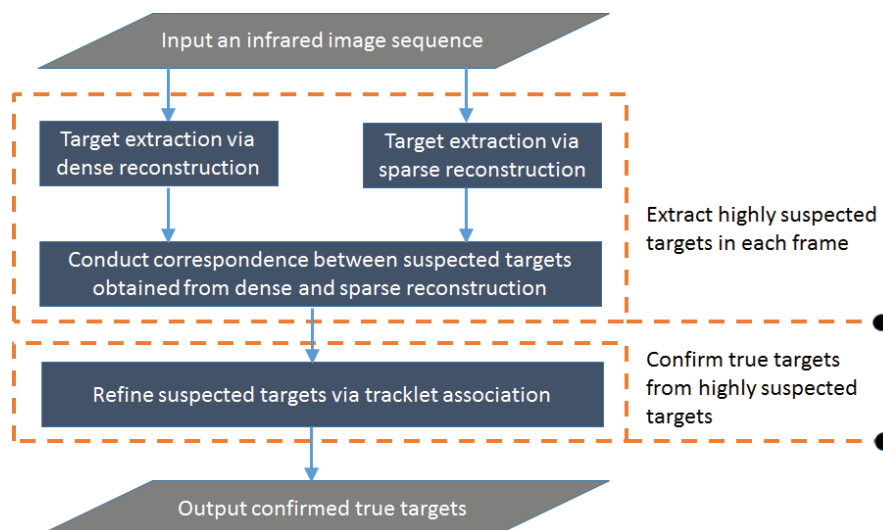


**Figure 1.** Flow chart of the proposed method.

### 3.1. Infrared Dim Target Model

An infrared image can be usually described as [8]

$$F(x,y) = B(x,y) + D(x,y) + N(x,y) \tag{1}$$

where $(x,y)$ denotes the coordinate of a pixel, $F(x,y), B(x,y), D(x,y)$ and $N(x,y)$ are the pixel intensity at coordinate $(x,y)$ for the original infrared image, the background image, the target image, and the random noise image respectively. We can get dense and sparse reconstruction maps from a single image, depending on whether there is a sparse constraint on the target image $D$ in the decomposition process.

### 3.1.1. Frequency Analysis of Infrared Images

It is well known that the singular value decomposition is a powerful tool, and is widely used in latent semantic analysis, recommendation system, defect detection, background suppression and so on [28,29]. The SVD of an infrared image $F$ with size $m \times n$ can be defined as

$$F = \sum_{i=1}^{r} \sigma_i u_i v_i^T \tag{2}$$

where $\{u_i\}$, $\{v_i\}$ and $\{\sigma_i\}$ are the left singular vectors, right singular vectors, and singular values respectively, and $\{\sigma_i\}$ is arranged in descending order.

We can find from (2) that an infrared image $F$ can be represented as a sum of different frequency components $\{u_i v_i^T\}$ regularized by $\{\sigma_i\}$, and the low frequency components correspond to the background part $B$ which always changes quite slowly, the medium frequency components correspond to the target part $D$ which usually appears as a bright area, and the high frequency components correspond to the noise part $N$. So, each part of an infrared image can be obtained by regularizing proper singular values for different purposes, such as background approximation with low pass filtering, and target detection with band pass filtering.

### 3.1.2. Low-Rank Analysis of Infrared Images

As discussed above, the infrared background image $B$ usually changes slowly and occupies most part of the original image $F$, and the image columns (rows) have the property of non-local self-correlation, i.e., columns (rows) are approximately linearly correlated with each other. Thus the background image $B$ can be well approximated by a low-rank matrix.

For a target image, the total number of target pixels is far less than the total number of pixels in the whole image, because of the small size of each target (usually no more than $10 \times 10$). So it is reasonable to assume that the target image is sparse.

Based on low-rank property of the background image and sparse property of the target image, the low-rank decomposition model can be used to build up a separation model for small targets and complex backgrounds.

### *3.2. Single Frame Target Extraction*

### 3.2.1. Target Extraction via Dense Reconstruction

A dense reconstruction model is developed here to extract target regions in an infrared image. For an original infrared image $F$, we can construct filters $v_i$ (orthogonal bases of row space) by performing SVD on $F$. Based on these filters, we can compute a one-dimensional response $R_i^r$ by filtering infrared image $F$

$$R_i^v = \frac{1}{\sigma_i} F v_i = u_i \tag{3}$$

Similarly, we can obtain a one-dimensional response $R_i^u$ by filtering infrared image $F$ with filters $u_i$ (orthogonal bases of column space)

$$R_i^u = \frac{1}{\sigma_i} F^T u_i = v_i \tag{4}$$

Combining Equations (3) and (4), we can compute a two-dimensional response $R_i$

$$R_i = R_i^v (R_i^u)^T = u_i v_i^T \tag{5}$$

Equations (3)–(5) show that a pair of filters $(u_i, v_i)$ can be used to generate a two-dimensional response $R_i$ which just corresponds to one frequency component of infrared image $F$. To better

utilize multiple frequencies of $F$ to extract target information, we choose to combine the obtained two-dimensional responses in a reasonable bound together to obtain the final target map

$$D^{\mathrm{d}} = \sum_{i=l}^{h} \rho_i R_i = \sum_{i=l}^{h} \rho_i u_i v_i^{\mathrm{T}} \tag{6}$$

where $l$ denotes the low cut subscript, $h$ denotes the high cut subscript, and $\rho_i$ denotes a linear weight or non-linear weight. When $\rho_i$ equals to $\sigma_i$, Equation (6) degenerates to Equation (2), up to the low and high cut subscripts. In Equation (6), $R_i$ denotes a two-dimensional response of dense filters $v_i$ and $u_i$, and all its elements is weighted by only one weight $\rho_i$, thus the map $D^{\mathrm{d}}$ is a global dense representation for infrared targets.

For each response $R_i$, the background clutter is still a key factor to influence the final target map, thus the corresponding weight $\rho_i$ should be regularized to suppress the background further. In Equation (6), weight $\rho_i$ is defined as the logistic sigmoid function of $\sigma_i$

$$\rho_i = \frac{1}{1 + exp(-\sigma_i)} \tag{7}$$

As mentioned above, each singular value indicates the ability of the corresponding response to approximate the original image. Thus, the singular values can be used to estimate parameters $l$ and $h$. Note that the first component always corresponds to the main part of the background, so we do not consider the first singular value in computing $l$ and $h$. Let $\widehat{\sigma}_i = \sigma_i - \sigma_{i+1}$, $i = 2, \ldots, r-1$, and $\overline{\sigma}_1 = \frac{1}{r-1}(\widehat{\sigma}_2 + \cdots + \widehat{\sigma}_{r-1})$, we can compute $l$ using the following equation

$$l = \max_i \ \{\widehat{\sigma}_i > \zeta_1 \overline{\sigma}_1\}_{i=1}^{r-1} + 1 \tag{8}$$

A similar consideration can be used to obtain the parameter $h$. Let $\overline{\sigma}_2 = \frac{1}{r-l}(\widehat{\sigma}_l + \cdots + \widehat{\sigma}_{r-1})$, we can calculate $h$ using the following equation

$$h = \max_i \ \{\widehat{\sigma}_i > \zeta_2 \overline{\sigma}_2\}_{i=l}^{r-1} \tag{9}$$

$\zeta_1$ and $\zeta_2$ are scaling factors.

For the final target map $D^{\mathrm{d}}$, a small part of background clutter and noise is still required to be removed, because $D^{\mathrm{d}}$ comprises of a series of responses $\{R_i\}$ which is regularized by a global weight $\rho_i$. In fact, the remaining clutter and noise is not necessarily to be gaussian. Therefore, we use Chebyshev's theorem to remove the clutter and noise in $D^{\mathrm{d}}$, and set $Th = \mu + c\tau$ as the global threshold, where $\mu$ and $\tau$ denote the mean and standard deviation of $D^{\mathrm{d}}$, and $c$ is a positive number and denotes the multiple of standard deviations [30].

### 3.2.2. Target Extraction via Sparse Reconstruction

As discussed above, it can be concluded from the view of low-rank representation that the task of target map computation can be formulated as a convex optimization problem:

$$\min_{B,D} \|B\|_* + \lambda \|W \circ D\|_1, \quad s.t. \quad \|F - B - D\|_{\mathrm{F}}^2 \leq \epsilon \tag{10}$$

where $\|.\|_*$ denotes the nuclear norm of a matrix, $\|.\|_{\mathrm{F}}$ is the Frobenius norm of a matrix, $\|.\|_1$ represents the sum of absolute values of matrix elements, $\circ$ denotes the entrywise product of the weighting matrix $W$ and the target image $D$, $\lambda > 0$ is a regularization parameter which controls the tradeoff between the background image $B$ and the target image $D$, and $\epsilon > 0$ is the upper bound of noise energy.

By introducing a multiplier $\mu > 0$, the optimization problem (10) can be relaxed as:

$$\min_{B,D} \|B\|_* + \lambda\|W \circ D\|_1 + \frac{1}{2\mu}\|F - B - D\|_F^2 \tag{11}$$

It can be shown that for some proper value $\mu(\epsilon)$, any solution of (11) is equivalent to the solution of (10) [27]. To achieve superior convergence, the accelerated proximal gradient (APG) algorithm with a continuation technique on $\mu$ is used to solve (11) [31–33]. The convex optimization problem (11) can be decomposed into two subproblems that minimize $B$ and $D$ respectively (for details please see the appendix):

$$B_{k+1} = \arg\min_B \{\|B\|_* + \frac{1}{\mu}\|B - Y_k^B + \frac{1}{2}(Y_k^B + Y_k^D - F)\|_F^2\} \tag{12}$$

$$D_{k+1} = \arg\min_D \{\lambda\|W \circ D\|_1 + \frac{1}{\mu}\|D - Y_k^D + \frac{1}{2}(Y_k^B + Y_k^D - F)\|_F^2\} \tag{13}$$

The subproblems (12) and (13) can be solved by the following equations respectively [34,35]

$$B_{k+1} = \mathcal{B}_{\frac{\mu}{2}}(Y_k^B - \frac{1}{2}(Y_k^B + Y_k^D - F)) \tag{14}$$

$$D_{k+1} = \mathcal{D}_{\frac{\mu\lambda W}{2}}(Y_k^D - \frac{1}{2}(Y_k^B + Y_k^D - F)) \tag{15}$$

where $\mathcal{B}_{\frac{\mu}{2}}(G_k^B) = U\text{diag}(\max\{\sigma_i - \frac{\mu}{2}, 0\})V^T$ in which $U, V$ and $\{\sigma_i\}$ are generated by the singular value decomposition of $G_k^B = Y_k^B - \frac{1}{2}(Y_k^B + Y_k^D - F)$, and $\mathcal{D}_{\frac{\mu\lambda W}{2}}(G_k^D) = \text{sign}(G_k^D)\max(\text{abs}(G_k^D) - \frac{\mu\lambda W}{2}, 0)$ in which $G_k^D = Y_k^D - \frac{1}{2}(Y_k^B + Y_k^D - F)$. The details of the solution via APG is described in Algorithm 1.

---

**Algorithm 1:** Target map extraction by APG.

    **Input:** Infrared image $F, \lambda, W$.

    **Output:** $B = B_k, D^s = D_k$

  1:   $B_0 = B_{-1} = \mathbf{0}; D_0 = D_{-1} = \mathbf{0}; t_0 = t_{-1} = 1; \mu_0 > 0; \overline{\mu} > 0$.

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

  3:      $Y_k^B = B_k + \frac{t_{k-1}-1}{t_k}(B_k - B_{k-1})$

  4:      $Y_k^D = D_k + \frac{t_{k-1}-1}{t_k}(D_k - D_{k-1})$

  5:      $B_{k+1} = \mathcal{B}_{\frac{\mu}{2}}(Y_k^B - \frac{1}{2}(Y_k^B + Y_k^D - F))$

  6:      $D_{k+1} = \mathcal{D}_{\frac{\mu\lambda W}{2}}(Y_k^D - \frac{1}{2}(Y_k^B + Y_k^D - F))$

  7:      $t_{k+1} = \frac{1}{2} + \frac{1}{2}\sqrt{4t_k^2 + 1}$

  8:      $\mu_{k+1} = \max(0.9\mu_k, \overline{\mu})$

  9:      $k = k + 1$

 10:  **end while**

---

The computation of the entry-wise weight matrix $W$ is based on the local difference features between the target and neighboring background, and these local difference features could be well measured by the local weighted entropy [19]. For a pixel $F(x, y)$ which has a small neighborhood containing $n$ kinds of gray values $f_1, f_2, \ldots, f_n$, its local weighted information entropy is expressed as

$$H(x, y) = -\sum_{i=l}^{n}(f_i - F(x, y))^2 p_i \log p_i, \ p_i = \frac{n_i}{m \times n} \tag{16}$$

where $n_i$ is the number of gray value $f_i$ in the neighborhood. Then $H$ is convolved with a two-dimensional Gaussian operator $Ga(x,y) = \exp\{-\frac{1}{2\sigma_0^2}(x^2 + y^2)\}$ to generate a new version $\overline{H}$. Consequently the weighting matrix W is computed as $\alpha(1 - \overline{H})$, where $\alpha$ denotes a regularization parameter which controls the prior impact in the weight matrix.

### 3.2.3. Target Confirmation via Location Correlation

In order to extract as many highly suspected targets as possible, we use the location cue to remove false alarms and validate candidates in each frame. The intuition is that real targets though obtained by different methods should be located at the same position, but the random noise is not necessarily so. Therefore, through location correlation, some random noise should be suppressed and the ones occurring both in the dense and sparse reconstruction maps may be real targets with high probability.

Suppose that the regions obtained from $D^d$ and $D^s$ are denoted by $G_j^d$ with coordinate set $\mathcal{X}_j^d$ and $G_i^s$ with coordinate set $\mathcal{X}_j^s$ respectively, each region corresponding to a suspected target, and that $D^c = \mathbf{0}$ has the same size with $D^d$ and $D^s$. The main steps of target confirmation are described as follows:

1. For each $G_i^s$ in the target map $D^s$, we find $G_j^d$ in $D^d$ whose coordinates are overlapped with that of $G_i^s$, namely $\mathcal{X}_j^d \cap \mathcal{X}_i^s \neq \varnothing$.

2. For each successfully correlated pair $(G_i^s, G_j^d)$, we select pixels in $D^d$ with coordinates $\mathcal{X}_i^s$ as the correlation result, namely $D^c(\mathcal{X}_i^s) = D^d(\mathcal{X}_i^s)$.

Note that in step 2, we choose the pixels with coordinates $\mathcal{X}_i^s$ in the dense reconstruction map $D^d$. This step could not only avoid the drawback of L1-norm target measure, which treats each pixel independently in terms of intensity and weakens the intensities of the boundary target pixels, but also avoid the drawback of dense reconstruction, which enlarges the target area by combining nearby false alarms together. In essence, the dense reconstruction extracts suspected targets from the view of L2 norm that measures the minimal residual, and thus the boundary target pixels obtained from the dense reconstruction have more bright values than the ones obtained by the sparse reconstruction.

### 3.3. Multiple Frame Target Refinement

After obtaining candidates from each target map $D_k^c$, we generate suspected target tracklets based on target location and appearance features of consecutive frames, then perform non-maxima suppression to remove false tracks formed by noisy or clutter, finally refine the suspected targets in each target map $D_k^c$ according to the obtained tracks.

Suppose that the candidates in the target map $D_p^c$ and $D_k^c$ are denoted by $G_j^p = \{x_j^p, y_j^p, E_j^p\}$ and $G_i^k = \{x_i^k, y_i^k, E_i^k\}$ respectively. Note that the italic symbol $k$ denotes the image index, and the upright symbol p indicates the association result before the *k*th frame. For each candidate, $E$ represents the energy of pixels, and $(x, y)$ denotes the average coordinate of pixels. Hence, the tracklets can be generated by repeating linking $\{G_j^p\}$ and $\{G_i^k\}$ together, whereas $\{G_j^p\}$ is updated with each association. The details of generating tracklets are described as follows:

1. We set $D_p^c = D_1^c, \{\mathcal{X}^t = (x_1^t, y_1^t)\}, k = 2$.
2. For each $G_i^k$ in the target map $S_k^c(k \leq L)$, we select $\{G_j^p\}$ in $D_p^c$ within a circular gate of $G_i^k$.
3. For each $G_i^k$, if $E_i^k > E_\theta$, we keep $G_i^k$ in $D_k^c$ and generate $\{((x_j^p, y_j^p), (x_i^k, y_i^k))\}$ as small tracklets, here $E_\theta$ denotes the average energy of all candidate regions.
4. Update each track $\mathcal{X}^t$ with a proper tracklet selected from the set $\{((x_j^p, y_j^p), (x_i^k, y_i^k))\}$.
5. We set $D_p^c = D_k^c, k = k + 1$.
6. Repeating the above steps until $k$ is greater than $L$, we finally obtain tracks $\{\mathcal{X}^t\}$.

After obtaining trajectories through the above process, a non-maximum suppression (NMS) scheme is used to prune false trajectories. For each trajectory $\mathcal{X}_t$, a displacement gain measure $\Delta_t^1$ and a length ratio measure $\Delta_t^2$ are computed as

$$\Delta_t^1 = \sum_i |x_i^t - x_{i-1}^t| + |y_i^t - y_{i-1}^t| \tag{17}$$

$$\Delta_t^2 = \frac{\text{len}(\mathcal{X}_t)}{L} \tag{18}$$

where $(x_i^t, y_i^t)$ and $(x_{i-1}^t, y_{i-1}^t)$ denote elements of track $\mathcal{X}_t$, and $\text{len}(\cdot)$ is a function to solve the length of track $\mathcal{X}_t$. Next, the trajectories that not only have greater displacement gain than a threshold $\theta_1$ but also have longer length gain than a threshold $\theta_2$, are selected. The intuition is that real targets and false alarms have different motion features, and the trajectories formed by false alarms are diverse from the true trajectories produced by real targets. Finally, each suspected target in the target map $D_k^c$ is refined by measuring the distance of its centroid to the valid tracks, and the final detection result $D_k^r$ is obtained accordingly.

## 4. Experimental Section

### 4.1. Experimental Configuration

#### 4.1.1. Data Sets

In order to fairly evaluate the performance of infrared detection methods, a representative data set consisting of three public infrared sequences with different complex backgrounds is used and the detailed features are listed in Table 1. In Sequence 1, the detection is influenced by strong noise, plants and trees [36]. In Sequence 2, the background changes rapidly due to the movement of imaging platform, and a plane moves from the thick cloud region to the sky [37]. In Sequence 3, the detection barrier is noise and changing wispy clouds [11]. On a whole, the data set contains various situations in airborne infrared target detection. Therefore, using the given data set could fairly show the performance of infrared detection methods.

**Table 1.** Details of the evaluation data set.

| Sequences | Number | Size | Target Type | Target Details | Background Details |
|---|---|---|---|---|---|
| Sequence 1 | 600 | $320 \times 240$ | Helicopter | Low SCR value<br>A smaller size | Heavy noise backgrounds<br>Skyline and plants backgrounds |
| Sequence 2 | 30 | $256 \times 200$ | Airplane | A long imaging distance<br>Keeping curved movement<br>A changing size with a big range | Heavy clouds backgrounds<br>Uniform backgrounds |
| Sequence 3 | 40 | $256 \times 200$ | Airplane | A long imaging distance<br>Keeping zigzag-shaped movement | Wispy cloud backgrounds<br>Uniform backgrounds |

#### 4.1.2. Evaluation Metrics

The main objective of the proposed method is to effectively suppress the background noise and clutters, and then significantly reduce false alarms to improve detection performance. In this paper, the metrics of signal-to-clutter ratio gain (SCRG), background suppression factor (BSF), precision, recall and F-measure (PRF) are used to evaluate the performance of infrared detection methods. More specifically, the SCRG measures the enhancement of targets relative to the backgrounds before and after detection, and is defined as follows [38,39]:

$$\text{SCRG} = \frac{\text{SCR}_{\text{out}}}{\text{SCR}_{\text{in}}} \tag{19}$$

where $\mathrm{SCR_{out}}$ and $\mathrm{SCR_{in}}$ are the local SCR values computed from the filtered and original images respectively. Moreover, the BSF evaluates the background inhibition degree, and is defined by [40]:

$$\mathrm{BSF} = \frac{\delta_{\mathrm{in}}}{\delta_{\mathrm{out}}} \tag{20}$$

where $\delta_{\mathrm{in}}$ and $\delta_{\mathrm{out}}$ denote the standard deviations of the original and processed images respectively. In addition, the precision and recall reflect the false alarm rate and miss rate respectively, and the F-measure is the weighted harmonic mean of precision and recall [41].

### 4.2. Overall Performance of the Proposed Method

In subsection of dense reconstruction, the scaling factors $\zeta_1$ and $\zeta_2$ are both set to 1, and the multiple $c$ of standard deviations from the mean is set to 3. In subsection of sparse reconstruction, the window of local weighted entropy is set to $5 \times 5$, $\lambda$ is set to $1/\mathrm{sqrt}(\max(m, n))$, $\mu_0$ and $\overline{\mu}$ are set to $\sigma_3$ and $\frac{1}{r-2} \sum_{i=3}^{r} \sigma_i$ respectively, $\alpha$ is set to 2, and $\sigma_0$ and the size for the Gaussian filter are set to $\frac{1}{36} \min(m, n)$ and $\frac{1}{6} \min(m, n)$ respectively. In addition, in subsection of multiple frame target refinement, the circular gate is set to 15, the displacement gain is set to $\frac{1}{2} \max(\Delta_t^1)$, and the length threshold is set to 0.6. we first verify the validity of all the procedures of the proposed method, and test the proposed method on all the sequences in the data set with the same configuration parameters. More specifically, the visual illustrations of the dense reconstruction map (DRM) $D_k^{\mathrm{d}}$, the sparse reconstruction map (SRM) $D_k^{\mathrm{s}}$, and the location correlation map (LCM) $D_k^{\mathrm{c}}$ of $D_k^{\mathrm{d}}$ and $D_k^{\mathrm{s}}$, and the refined map (RM) $D_k^{\mathrm{r}}$ are shown in Figure 2, and the quantitative evaluations are given in Table 2 and Figure 3. Note that in Figure 2 the red circles denote the detected real targets, and the blue circles denote some target-like false alarms.

As depicted in Figure 2, $D_k^{\mathrm{d}}$ and $D_k^{\mathrm{s}}$ display the ability to reveal the small target in the preliminary results. In addition, the location correlation map $D_k^{\mathrm{c}}$ of $D_k^{\mathrm{d}}$ and $D_k^{\mathrm{s}}$ is shown to contribute to the results further by suppressing the false alarms and enhancing the target. Consistently, as illustrated in Table 2, the SCRG and BSF of $D_k^{\mathrm{c}}$ was a good tradeoff between that of $D_k^{\mathrm{d}}$ and $D_k^{\mathrm{s}}$. Although $D_k^{\mathrm{s}}$ obtained higher SCRG than $D_k^{\mathrm{d}}$ did on Sequence 3, it obtained lower SCRG than $D_k^{\mathrm{d}}$ did on Sequence 1 and Sequence 2, because the very challenging dataset based on an actual application is diverse and characteristic. Hence, a proper tradeoff based on location correlation is necessary for suppressing false alarms in $D_k^{\mathrm{d}}$ and keeping target pixels in $D_k^{\mathrm{s}}$. However, after correlation of $D_k^{\mathrm{d}}$ and $D_k^{\mathrm{s}}$, there still exist highly suspected targets in the correlation map $D_k^{\mathrm{c}}$, and some representatives are labeled by blue circles. So in this paper, we use multiple frame target refinement to suppress these false alarms. From the visual and quantitative results in Figures 2 and 3 and Table 2, the refined map $D_k^{\mathrm{r}}$ produced the best target detection performance.

**Table 2.** Average scores of the procedures of the proposed method.

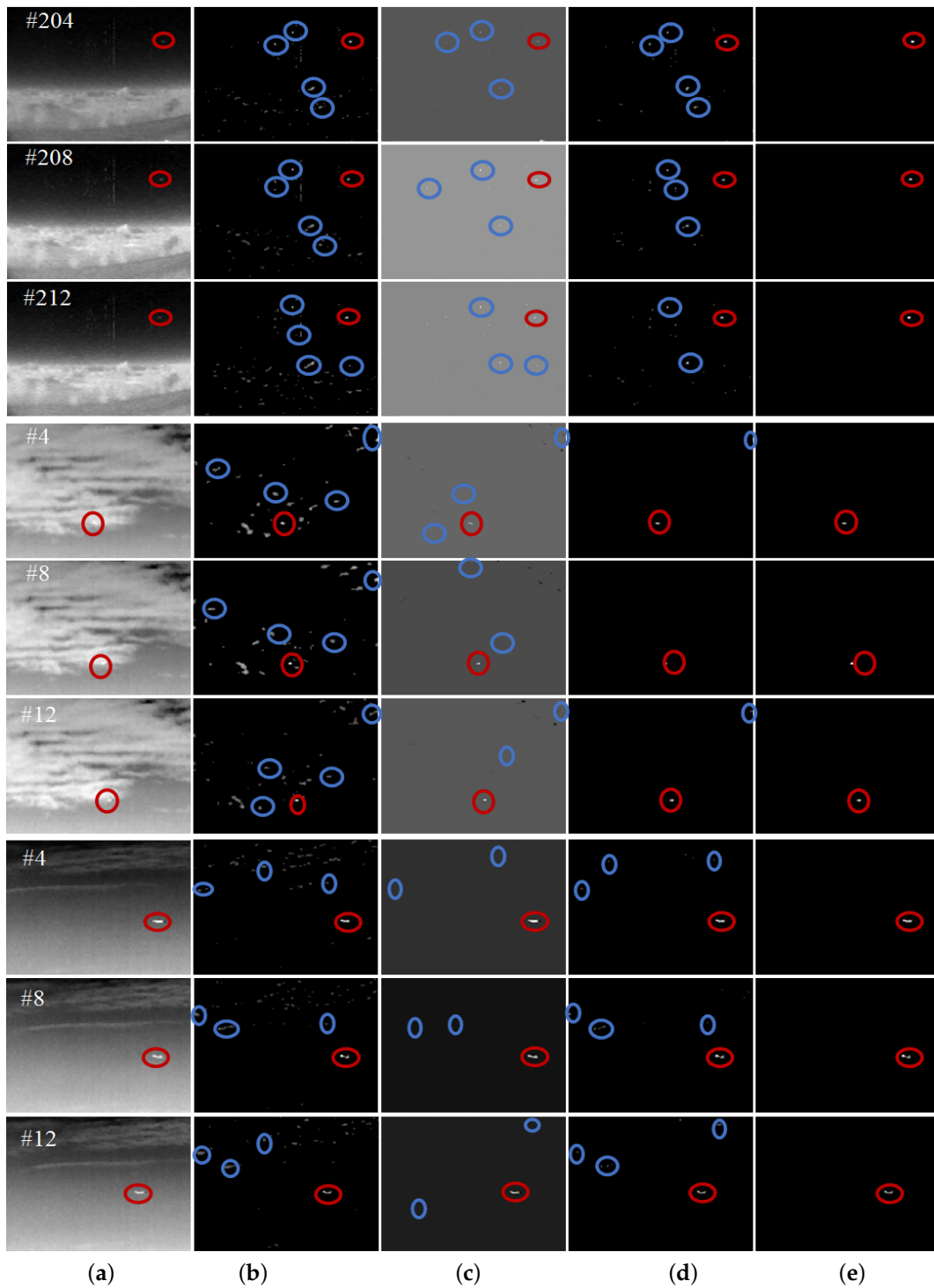| Sequences | Metrics | DRM | SRM | LCM | RM |
|---|---|---|---|---|---|
| Sequence 1 | SCRG | 2.37 | 1.71 | 2.27 | 2.31 |
|  | BSF | 7.01 | 25.44 | 10.91 | 19.12 |
| Sequence 2 | SCRG | 3.25 | 3.14 | 3.60 | 3.60 |
|  | BSF | 1.89 | 9.20 | 6.04 | 6.51 |
| Sequence 3 | SCRG | 1.33 | 1.38 | 1.34 | 1.34 |
|  | BSF | 4.19 | 8.37 | 6.67 | 7.84 |

**Figure 2.** Visual illustration of the procedures of the proposed method. The first three rows, the middle three rows, and the last three rows illustrate the results of the proposed method on Sequence 1, Sequences 2 and Sequence 3 respectively. (**a**) denotes the original images selected from three sequences, (**b**) denotes the dense reconstruction map $D_k^{\mathrm{d}}$, (**c**) denotes the sparse reconstruction map $D_k^{\mathrm{s}}$, (**d**) denotes the location correlation map $D_k^{\mathrm{c}}$ of $D_k^{\mathrm{d}}$ and $D_k^{\mathrm{s}}$, and (**e**) denotes the refined map $D_k^{\mathrm{r}}$.
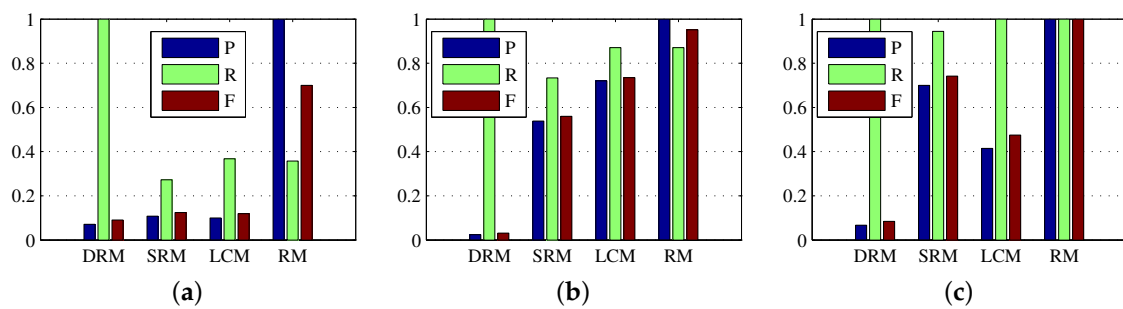
**Figure 3.** Precision, recall, and F-measure bars for the procedures of the proposed method. (**a**–**c**) denote the results of the procedures of the proposed method on Sequence 1, Sequences 2 and Sequence 3 respectively.

## 4.3. Parameters Analysis

For the dense reconstruction model, the main relevant parameters include the regularization weight $\rho_i$ and multiple $c$. As previously discussed, the regularization weight $\rho_i$ helped to eliminate the background influence and enhance the target further, and the multiple parameter $c$ controlled the global threshold for target extraction in the dense reconstruction map $D^d$. To show the benefits of using $\rho_i$ rather than $\sigma_i$ in the dense reconstruction model, we consider several combinations $(\rho_i, c)$ and $(\sigma_i, c)$ in the experiment, and set $c = 1, 3, 5, 7$. As seen from Figure 4a–d and the first four rows of Table 3, for Sequence 1 and Sequence 2, the combinations $(\rho_i, c)$ helped target map $D^d$ obtain the better precision, recall, and F-measure, and achieve the higher average SCRG and BSF values. Note that the symbol $\infty$ in Table 3 indicates that the target map obtained by the combination $(\sigma_i, c = 7)$ is a zero matrix which means that there is no target or background information in the target map, hence the choice of $\rho_i$ is superior to $\sigma_i$ in the process of target reconstruction. In addition, we found from Figure 4e–f and the last two rows of Table 3 that there was no diverse difference between the results obtained by using combinations $(\rho_i, c)$ and $(\sigma_i, c)$ on Sequence 3, and the combination $(\rho_i, c)$ obtained a slightly worse result than $(\sigma_i, c)$. The reason is that $\rho_i$ could also suppress the target when suppressing the background, and the background in Sequence 3 is not as complex as that in Sequence 1 and Sequence 2. As a whole, the regularization weight $\rho_i$ is more suitable for more complex background suppression than $\sigma_i$. As mentioned, $c$ is also an important parameter. As shown in Table 3, the BSF increased with increasing $c$ value, while the SCRG decreased with increasing $c$ value. Hence, an intermediate $c$ value could be a good choice. The consistent conclusion can be seen from Figure 4, because for example the case $c = 7$ could cause high miss rate while the case $c = 1$ could result in lots of false alarms.

**Table 3.** Average scores of the dense reconstruction maps with different combinations.

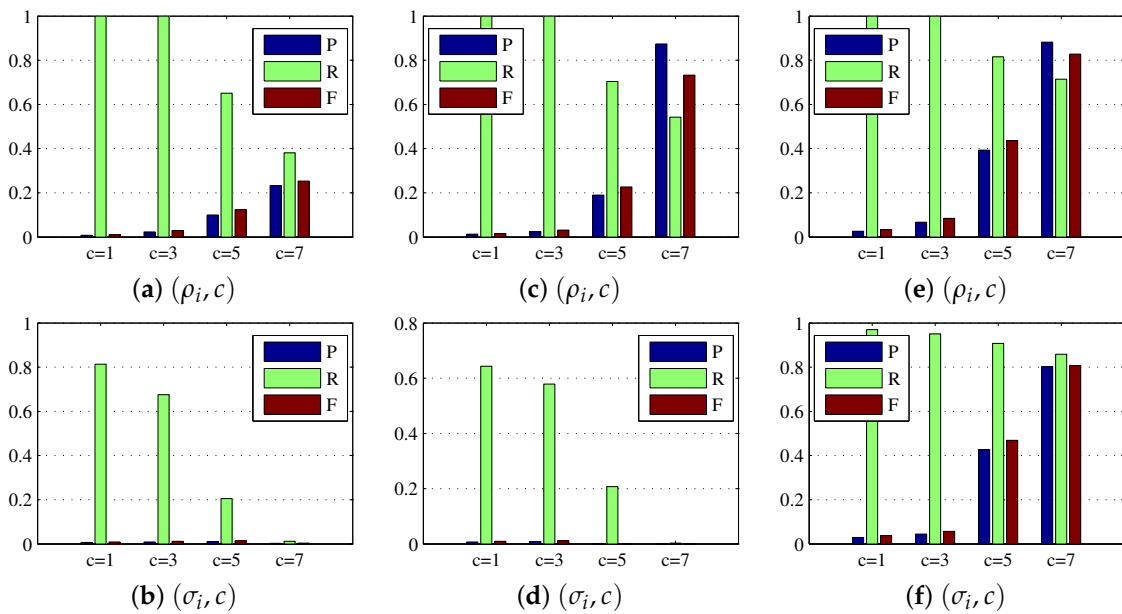| Sequences | Metrics | $\rho_i\ (\sigma_i), c = 1$ | $\rho_i\ (\sigma_i), c = 3$ | $\rho_i\ (\sigma_i), c = 5$ | $\rho_i\ (\sigma_i), c = 7$ |
|---|---|---|---|---|---|
| Sequence 1 | SCRG | 2.86(2.72) | 2.60(1.99) | 2.27(0.97) | 1.77(0.08) |
|  | BSF | 4.49(2.56) | 7.01(3.42) | 12.05(5.08) | 18.44(8.81) |
| Sequence 2 | SCRG | 2.83(1.79) | 3.25(1.60) | 3.20(1.12) | 2.74(0.03) |
|  | BSF | 1.17(0.50) | 1.89(0.77) | 4.14($\infty$) | 6.83($\infty$) |
| Sequence 3 | SCRG | 1.37(1.49) | 1.33(1.43) | 1.25(1.39) | 1.18(1.34) |
|  | BSF | 2.86(3.87) | 4.19(4.71) | 6.50(6.24) | 7.90(6.81) |

**Figure 4.** Precision, recall, and F-measure bars for the dense reconstruction maps under different combinations $(\rho_i, c)$ and $(\sigma_i, c)$. (**a,b**) denote the results on Sequence 1, (**c,d**) denote the results on Sequence 2, and (**e,f**) denote the results on Sequence 3.

For the sparse reconstruction model, the regularization parameter $\alpha$ controlled the prior impact in the weight matrix. We varied $\alpha$ from 1 to 4 in the experiment, and illustrated the precision, recall, and F-measure in Figure 5. From the illustration, it could be observed that an intermediate $\alpha$ would give a better tradeoff between miss rate and false alarms. For example, the precision, recall, and F-measure of $\alpha = 4$ showed a high miss rate on Sequence 1 and Sequence 2, because the real targets were suppressed by the overlarge prior weight. In contrast, when the low miss rate was guaranteed, the false alarm rate of $\alpha = 1$ is higher than other settings, suggesting that a too small $\alpha$ is also not a proper choice.
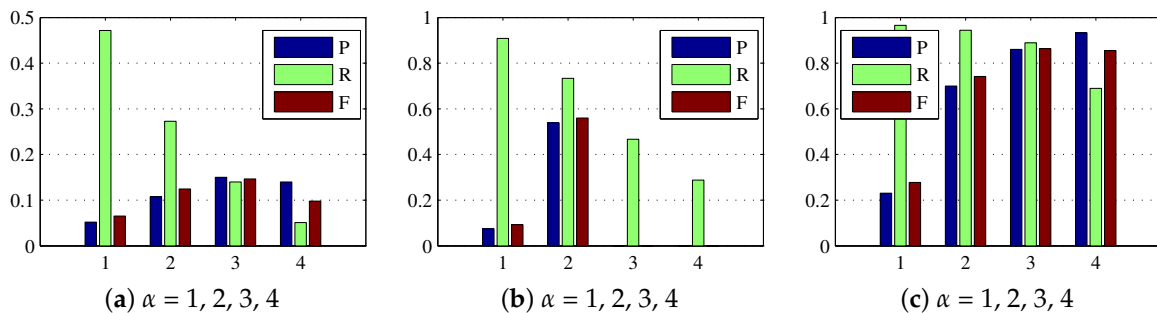


**Figure 5.** Precision, recall, and F-measure bars for the sparse reconstruction maps under different values $\alpha$. (**a**–**c**) denote the results on Sequence 1, Sequences 2 and Sequence 3 respectively.

As described in subsection of multiple frame target refinement, a trajectory comprises of small tracklets which mainly depend on the gate size $\delta$ of $G_i^k$. We varied the gate size $\delta$ from 5 to 25 in the experiment, and showed the relationship between the trajectory detection rate and gate size in Figure 6. It could be observed that a small gate size $\delta < 10$ could result in a low trajectory detection rate on Sequence 2, a small gate size $\delta < 15$ could cause a low trajectory detection rate on Sequence 3, and a large gate size $\delta > 20$ could lead to a low trajectory detection rate on Sequence 1, because the real targets in Sequence 1, Sequence 2, and Sequence 3 have different velocities which can be inferred from Figure 2. For example, because the target in Sequence 1 has the lowest velocity, a very large gate

size could lead to false association between the current frame and the former frame, and moreover the target in Sequence 3 has the highest velocity, thus a very small gate size could result in association failure between consecutive frames. From the illustration, we can find that the interval [15,20] could be a proper scope of the gate size for Sequence 1, Sequence 2, and Sequence 3.
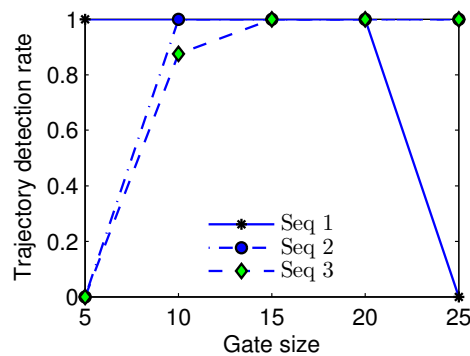


**Figure 6.** Relationship curve of the trajectory detection rate and gate size.

### 4.4. Comparison with State-Of-The-Art Approaches

In order to show the performance of the proposed method, we selected five state-of-the-art infrared small target detection methods, including two classical methods (MM [18], TH [42]), and three recent methods (IPI [8], MPC [22], WLD [10]) based on the patch-image model, the multiscale contrast measure and the local difference measure. As a whole, the adopted comparison approaches are representatives of the advanced level of infrared small target detection at present. The detailed parameter settings used in the experiments are described in Table 4 for reproduction. The detection results of the proposed method and the existing methods are visually shown in Figure 7 in which the red circles denote the detected real targets, the blue circles denote the false alarms, and the green circles indicate that the real targets were lost. Note that the results of the existing methods in Figure 7 are obtained using a global threshold $\mu + 3\tau$, and the original images corresponding to Figure 7 are the same as those of Figure 2. In addition, the quantitative evaluation results are provided in Table 5 and Figure 8.

**Table 4.** Detailed parameter settings of the six test methods.

| No. | Methods | Acronyms | Parameter Settings |
|-----|---------|----------|--------------------|
| 1 | Max-Mean filter | MM | Support size: $5 \times 5$ |
| 2 | Top-Hat approach | TH | Structure element: disk, radius: 6 |
| 3 | Infrared patch image model | IPI | Patch size: $50 \times 50$, sliding step: 10, $\lambda = 1/\sqrt{\max(m,n)}$ |
| 4 | Multiscale patch-based contrast measure | MPC | N = 2, 3, 4, 5 |
| 5 | Weighted local difference measure | WLD | L = 4, entropy neighborhood $5 \times 5$ |
| 6 | Proposed co-detection method | COD | $\lambda = 1/\sqrt{\max(m,n)}$, entropy neighborhood $5 \times 5$ |

**Table 5.** Average scores of the proposed method and other existing methods.

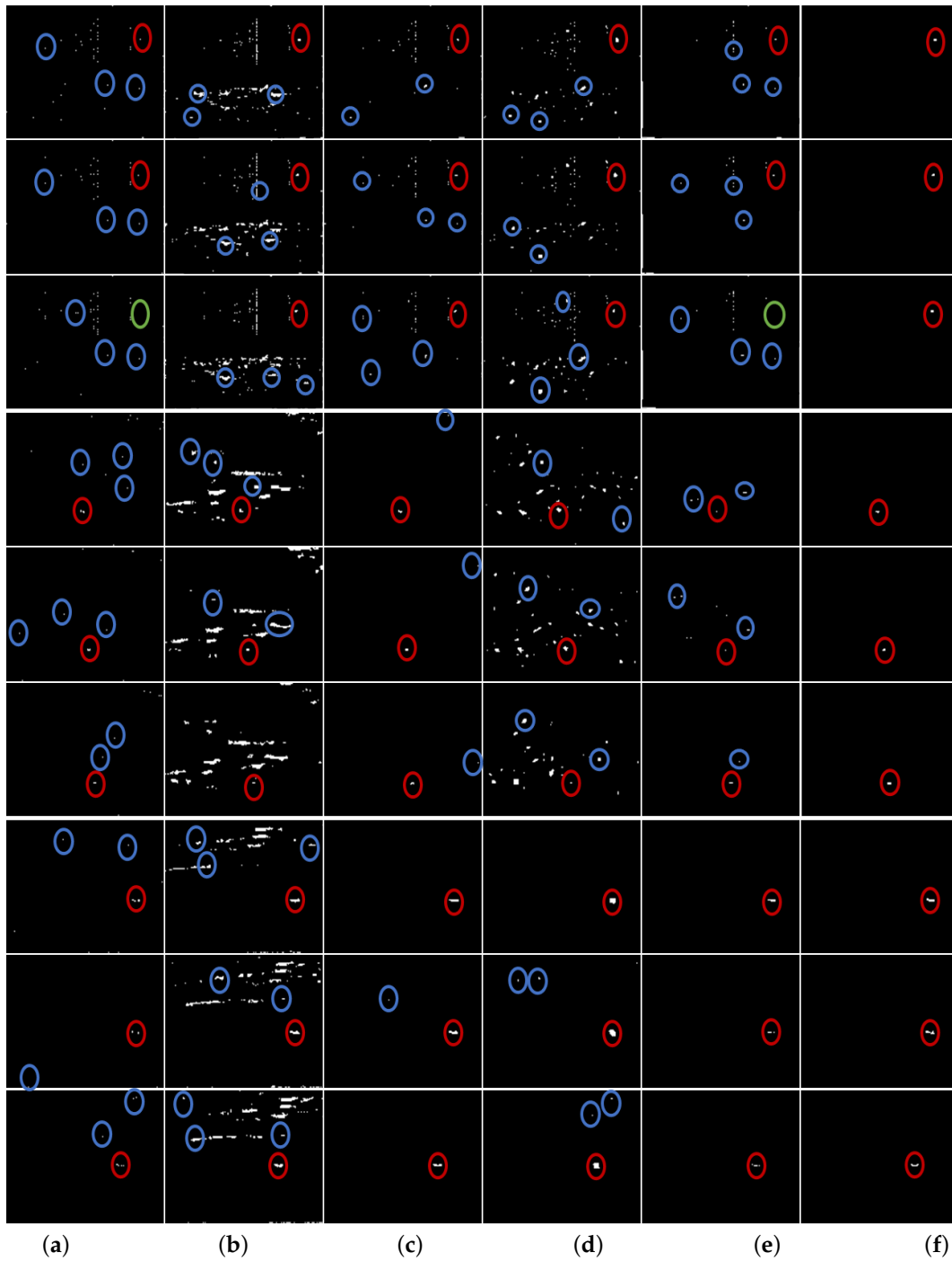| Sequences | Metrics | MM | TH | IPI | MPC | WLD | COD |
|-----------|---------|------|------|-------|-------|-------|-------|
| Sequence 1 | SCRG | 0.80 | 1.99 | 1.55 | 2.45 | 2.27 | 2.31 |
|  | BSF | 9.59 | 5.26 | 23.34 | 16.14 | 9.81 | 19.12 |
| Sequence 2 | SCRG | 2.10 | 2.67 | 3.13 | 2.91 | 2.35 | 3.60 |
|  | BSF | 2.30 | 0.77 | 9.39 | 7.44 | 8.36 | 6.51 |
| Sequence 3 | SCRG | 0.78 | 1.51 | 1.37 | 1.45 | 1.01 | 1.34 |
|  | BSF | 6.08 | 3.54 | 8.35 | 7.54 | 10.37 | 7.84 |

**Figure 7.** Visual illustration of the results of the proposed method and other existing methods. The first three rows, the middle three rows, and the last three rows illustrate the results of the proposed method and other existing methods on Sequence 1, Sequences 2 and Sequence 3 respectively. (**a**) denotes the detection results of MM method, (**b**) denotes the detection results of TH method, (**c**) denotes the detection results of IPI model, (**d**) denotes the detection results of MPC method, (**e**) denotes the detection results of WLD method, and (**f**) denotes the detection results of the proposed method.
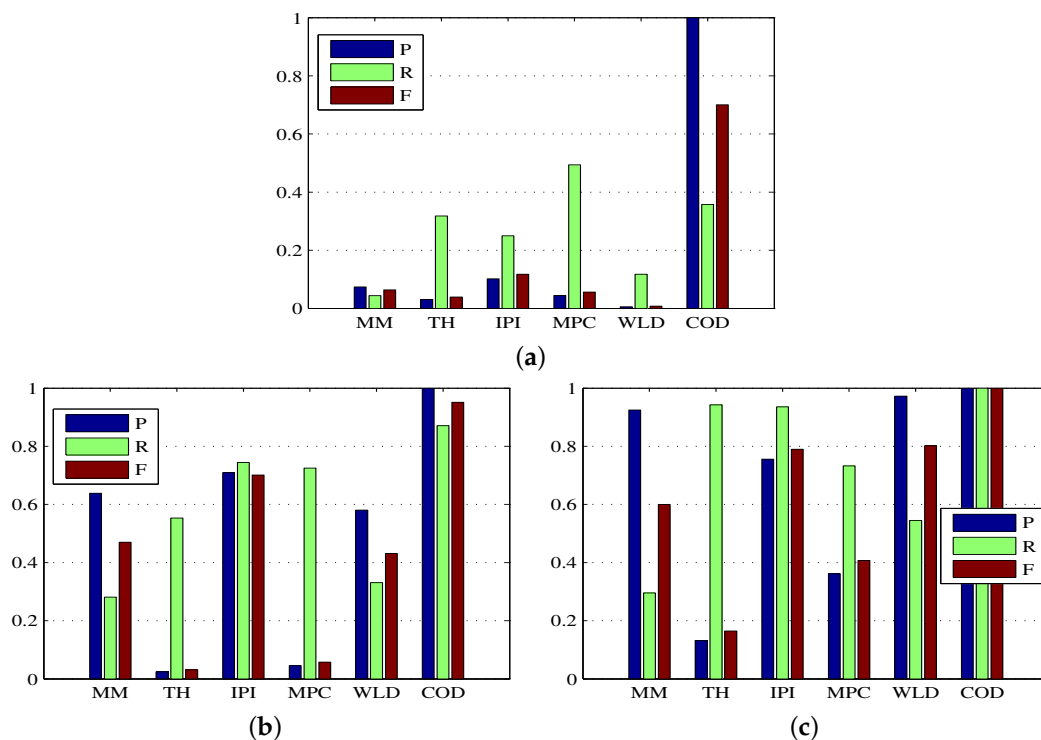
**Figure 8.** Precision, recall, and F-measure bars for the proposed method and other existing methods. (**a**–**c**) denote the results of the proposed method and other existing methods on Sequence 1, Sequences 2 and Sequence 3 respectively.

As depicted in Table 5, the MM achieved higher BSF than TH on Sequence 1, Sequence 2 and Sequence 3, but gave the worst performance in terms of the SCRG and the visual illustration of Figure 7. In addition, the TH and MPC achieved the top two SCRG on Sequence 3, but gave the poor detection results as seen from the visual result of Figure 7 that many target pixels were lost in the detection result of MPC, and that much background cloud was left in the detection of TH. Moreover, the IPI model exhibited excellent background suppression performance, but the SCRG is very low when the background was complex in Sequence 1. Furthermore, the WLD achieved the highest BSF on Sequence 3, but obtained less SCRG which was slightly better than that of MM. As shown in Figures 7 and 8b–c, for the IPI and WLD methods, the false alarms can be well suppressed with an appropriate threshold on Sequence 2 and Sequence 3. However, as shown in Figures 7 and 8a, the existing five state-of-the-art methods still performed poorly on Sequence 1, because a very challenging test sequence based on an actual application (existence of targe-like false alarms as shown in Figure 7) was used for the comparative testing in this paper. Thus, the success of the existing five methods based on only a single image was restricted to its own specific application. Therefore, exploitation of the motion and appearance cues from the image sequences was necessary to further improve the detection performance of a single frame. Although the proposed method do not have the highest SCRG and BSF, based on the visual comparison in Figure 7 and the quantitative comparison in Figure 8, it is clear that the proposed method consistently performed well on all three sequences and outperformed other test methods from the view that all false alarms including the target-like ones in each image are well suppressed.

*4.5. Computational Complexity*

The computational complexity and time for the proposed method and other existing methods were given in Table 6. All the experiments were carried out on a computer with a 3.2 GHz Intel CPU and 4-GB memory. The image size is $m \times n$, the patch-image size in the IPI model is $\overline{m} \times \overline{n}$,

$\beta$ denotes the iteration number of the algorithm, and $p$ denotes the pixel number in the support region. As depicted in Table 6, the six test methods differed greatly in running time, though there existed a little difference in the computational complexity. For the MM and TH methods, the time difference is mainly caused by the max operation in MM. For the MPC and WLD methods, the time difference lies in the sort operation in the computation of local entropy. In addition, the IPI method took the longest time in the test methods, the cost is mainly caused by the vectorization and median operations. However, the running time of the proposed method is only about one-fiftieth the time of IPI method. The essential reason is that the non-patch scheme and local prior weight contribute to improving convergence speed. Although the proposed method took more time than the MM, TH and MPC methods did, it is acceptable from the view of detection performance.

**Table 6.** Algorithm complexity and running time of the proposed method and other existing methods.

|  | MM | TH | IPI | MPC | WLD | COD |
|---|---|---|---|---|---|---|
| Complexity | O($4mn$) | O($2pmn$) | O($\beta k \overline{mn} \log(\overline{mn}) + mnp$) | O($(p+1)Lmn$) | O($(L+p)mn$) | O($mn(p + k + \beta k \log(mn))$) |
| Time (s) | 8.30 | 0.17 | 503.65 | 0.28 | 18.20 | 9.2 |

## 5. Conclusions

In this paper, a novel infrared target co-detection model, which combines the self-correlation features of backgrounds and the commonality features of targets in the spatio-temporal domain, is proposed to detect small targets in a sequence of images with complex backgrounds. On one hand, the nonlinear weights has been constructed based on the logistic sigmoid function, and has more advantages than weights of singular values in suppressing background and keeping small targets. On the other hand, the entry-wise weight has been designed based on the local weighted entropy, and can extract real targets accurately and suppress background clutters efficiently. Finally, the location correlation of the dense and sparse reconstruction maps for a single frame and tracklet association of the location correlation maps for successive frames are performed to suppress false alarms and confirm suspected targets. The experiments have testified the effectiveness of the proposed co-detection model.

**Author Contributions:** The work presented here was carried out in collaboration with all of the authors. Jingli Gao and Chenglin Wen conceived the idea and research theme. Jingli Gao designed and performed the experiments, and wrote the paper. Meiqin Liu supervised the work and analyzed the experimental results.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

With $X = (B, D)^{\mathrm{T}}, g(X) = \|B\|_* + \lambda \|W \circ D\|_1$, and $f(X) = \frac{1}{2\mu} \|F - B - D\|_{\mathrm{F}}^2$ in (11), thus $\nabla_{\mathrm{B}} f(X) = \nabla_{\mathrm{D}} f(X) = \frac{1}{\mu}(B + D - F)$, and $\nabla f(X) = (\nabla_{\mathrm{B}} f(X), \nabla_{\mathrm{D}} f(X))^{\mathrm{T}}$, we can obtain

$$
\begin{aligned}
\|\nabla f(X_1) - \nabla f(X_2)\|_{\mathrm{F}} &= \frac{1}{\mu} \left\| \begin{bmatrix} B_1 + D_1 - F \\ B_1 + D_1 - F \end{bmatrix} - \begin{bmatrix} B_2 + D_2 - F \\ B_2 + D_2 - F \end{bmatrix} \right\|_{\mathrm{F}} = \frac{1}{\mu} \left\| \begin{bmatrix} B_1 - B_2 \\ D_1 - D_2 \end{bmatrix} + \begin{bmatrix} D_1 - D_2 \\ B_1 - B_2 \end{bmatrix} \right\|_{\mathrm{F}} \\
&\leq \frac{1}{\mu} \left\| \begin{bmatrix} B_1 - B_2 \\ D_1 - D_2 \end{bmatrix} \right\|_{\mathrm{F}} + \frac{1}{\mu} \left\| \begin{bmatrix} D_1 - D_2 \\ B_1 - B_2 \end{bmatrix} \right\|_{\mathrm{F}} = \frac{2}{\mu} \|X_1 - X_2\|_{\mathrm{F}}
\end{aligned}
\tag{A1}
$$

So $f$ is Lipschitz continuous, and the Lipschitz constant $L_f = \frac{2}{\mu}$. According to the proximal gradient approach, we could approximate $f(X)$ locally as a quadratic function

$$
\begin{aligned}
(B_{k+1}, D_{k+1}) = X_{k+1} &= \arg\min_X \{g(X) + f(Y_k) + <\nabla f(Y_k), X - Y_k> + \frac{L_f}{2}\|X - Y_k\|_F^2\} \\
&= \arg\min_X \{g(X) + \frac{1}{\mu}\|X - Y_k + \frac{\mu}{2}\nabla f(Y_k))\|_F^2\} \\
&= \arg\min_{B,D} \{\|B\|_* + \lambda\|W \circ T\|_1 + \frac{1}{\mu}\left\|\begin{bmatrix} B \\ D \end{bmatrix} - \begin{bmatrix} Y_k^B - \frac{\mu}{2}\nabla_B f(Y_k) \\ Y_k^D - \frac{\mu}{2}\nabla_D f(Y_k) \end{bmatrix}\right\|_F^2\} \\
&= \arg\min_{B,D} \{\|B\|_* + \frac{1}{\mu}\|B - Y_k^B + \frac{\mu}{2}\nabla_B f(Y_k))\|_F^2 + \lambda\|W \circ D\|_1 \\
&\quad + \frac{1}{\mu}\|D - Y_k^D + \frac{\mu}{2}\nabla_D f(Y_k))\|_F^2\},
\end{aligned}
\tag{A2}
$$

and then solve (A2) to update the solution *X*. It can conclude that the problem (A2) is separable and can be decomposed into two subproblems (12) and (13).

**References**

1. Li, Z.Z.; Chen, J.; Hou, Q.; Fu, H.X.; Dai, Z.; Jin, G.; Li, R.Z.; Liu, C.J. Sparse Representation for Infrared Dim Target Detection via a Discriminative Over-Complete Dictionary Learned Online. *Sensors* **2014**, *14*, 9451–9470.
2. Sanna, A.; Lamberti, F. Advances in target detection and tracking in Forward-Looking InfraRed (FLIR) imagery. *Sensors* **2014**, *14*, 20297–20303.
3. Kim, S. High-speed incoming infrared target detection by fusion of spatial and temporal detectors. *Sensors* **2015**, *15*, 7267–7293.
4. Zhao, H.; Ji, Z.; Li, N.; Gu, J.; Li, Y. Target Detection over the Diurnal Cycle Using a Multispectral Infrared Sensor. *Sensors* **2017**, *17*, 56.
5. Liu, D.; Li, Z.; Wang, X.; Zhang, J. Moving target detection by nonlinear adaptive filtering on temporal profiles in infrared image sequences. *Infrared Phys. Technol.* **2015**, *73*, 41–48.
6. Bae, T.W. Small target detection using bilateral filter and temporal cross product in infrared images. *Infrared Phys. Technol.* **2011**, *54*, 403–411.
7. Bae, T.W. Spatial and temporal bilateral filter for infrared small target enhancement. *Infrared Phys. Technol.* **2014**, *63*, 42–53.
8. Gao, C.; Meng, D.; Yang, Y.; Wang, Y.; Zhou, X.; Hauptmann, A.G. Infrared Patch-Image Model for Small Target Detection in a Single Image. *IEEE Trans. Image Process.* **2013**, *22*, 4996–5009.
9. Chen, C.P.; Li, H.; Wei, Y.; Xia, T.; Tang, Y.Y. A local contrast method for small infrared target detection. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 574–581.
10. Deng, H.; Sun, X.; Liu, M.; Ye, C.; Zhou, X. Small Infrared Target Detection Based on Weighted Local Difference Measure. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4204–4214.
11. He, Y.J.; Li, M.; Zhang, J.L.; Yao, J.P. Infrared target tracking via weighted correlation filter. *Infrared Phys. Technol.* **2015**, *73*, 103–114.
12. He, Y.; Li, M.; Zhang, J.; Yao, J. Infrared Target Tracking Based on Robust Low-Rank Sparse Learning. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 232–236.
13. Jiang, X.; Lai, J. Sparse and Dense Hybrid Representation via Dictionary Decomposition for Face Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1067–1079.
14. Lu, H.; Li, X.; Zhang, L.; Xiang, R.; Yang, M.H. Dense and Sparse Reconstruction Error Based Saliency Descriptor. *IEEE Trans. Image Process.* **2016**, *25*, 1592–1603.
15. Ma, X.; Xie, X.; Lam, K.M.; Hu, J.; Zhong, Y. Saliency detection based on singular value decomposition. *J. Vis. Commun. Image Represent.* **2015**, *32*, 95–106.
16. Dai, Y.; Wu, Y.; Song, Y. Infrared small target and background separation via column-wise weighted robust principal component analysis. *Infrared Phys. Technol.* **2016**, *77*, 421–430.
17. Tom, V.T.; Peli, T.; Leung, M.; Bondaryk, J.E. Morphology-based algorithm for point target detection in infrared backgrounds. *Proc. SPIE* **1993**, *1954*, 2–11.

18. Deshpande, S.D.; Meng, H.E.; Venkateswarlu, R.; Chan, P. Max-mean and max-median filters for detection of small targets. In Proceedings of the SPIE's International Symposium on Optical Science, Engineering, and Instrumentation, Denver, CO, USA, 4 October 1999; pp. 74–83.

19. Qu, X.; Chen, H.; Peng, G. Novel detection method for infrared small targets using weighted information entropy. *J. Syst. Eng. Electron.* **2012**, *23*, 838–842.

20. Yang, C.; Ma, J.; Zhang, M.; Zheng, S.; Tian, X. Multiscale facet model for infrared small target detection. *Infrared Phys. Technol.* **2014**, *67*, 202–209.

21. Qi, S.; Xu, G.; Mou, Z.; Huang, D.; Zheng, X. A fast-saliency method for real-time infrared small target detection. *Infrared Phys. Technol.* **2016**, *77*, 440–450.

22. Wei, Y.; You, X.; Li, H. Multiscale patch-based contrast measure for small infrared target detection. *Pattern Recognit.* **2016**, *58*, 216–226.

23. Silverman, J.; Mooney, J.M.; Caefer, C.E. Temporal filters for tracking weak slow point targets in evolving cloud clutter. *Infrared Phys. Technol.* **1996**, *37*, 695–710.

24. Wang, C.; Qin, S. Adaptive detection method of infrared small target based on target-background separation via robust principal component analysis. *Infrared Phys. Technol.* **2015**, *69*, 123–135.

25. Oh, T.H.; Kim, H.; Tai, Y.W.; Bazin, J.C.; Kweon, I.S. Partial Sum Minimization of Singular Values in RPCA for Low-Level Vision. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 145–152.

26. Oh, T.H.; Tai, Y.W.; Bazin, J.C.; Kim, H.; Kweon, I.S. Partial Sum Minimization of Singular Values in Robust PCA: Algorithm and Applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 744–758.

27. Wright, J.; Peng, Y.; Ma, Y.; Ganesh, A.; Rao, S. Robust principal component analysis: exact recovery of corrupted low-rank matrices by convex optimization. In Proceedings of the International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 7–10 December 2009; pp. 2080–2088.

28. Manning, C.D.; Raghavan, P.; Schütze, H. *Introduction to Information Retrieval*; Cambridge University Press: New York, NY, USA, 2008.

29. Lu, C.J.; Tsai, D.M. Automatic defect inspection for LCDs using singular value decomposition. *Int. J. Adv. Manuf. Technol.* **2005**, *25*, 53–61.

30. Beigi, H. Probability Theory and Statistics. In *Fundamentals of Speaker Recognition*; Springer: Boston, MA, USA, 2011; pp. 205–263.

31. Lin, Z.; Ganesh, A.; Wright, J.; Wu, L.; Chen, M.; Ma, Y. Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix. In Proceedings of the Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), Aruba, Dutch Antilles, The Netherlands, 13–16 December 2009; Volume 61.

32. Parikh, N.; Boyd, S. Proximal Algorithms. *Found. Trends Optim.* **2014**, *1*, 127–239.

33. Beck, A.; Teboulle, M. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM J. Imaging Sci.* **2009**, *2*, 183–202.

34. Hale, E.T.; Yin, W.; Zhang, Y. Fixed-point continuation for l1-minimization: Methodology and convergence. *SIAM J. Optim.* **2008**, *19*, 1107–1130.

35. Cai, J.F.; Candès, E.J.; Shen, Z. A Singular Value Thresholding Algorithm for Matrix Completion. *SIAM J. Optim.* **2010**, *20*, 1956–1982.

36. Miezianko, R. IEEE OTCBVS WS Series Bench, 2006. Available online: http://vcipl-okstate.org/pbvs/bench/Data/05/download.html (accessed on 20 December 2016).

37. Li, Y.; Li, P.; Shen, Q. Real-time infrared target tracking based on $\ell_1$ minimization and compressive features. *Appl. Opt.* **2014**, *53*, 6518–26.

38. Caefer, C.E.; Silverman, J.; Mooney, J.M. Optimization of point target tracking filters. *IEEE Trans. Aerosp. Electron. Syst.* **2000**, *36*, 15–25.

39. Kim, S.; Lee, J. Scale invariant small target detection by optimizing signal-to-clutter ratio in heterogeneous background for infrared search and track. *Pattern Recognit.* **2012**, *45*, 393–406.

40. Gu, Y.; Wang, C.; Liu, B.X.; Zhang, Y. A Kernel-Based Nonparametric Regression Method for Clutter Removal in Infrared Small-Target Detection Applications. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 469–473.

41. Borij, A.; Cheng, M.-M.; Jiang, H. Salient Object Detection: A Benchmark. *IEEE Trans. Image Process.* **2015**, *24*, 5706.

42. Fortin, R. Detection of dim targets in digital infrared imagery by morphological image processing. *Opt. Eng.* **1996**, *35*, 1886–1893.