

# A novel hybrid method for robust infrared target detection

Xin Wang<sup>1,2\*</sup>, Lingling Xu<sup>1</sup>, Yuzhen Zhang<sup>2</sup> and Chen Ning<sup>3</sup>

<sup>1</sup>College of Computer and Information, Hohai University  
Nanjing, Jiangsu 211100 - China  
[e-mail: wang\_xin@hhu.edu.cn]

<sup>2</sup>Jiangsu Key Laboratory of Spectral Imaging and Intelligent Sense, Nanjing University of Science and Technology  
Nanjing, Jiangsu 210094 - China

<sup>3</sup>School of Physics and Technology, Nanjing Normal University  
Nanjing, Jiangsu 210023 - China

\*Corresponding author: Xin Wang

*Received February 16, 2017; revised April 13, 2017; accepted May 28, 2017;  
published October 31, 2017*

---

## Abstract

Effect and robust detection of targets in infrared images has crucial meaning for many applications, such as infrared guidance, early warning, and video surveillance. However, it is not an easy task due to the special characteristics of the infrared images, in which the background clutters are severe and the targets are weak. The recent literature demonstrates that sparse representation can help handle the detection problem, however, the detection performance should be improved. To this end, in this text, a hybrid method based on local sparse representation and contrast is proposed, which can effectively and robustly detect the infrared targets. First, a residual image is calculated based on local sparse representation for the original image, in which the target can be effectively highlighted. Then, a local contrast based method is adopted to compute the target prediction image, in which the background clutters can be highly suppressed. Subsequently, the residual image and the target prediction image are combined together adaptively so as to accurately and robustly locate the targets. Based on a set of comprehensive experiments, our algorithm has demonstrated better performance than other existing alternatives.

---

**Keywords:** Infrared, target detection, local sparse representation, contrast

---

This work is supported by the Fundamental Research Funds for the Central Universities (Grant No. 2015B19014), National Natural Science Foundation of China (Grant No. 61603124, 61374019), Six Talents Peak Project of Jiangsu Province (Grant No. XYDXX-007), 333 High-Level Talent Training Program of Jiangsu Province, and Open Research Fund of Jiangsu Key Laboratory of Spectral Imaging & Intelligent Sense (Grant No. 3091601410401).

## 1. Introduction

With the development of infrared imaging technology, infrared target detection has been widely applied not only to the military field, but also to the transportation, industry, agriculture, and other civilian areas. Infrared imaging technology has the advantages of good concealment, strong anti-interference ability, and strong ability to adapt to the environment. However, it is still a challenging problem of the accurate and robust infrared target detection under the complex background. Compared with the visible light images, there exist strong background clutters and noise interferences in the infrared imaging system, leading to low contrast and signal to noise ratio (SNR) in the infrared images. Therefore, effective algorithms should be specially explored to handle the difficulties of infrared object detection.

Infrared target detection has been a hot research topic in the field of image processing and computer vision. A number of infrared target detection methods have been emerging, including Gaussian mixture model method, wavelet transform method, information entropy method, bilateral filtering method, and so on. For example, [1] proposed finite asymmetric generalized Gaussian mixture models for infrared object detection. [2] proposed a knowledge model for target recognition in infrared image. [3] presented an infrared target detection algorithm based on wavelet multi-scale and maximum likelihood. [4] introduced to detect small targets in infrared images by using the self-information map. [5] designed a local mutation weighted information entropy based method to detect the infrared targets. [6] proposed a small infrared target detection algorithm based on bilateral filter and temporal cross product.

In recent years, due to the development of the theory of sparse representation, many target detection algorithms based on sparse representation have been also proposed by researchers. For instance, [7] proposed a computing model based on visual sparse overcomplete representation method, which achieves the object detection in unstructured scenes. In this method, firstly, a sparse overcomplete computing model was established based on energy model and score matching method. Then the object detection algorithm based on neuron response and dynamic threshold strategies was designed. [8] introduced an objects detection method based on local feature and sparse representation. By employing supervised random tree method to learn local features of sample images, a dictionary could be formed. The combination of sub-image blocks of test image and well trained dictionary in first stage could predict the location of the object in the test image, in this way it could obtain a sparse representation of the test image as well as the object detection goal. [9] presented an infrared small target detection method based on image sparse representation. The over-complete target dictionary was first constructed with atoms which were produced by two-dimensional Gaussian model. Then the sub-image blocks of the test image were extracted successively and the corresponding coefficients were calculated with the constructed over-complete target dictionary. There was a significant difference between the coefficients of objective and background. Whether the sub-image block contains small target or not could be determined by the index of sparse concentration. [10] proposed a low-rank and sparse representation (LRSR) model. This model could describe the specific structure of noise data by utilizing sparse representation theory on the basis of low-rank matrix representation. In addition, LRSR based infrared small target detection algorithm was presented. First, a two-dimensional Gaussian model was used to produce the atoms that construct over-complete target dictionary. Then, the reset image data matrix was decomposed by the LRSR model to obtain the background, noise

and target components of the image. Finally, the target position could be determined by threshold processing for the target component data. The above methods may be effective for detecting objects in infrared images to some extent. However, since the size of infrared target under complex background is uncertain, the methods, which usually construct dictionaries by training target sample set, may cause the target features in the dictionary to be submerged by the background features, or cause the phenomenon of overmatching in the process of searching for the matched targets. In addition, due to the special imaging mechanism, infrared images usually have low contrast and lots of background noise. The detection results obtained by the traditional sparse representation based methods always contain a lot of false alarms. These will affect the final performance of the target detection.

Thus, to improve the effectiveness and robustness of infrared target detection, this paper proposes an improved hybrid method which is based on local sparse representation and contrast. Given an original infrared image, it is first processed from two aspects. On the one hand, a residual image is calculated based on local sparse representation for the original image. Specifically, the original image is first divided into a number of image blocks. Then for each block, a dissimilar dictionary is learned by using the inner product approach. The residual image block for each image block is subsequently calculated with the dissimilar dictionary. After that, the residual image for the original image can be obtained. In the residual image, the target can be effectively highlighted. On the other hand, for the original image, a local contrast based method is adopted to compute the target prediction image, in which the background clutters can be highly suppressed. Second, after obtaining the residual image as well as the target prediction image, we propose to combine them together adaptively so as to accurately and robustly locate the targets. Finally, the true targets are detected by the binary threshold segmentation. Experiments based on the real infrared images show that our detection method has a good performance.

The remaining part of this paper is organized as follows. Section 2 reviews the sparse representation theory. Section 3 presents the improved hybrid target detection method. Experimental results are presented and analyzed in Section 4. Finally, we conclude this paper in Section 5.

## 2. Sparse Representation

Image sparse representation, which approximates an input image by a linear combination of a few items from an overcomplete dictionary, has been successfully applied to a variety of problems in image processing and analysis, such as image denoising [11], image restoration [12], object recognition [13], saliency detection [14], visual tracking [15, 16], and image classification [17-19].

Let  $Y$  be a set of  $n$ -dimensional  $N$  input signals, i.e.,  $Y = [y_1, y_2, \dots, y_N] \in R^{n \times N}$ . We wish to learn a reconstructive dictionary with  $K$  elements for sparse representation of  $Y$ . This is formally written as the following optimized problem [20]:

$$\langle D, X \rangle = \arg \min_{D, X} \|Y - DX\|_2^2 \quad s.t. \quad \forall i, \|x_i\|_0 \leq T \quad (1)$$

where  $\|Y - DX\|_2^2$  denotes the reconstruction error.  $D = [d_1, d_2, \dots, d_K] \in R^{n \times K}$  is the learned dictionary and the  $K$  is the number of atoms in the dictionary.  $X = [x_1, x_2, \dots, x_N] \in R^{K \times N}$  is the sparse representation of input signals  $Y$ .  $\|\cdot\|_0$  is the  $L_0$  norm that counts the number of

nonzero elements in a vector.  $T$  is a sparsity threshold, and each  $x_i$  contains  $T$  ( $T \ll n$ ) or fewer nonzero elements.

To train the reconstructive dictionary efficiently, the K-SVD algorithm [21] can be adopted. The characteristic of K-SVD is that it can update the dictionary by column simply and efficiently. As an iterative approach, the K-SVD algorithm includes two steps, i.e., sparse coding and dictionary update, at each iteration.

In the sparse coding step,  $D$  is fixed, the sparse representation  $x_i$  of an input signal  $y_i$  is computed as:

$$x_i = x^*(y_i, D) \equiv \arg \min_x \|y_i - Dx\|_2^2 \quad s.t. \quad \|x\|_0 \leq T \quad (2)$$

Note that, although  $L_0$  provides a simple notion of sparsity, the optimization problem of  $L_0$  norm is nonconvex [22]. That is to say, the direct solution of the Eq. 2 is an NP-Hard problem. There is a way to solve this problem, that is to replace  $L_0$  norm with  $L_1$  norm, so that the problem can be transformed into a convex:

$$\langle D, X \rangle = \arg \min_{D, X} \|Y - DX\|_2^2 + \alpha \|X\|_1 \quad (3)$$

where  $\alpha$  is a trade-off parameter [21], which can balance the reconstruction error and sparsity. Therefore, the Eq. 2 can be rewritten as:

$$x_i = x^*(y_i, D) \equiv \arg \min_x \|y_i - Dx\|_2^2 + \alpha \|x\|_1 \quad (4)$$

In the dictionary update step, the columns, which are known as the dictionary atoms of  $D$ , are updated by SVD [21]. At the same time, the relevant coefficients in  $X$  will be updated.

### 3. Proposed Method

For target detection problems, the traditional sparse representation based methods usually adopt the target samples to train a dictionary. So the dictionary can be thought as a matrix, which contains the target features. Different from such idea, in this paper, we propose to divide the input image into a number of image blocks, and then search the image blocks which are dissimilar to the central image block to learn a dissimilar dictionary. Based on such dictionary, the residual image of the original image can be generated, in which the target can be effectively highlighted. Furthermore, the paper also presents to predict the candidate target areas by using the gray contrast. This may form the target prediction image, in which the background clutters can be highly suppressed. At last, the precise target location can be determined by combining the residual image and the target prediction image, and the final detection results are gotten by the binary threshold segmentation. Fig. 1 illustrates the framework of the proposed method.

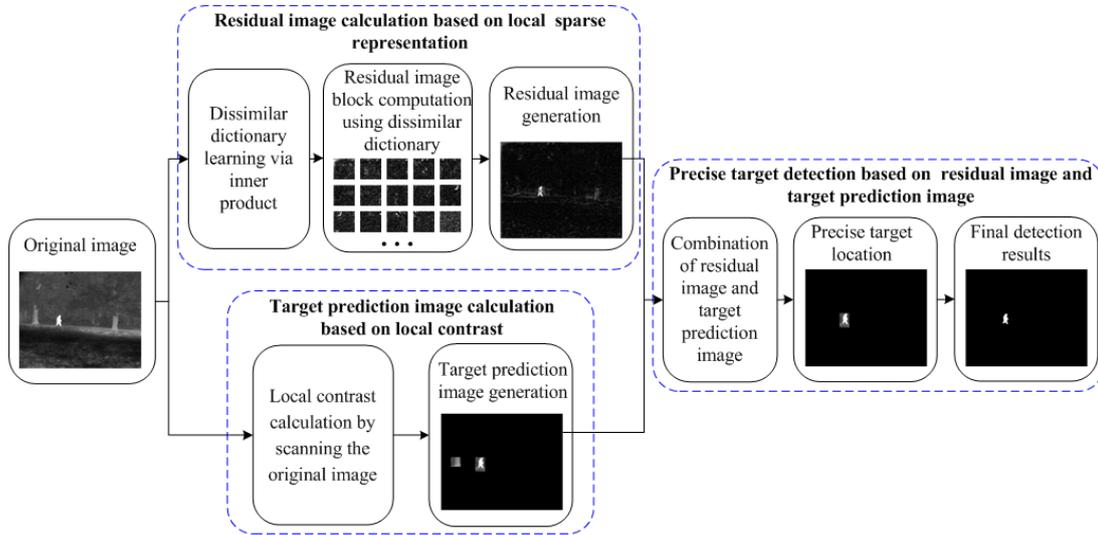


Fig. 1. Framework of the proposed method.

### 3.1 Residual Image Calculation based on Local Sparse Representation

In infrared target detection problems, in contrast with the background area, the target area is usually thought to occupy a small proportion of the whole image. When the input image is divided into a set of images blocks, the blocks, which are around the target area, have great differences from the target area. i.e., the central block. On the contrary, the blocks, which are around the background area, have little differences from its central image block. According to it, in this section, we propose to search the image blocks which are dissimilar to the central image block to learn a dissimilar dictionary. Based on such dictionary, the residual image of the original image can be generated.

#### 3.1.1 Dissimilar Dictionary Learning via Inner Product

As shown in Fig. 2, the process of dissimilar dictionary learning consists of three steps: dividing, searching, and stacking. The specific description is given as below.

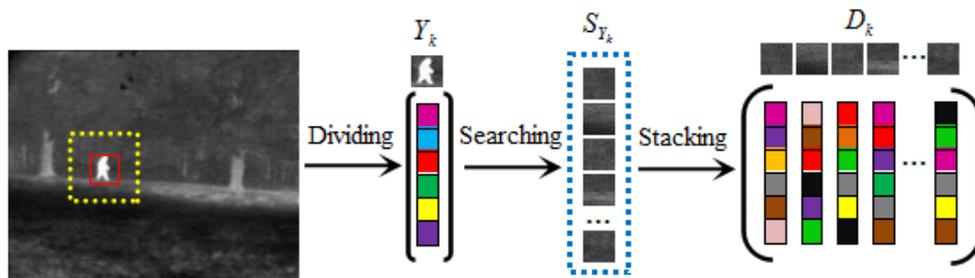


Fig. 2. Dissimilar dictionary learning process.

(1) First, given an input original infrared image  $Y$  with the size of  $n \times N$ , set an empty residual image set  $repatchset = \emptyset$  and a residual image  $G \in R^{n \times N}$ . Each element of  $G$  is initialized to 0. As shown in Fig. 2, a sliding window (i.e., the red window) is used to divide

the original image  $Y$  into  $m$  overlapped image blocks with the size of  $\sqrt{B_s} \times \sqrt{B_s}$ . Each block is column vectorized and then be denoted by a vector  $Y_k \in R^{B_s}$ ,  $k = 1, 2, \dots, m$ .

For each image block, a training window (i.e., the yellow window) with the size of  $L \times L$  is designed. Then it is divided into  $h$  overlapped image blocks with the size of  $\sqrt{B_s} \times \sqrt{B_s}$ . Each block is also column vectorized and then denoted by a vector  $f_g \in R^{B_s}$ ,  $g = 1, 2, \dots, h$ , and ultimately they form a matrix  $F = [f_1, f_2, \dots, f_h] \in R^{B_s \times h}$ .

(2) Second, after obtaining the image block  $Y_k$  and its surrounding blocks  $f_g$ ,  $g = 1, 2, \dots, h$ , calculate the inner product of  $Y_k$  with each  $f_g$ , i.e.:

$$\rho_g = \langle Y_k, f_g \rangle \quad (5)$$

where  $\langle \cdot, \cdot \rangle$  denotes the inner product.  $\rho_g$  represents the inner product of  $Y_k$  with the  $g$ -th column vector in the matrix  $F$  (i.e.,  $f_g$ ). Eventually, we can get the inner product set  $\rho = [\rho_1, \rho_2, \dots, \rho_h]$ . Obviously, the more similar  $f_g$  with  $Y_k$ , the larger the  $\rho_g$  will be.

Moreover, compute the inner product of the  $Y_k$  with itself by:

$$\beta = \langle Y_k, Y_k \rangle \quad (6)$$

To search the dissimilar image block set  $S_{Y_k}$  for  $Y_k$ , compute the difference between  $\beta$  and  $\rho_g$  as follows:

$$E_{r_g} = |\beta - \rho_g| \quad (7)$$

where  $E_{r_g}$  denotes the absolute value of difference between  $\beta$  and the  $g$ -th elements in  $\rho$ . Thus, according to Eq. 7, a set of absolute values of differences between the  $\beta$  and the all elements in the set  $\rho$  can be obtained. Clearly, a higher value of  $E_{r_g}$  indicates a bigger difference between  $f_g$  and  $Y_k$ . Thus, extract  $c$  image blocks with higher absolute values of differences, that is to say, search  $c$  least matched blocks, to form the dissimilar set  $S_{Y_k}$ .

(3) Third, stack the  $c$  least matched blocks in the dissimilar set  $S_{Y_k}$  into a matrix of size  $B_s \times c$ , which is denoted  $D_k \in R^{B_s \times c}$ .  $D_k$  is ultimately regarded as the constructed dissimilar dictionary.

By using such dissimilar dictionary  $D_k$ , the residual image can be obtained in the following sections.

### 3.1.2 Residual Image Block Computation using Dissimilar Dictionary

Based on the dissimilar dictionary  $D_k$ , the residual image block corresponding to the image block  $Y_k$  is calculated as follows.

(1) First, the orthogonal matching pursuit (OMP) algorithm is used to calculate the sparse coding coefficients  $X_k$  of  $Y_k$  with the dictionary  $D_k$ .

(2) Second, compute the reconstruction residual vector  $\varepsilon_k$  by:

$$\varepsilon_k = Y_k - D_k X_k \quad (8)$$

(3) Third, restore the reconstruction residual vector  $\varepsilon_k$  into a matrix of size  $\sqrt{B_s} \times \sqrt{B_s}$ , which is represented by the residual image block *repatch*.

(4) Fourth, deposit *repatch* into the residual image set *repatchset*, i.e.:

$$\text{repatchset} = [\text{repatchset}, \text{repatch}] \quad (9)$$

### 3.1.3 Residual Image Generation

A series of image blocks  $Y_k$  in  $Y$  are processed in accordance with Sections 3.1.1 and 3.1.2. Then the whole residual image set *repatchset* can be obtained.

According to the position of each corresponding original image block in the image  $Y$ , each image block in *repatchset* is put into the residual image  $G$ , where the overlapped parts are calculated by the mean values.

Finally, the residual image  $G$  can be obtained. Compared to the original  $Y$ , the object in  $G$  can be effectively highlighted.

### 3.2 Target Prediction Image Computation based on Local Contrast

Although the target can be highlighted in the residual image obtained above, the residual image still contains the background noise. To suppress the background clutters, a local contrast method is applied to the original infrared image to generate a target prediction image, in which the background clutters can be effectively suppressed.

In this method, the candidate target regions in the original image are detected by a search window as shown in Fig. 3. The search window is divided into nine cells with the same size. The central cell, referred to as the region under test, is denoted by “9”, where the target is likely to appear; other eight cells, i.e., cells “1” to “8” are referred to as the background regions. The specific steps are given as follows.

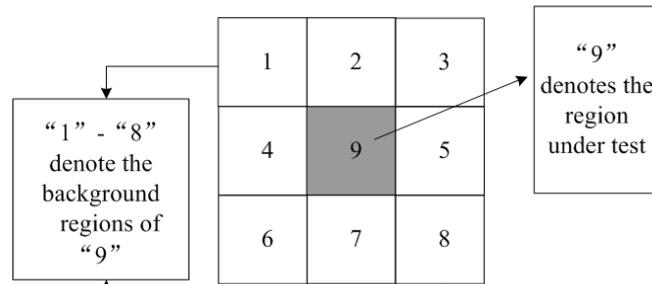


Fig. 3. The search window for target prediction.

(1) First, scan the original infrared image from top to bottom, left to right, by a sliding window with the size of  $\sqrt{U} \times \sqrt{U}$  to get a number of image blocks. Considering that the sizes of infrared targets in many real world applications generally vary from 1 to 100 pixels [23],  $\sqrt{U}$  is here set to 16 to ensure the integrity of the detected target.

In addition, in order to improve the detection efficiency, the sliding step is set to 8, i.e.,  $step = 8$ . The selection of this value of  $step$  will be discussed in Section 4.1.

(2) Second, for each image block obtained above, regarding it as a region under test, expand it to obtain a search window as shown in Fig. 3. Then, the region under test is judged by:

$$\text{the region under test is} = \begin{cases} \text{the candidate target region} & \mu_t > \mu_b + \lambda \times \sigma_b \\ \text{the background region} & \mu_t \leq \mu_b + \lambda \times \sigma_b \end{cases} \quad (10)$$

where  $\mu_t$  denotes the mean value of the region under test.  $\mu_b$  and  $\sigma_b$  are the mean value and the standard deviation value of the background regions.  $\lambda$  is a parameter. Obviously, the greater the value of  $\lambda$  is, the smaller number of candidates is.  $\lambda$  often belongs to [0.5, 3] in order to detect the targets with different sizes of the original image [24, 25]. In this paper,  $\lambda = 0.9$ . Its specific analysis will be also given in section 4.1.

(3) If a region under test is judged as a candidate target region, its pixel values are preserved. Otherwise, its pixel values are set to zeros. Thus, an image that only contains the possible candidate objects is generated, which is called the target prediction image. Compared to the original infrared image, the background clutters in the target prediction image are well suppressed.

### 3.3 Precise Target Detection based on Residual Image and Target Prediction Image

After obtaining the residual image and the target prediction image, to effectively and robustly detect the infrared targets, we combine these two kinds of images together automatically. The specific process is as follows.

(1) First, based on the computed residual image and target prediction image, the pixel values of nonzero regions in the target prediction image are converted to the pixel values of the corresponding regions in the residual image, so that a novel target prediction residual image  $S$  is generated.

(2) Second, since the method proposed in this paper is designed for single target detection problem, the real target can be detected by searching the region which has the highest mean value in the target prediction residual image  $S$ . Therefore, it is necessary to compare the mean values of all unconnected nonzero regions in the image  $S$ . The mean value is computed as follows:

$$\text{mean} = \frac{1}{n_p} \sum_{i=1}^{n_p} p_i \quad (11)$$

where  $p_i$  denotes the  $i$ -th pixel value of the nonzero region.  $n_p$  denotes the number of pixels in this region. Consequently, the real target region is the nonzero region which has the maximum mean value. The real target region is finally mapped on the original image, and the nonzero target regions in the original image are set to zeros.

(3) Third, apply a binary threshold segmentation method to the detected target region, and the final target detection results can be obtained.

## 4. Experimental Results and Analysis

We evaluate our proposed infrared target detection method on a dataset of 1000 real-life infrared images collected from the Internet [26] and compare its performance with several other classical algorithms. All the experiments are implemented on an Intel Dual Core 2.3 GHz laptop with 4 GB RAM. The programming platform is MATLAB R2013b.

By following most previous works [27], two standard criterions, which are called detection rate (DR) and false alarm rate (FAR), are used to measure the detection performance:

$$DR = \frac{DT}{DT + MT} \quad (12)$$

$$FAR = \frac{FT}{FT + DT} \quad (13)$$

where  $DT$  is the number of accurately detected targets.  $MT$  represents the number of miss targets.  $FT$  denotes the number of false alarms, that is, the object is not a real target but is detected incorrectly. Obviously, in the ideal stat,  $DR = 1$ ,  $FAR = 0$ . Besides,  $DR$  measures the effectiveness of the detection method, and  $FAR$  reflects the robustness of the detection method.

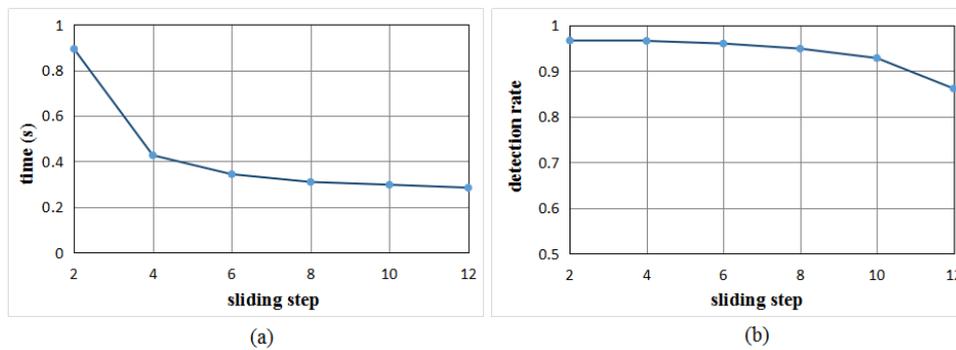
#### 4.1 Parameter Analysis

Before evaluating our method, we briefly discuss the question of choosing the parameters in our formulation. Performing cross validation on all parameters would of course be cumbersome, so we use a few heuristics to either fix arbitrarily some of the parameters or reduce the search space for them.

First, in the local sparse representation based residual image calculation stage, considering that the size of the original images is  $160 \times 200$ , we first set the size of each image block, i.e.,  $\sqrt{B_s} \times \sqrt{B_s}$  to  $40 \times 40$ . The width of overlapping between adjacent image blocks is set to half of the image block size, i.e., 20. Second, the size of the training window, i.e.,  $L \times L$  is set to be  $98 \times 98$ , which takes into account the size of the original image. In the training widow, the width of overlapping between adjacent image blocks is set to 1. Then, we can get  $h = 3481$  overlapped image blocks. Since each block is column vectorized to be a vector  $Y_k \in R^{B_s}$  ( $B_s = 1600$ ), the number of dissimilar blocks for each original image block, i.e.,  $c$  is set to 1690, which should be larger than  $B_s$ .

Second, in the stage of local contrast based target prediction image computation, there are two crucial parameters that may influence the detection performance: the first is the sliding step  $step$  and the second is the parameter  $\lambda$ . To select the optimal values of these two parameters, we conducted a series of experiments.

(1) The parameter  $step$



**Fig. 4.** Analysis of the parameter  $step$ . (a) Effect of  $step$  on detection efficiency. (b) Effect of  $step$  on detection rate.

The parameter *step* should not be too large. If it is too large, the real targets may be missed, which will result in low detection rates. Also, it should not be too small; otherwise, the computational complexity of the algorithm will increase. To get the optimal value of the parameter *step*, we try a few different values, for instance,  $step = 2 + 2 \times k$ , with  $k \in \{0, \dots, 5\}$ . For each value, we carry out experiments on all of the test images and calculate the efficiencies and the detection rates. The results are shown in Fig. 4. To balance the efficiency and detection rate, we select  $step = 8$  in our experiments.

(2) The parameter  $\lambda$

Next, we determine the value of  $\lambda$ . Starting with  $\lambda = 0$ , we gradually increase its value by 0.3 until  $\lambda = 1.8$ . For each value, we do the experiments on all of the test images to compute the detection rate and the number of candidate target regions. The results are given in Fig. 5. From Fig. 5 (a), we can find that when the parameter  $\lambda$  increases from 0 to 0.9, the detection rate has no obvious change, but the number of candidate target regions decreases distinctly. After that, with the increase of parameter  $\lambda$ , the detection rate begins to decline significantly and the number of the candidate target regions continues to reduce but the decline gradually decreases. Thus, to make a balance between the detection rate and the number of candidate target regions, we select  $\lambda = 0.9$ .

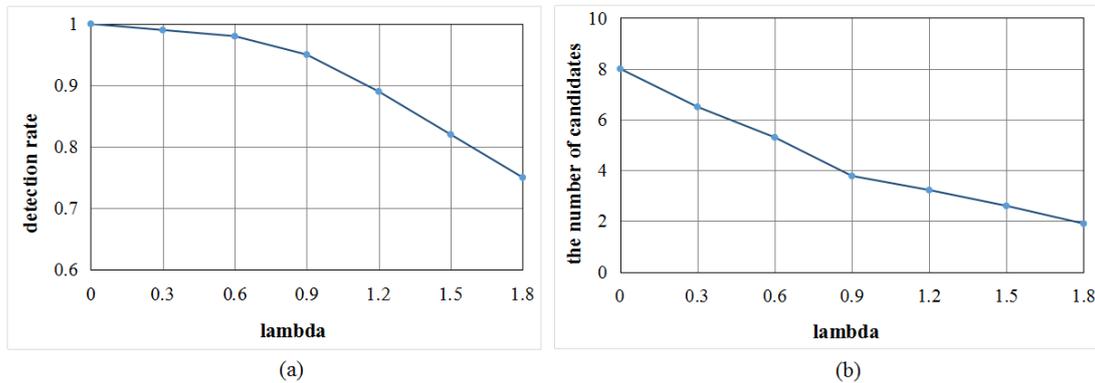


Fig. 5. Analysis of the parameter  $\lambda$ . (a) Effect of  $\lambda$  on detection rate. (b) Effect of  $\lambda$  on the number of candidate target regions.

## 4.2 Target Detection Results of Proposed Method

Now, we start to verify the performance of the proposed algorithm by using some examples. The original infrared images are shown in Fig. 6. These six original infrared images involve various categories of targets with different backgrounds. As shown in Fig. 6(a), a person is in the woods. Fig. 6(b) shows a person is on the street. Fig. 6(c) represents a person with woods clutters. Fig. 6(d) shows a plane under the sky. In Fig. 6(e), an aircraft exists in the sky with clutters. Fig. 6(f) represents a vessel in the sea. Besides, the size information, which is about the original images and targets, are given in the Table 1.

The detection results of the six infrared images are shown in Fig. 7. The first column of Fig. 7 shows the residual image of each original image obtained based on local sparse representation. The second column displays the target prediction images gotten by local contrast. The third column shows the images that only contain the target regions without other interference regions. The fourth column indicates the final target detection results. Fig. 7(a)

shows the proposed algorithm can accurately detect the human target in the forest. In the Fig. 7(b) and Fig. 7(d), the targets are small, but they are still detected correctly. Fig. 7(c) shows that since the background is complex, there exist many candidate target areas. Nevertheless, the values of the disturbance areas in the residual image are effectively reduced, so the interference areas can be well eliminated. Fig. 7(e) illustrates that the plane target with a large size can be detected accurately as well. As shown in Fig. 7(f), even though the values of the target in the residual image is not very high, the local contrast based approach can well solve this problem, and finally the target is successfully detected. In all, the experimental results show that our algorithm can obtain very good detection results.

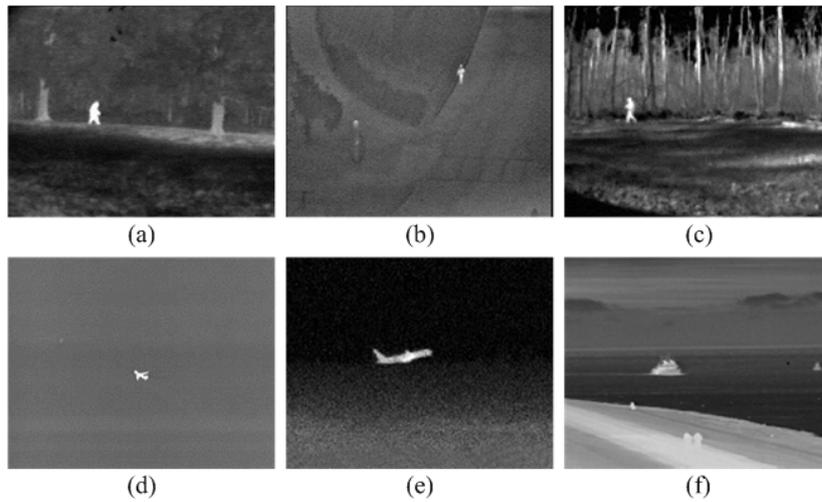


Fig. 6. Examples of the original infrared images.

Table 1. Information of the six images.

Images	Fig. 6(a)	Fig. 6(b)	Fig. 6(c)	Fig. 6(d)	Fig. 6(e)	Fig. 6(f)
Image size	160×200	160×200	160×200	160×200	160×200	160×200
Target size	15×10	15×6	19×9	8×13	12×44	14×28

### 4.3 Qualitative Comparison with Other Approaches

To demonstrate the effectiveness and the robustness of our method, it is compared with four existing infrared target detection algorithms.

The first one is a background prediction based detection method [28] (referred to as BP). It uses the local area images and the estimated images to predict the background adaptively, and then adopts the threshold selection algorithm based on two-dimensional exponential entropy to further identify the target. The second approach is based on top-hat transformations [29], and it is referred to as Top-hat in this paper. The third algorithm is based on mathematical morphology [30]. It firstly uses median filter to filter the image, then uses mathematical morphology to suppress the background noise, and finally the target is detected by using adaptive threshold. We call this method MFMM. The fourth one is based on local contrast [31]. First, the local contrast map of the input image is obtained using the local contrast measure. Then, an adaptive threshold is adopted to segment the target. We refer to this method as LC.

For a fair comparison, all methods are applied to the 1000 infrared images in the dataset.

Ten images and their corresponding detection results obtained by different methods are presented in Fig. 8. The basic information of the original images in Fig. 8(e)-(j) has been given in Table 1. Table 2 shows the basic information of the original images in Fig. 8(a)-(d). From Fig. 8, we can see that our method achieves the best performance, while the results of the Top-hat algorithm and LC method contain too much noise and clutter. In addition, the targets detected by BP and the MFMM algorithms are not complete.

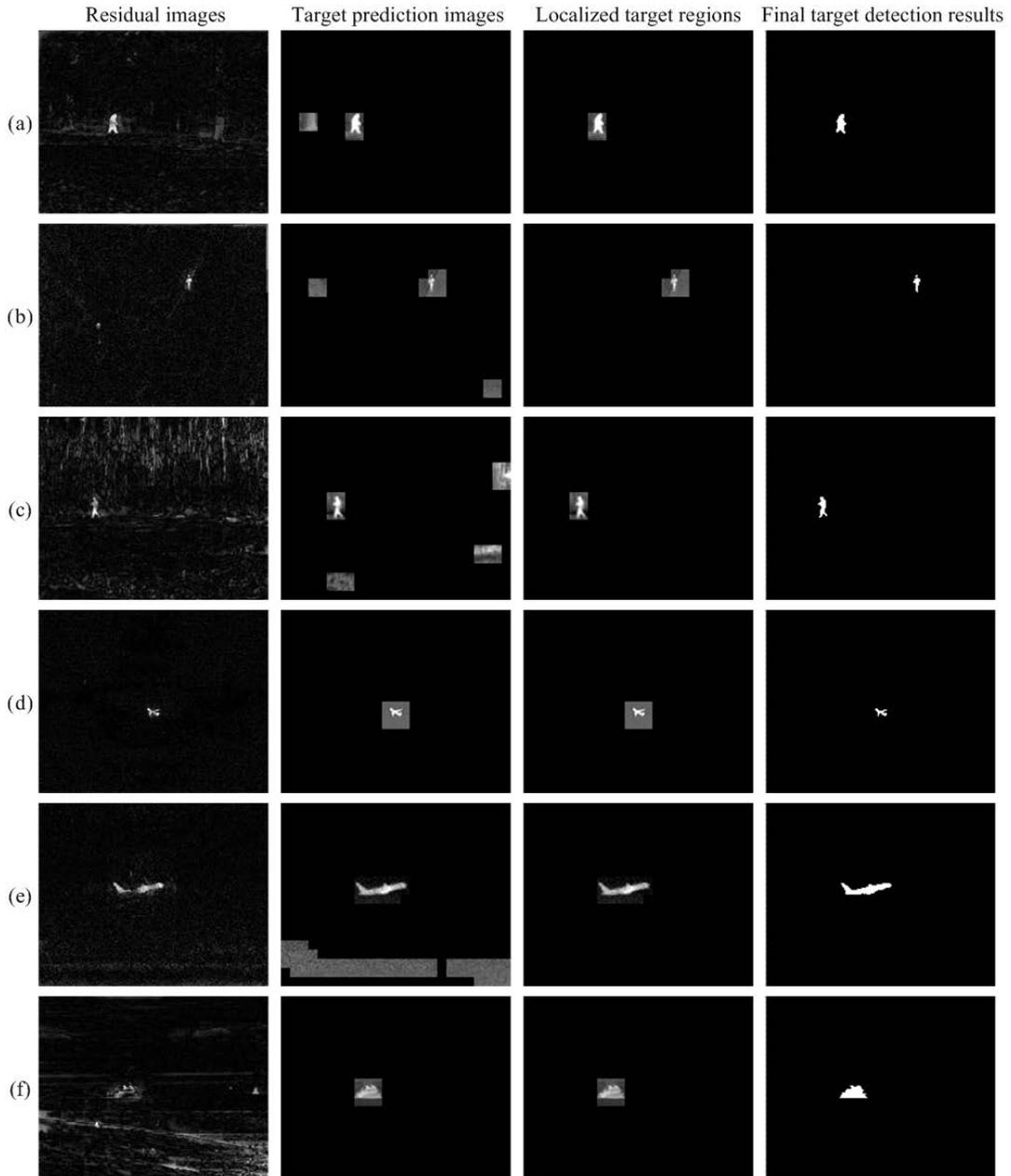
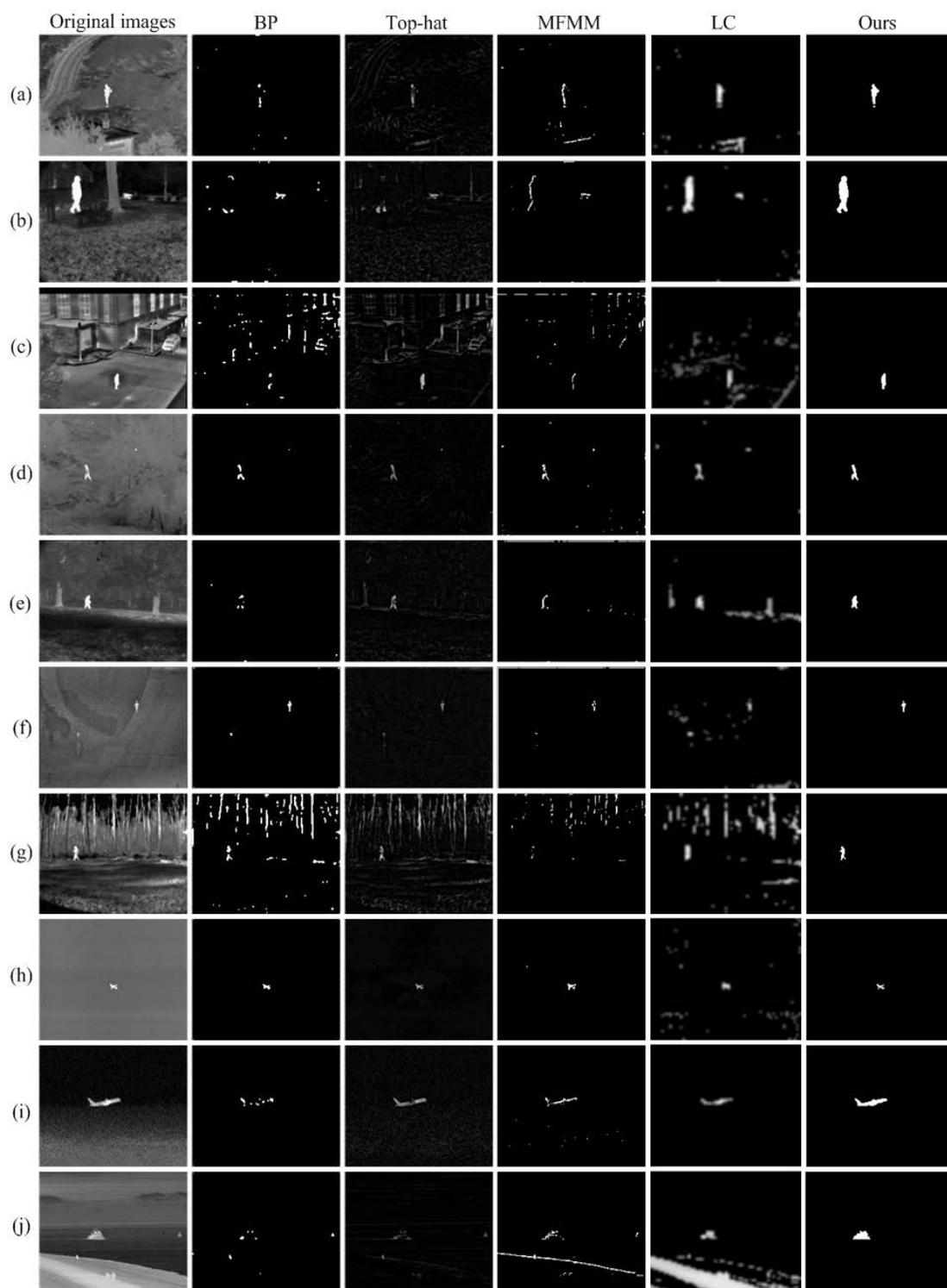


Fig. 7. Target detection results of the proposed method.



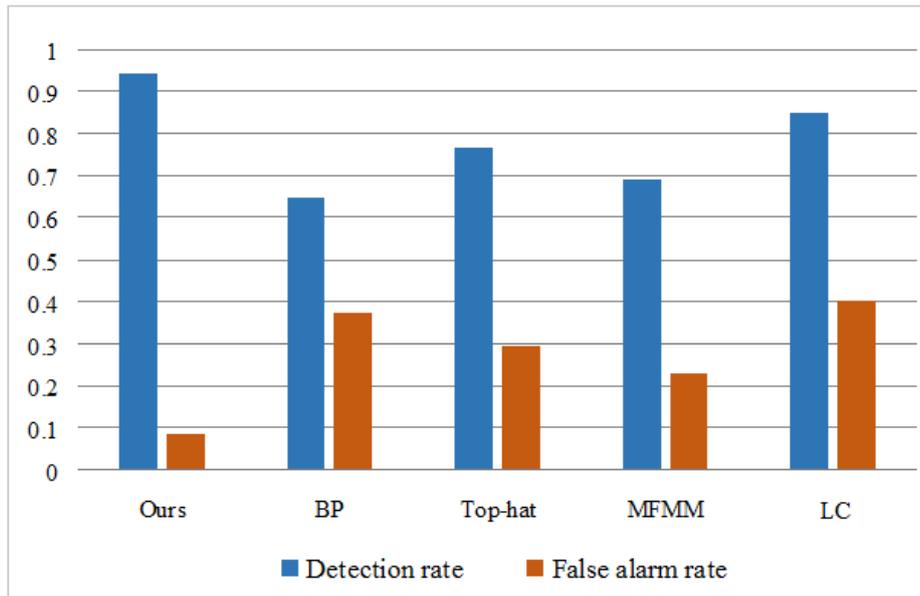
**Fig. 8.** Qualitative comparison results between the proposed method and other state-of-the-art approaches.

**Table 2.** Information of other four images.

Images	Fig. 8(a)	Fig. 8(b)	Fig. 8(c)	Fig. 8(d)
Image size	160×200	160×200	160×200	160×200
Target size	31×13	51×20	24×9	19×11

#### 4.4 Quantitative Comparison with Other Approaches

Also, to further illustrate the superiority of our algorithm, the two criteria DR and FAR are used for quantitative comparison. The comparison results are shown in Fig. 9. For our method, in the residual image generation stage, some dissimilar image blocks of each original image block make up the overcomplete dictionary, so the background noise of the residual image is weakened, which lays the foundation for the accuracy of the target detection. Moreover, the combination of the target prediction image and residual image can make the target to be detected correctly, and it improves the robustness of the algorithm and reduces the false alarm rate. As shown in Fig. 9, compared with the other four algorithms, our method produces the highest detection rate and the lowest false alarm rate for infrared target detection.



**Fig. 9.** Quantitative comparison results between the proposed method and other state-of-the-art approaches.

## 5. Conclusion

In this paper, we designed a hybrid method for target detection in infrared images. First, a residual image is calculated based on local sparse representation for the original image, in which the target can be effectively highlighted. Then, a local contrast based method is adopted to compute the target prediction image, in which the background clutters can be highly suppressed. Subsequently, the residual image and the target prediction image are combined together adaptively so as to accurately and robustly locate the targets. Experimental results

compared to several classical methods demonstrate the effectiveness and robustness of the proposed method.

## References

- [1] Tarek Elguebaly and Nizar Bouguila, "Finite asymmetric generalized Gaussian mixture models learning for infrared object detection," *Computer Vision & Image Understanding*, vol. 117, no. 12, pp. 1659-1671, December, 2013. [Article \(CrossRef Link\)](#).
- [2] Juan Su, Luo Yang and Jun Lu, "An Infrared Target Detection Algorithm Based on Knowledge Model," *Acta Photonica Sinica*, vol. 42, no. 7, pp. 839-844, July 2013. [Article \(CrossRef Link\)](#).
- [3] Zhaohui Li, Bing Wang and Ming Chen, "Infrared Target Detection Based on Wavelet Multi Scale and Maximum Likelihood," *Acta Optica Sinica*, vol. 29, no. 12, pp. 3404-3408, December 2009. [Article \(CrossRef Link\)](#).
- [4] He Deng and Jianguo Liu, "Infrared small target detection based on the self-information map," *Infrared Physics & Technology*, vol. 54, no. 2, pp. 100-107, March 2011. [Article \(CrossRef Link\)](#).
- [5] Xiujie Qu, He Chen and Guihua Peng, "Novel detection method for infrared small targets using weighted information entropy," *Journal of Systems engineering and electronics*, vol. 23, no.6, pp. 838-842, December 2012. [Article \(CrossRef Link\)](#).
- [6] Tae-Wuk Bae, "Small target detection using bilateral filter and temporal cross product in infrared images," *Infrared Physics & Technology*, vol. 54, no. 5, pp. 403-411, September 2011. [Article \(CrossRef Link\)](#).
- [7] Peng Lu, Shanshan Zhang, Chi Liu, Shilei Huang and Yuhe Tang, "Object detection algorithm based on sparse overcomplete representation," *Chinese Journal of Scientific Instrument*, vol. 34, no. 6, pp. 1273-1278, June, 2013. [Article \(CrossRef Link\)](#).
- [8] Yuanrong Tian, Song Tian and Yuelel Xu, "Image object detection based on local feature and sparse representation," *Journal of Computer Applications*, vol. 33, no. 6, pp. 1670-1673, June 2013. [Article \(CrossRef Link\)](#).
- [9] Jiajia Zhao, Zhengyuan Tang and Jie Yang, "Infrared small target detection based on image sparse representation," *Journal of Infrared & Millimeter Waves*, vol. 30, no. 2, pp. 156-161, February 2012. [Article \(CrossRef Link\)](#).
- [10] Yujie He, Min Li and Jinli Zhang, "Small infrared target detection based on low-rank and sparse representation," *Infrared Physics & Technology*, vol. 68, pp. 98-109, January 2015. [Article \(CrossRef Link\)](#).
- [11] Michael. Elad and Michal Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736-3745, December 2006. [Article \(CrossRef Link\)](#).
- [12] Jian Zhang, Debin Zhao and Wen Gao, "Group-Based Sparse Representation for Image Restoration," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3336-3351, March, 2014. [Article \(CrossRef Link\)](#).
- [13] Xin Wang, Siqiu Shen, Chen Ning, Fengchen Huang and Hongmin Gao, "Multi-class remote sensing object recognition based on discriminative sparse representation," *Applied Optics*, vol. 55, no. 6, pp. 1381-1394, June, 2016. [Article \(CrossRef Link\)](#).
- [14] Zhenfeng Zhu, Qian Chen, Yao Zhao, "Ensemble dictionary learning for saliency detection," *Image & Vision Computing*, vol. 32, no. 3, pp. 180-188, March, 2014. [Article \(CrossRef Link\)](#).
- [15] Tianzhu Zhang, Si Liu, Changsheng Xu, Shuicheng Yan, Bernard Ghanem, Narendra Ahuja, and Ming-Hsuan Yang, "Structural Sparse Tracking," in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*, pp.150-158, June 7-12, 2015. [Article \(CrossRef Link\)](#).
- [16] Tianzhu Zhang, Si Liu, Narendra Ahuja, Ming-Hsuan Yang, and Bernard Ghanem, "Robust Visual Tracking Via Consistent Low-Rank Sparse Learning," *International Journal of Computer Vision*, vol. 111, no. 2, pp. 171-190, January, 2015. [Article \(CrossRef Link\)](#).

- [17] Tianzhu Zhang, Bernard Ghanem, Si Liu, "Low-Rank Sparse Coding for Image Classification," in *Proc. of IEEE International Conference on Computer Vision*, pp. 281-288, December 1-8, 2013. [Article \(CrossRef Link\)](#).
- [18] Xiaoshan Yang, Tianzhu Zhang and Changsheng Xu, "Locality discriminative coding for image classification," in *Proc. of International Conference on Internet Multimedia Computing and Service*, pp.52-55, August 17-19, 2013. [Article \(CrossRef Link\)](#).
- [19] Xiaoshan Yang, Tianzhu Zhang, Changsheng Xu, and Min Xu, "Graph-Guided Fusion Penalty Based Sparse Coding for Image Classification," in *Proc. of Pacific-Rim Conference on Advances in Multimedia Information Processing*, pp. 475-484, December 13-16, 2013. [Article \(CrossRef Link\)](#).
- [20] Zhuolin Jiang, Zhe Lin and Larry S. Davis, "Label consistent K-SVD: learning a discriminative dictionary for recognition," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 35, no. 35, pp. 2651-2664, November 2013.. [Article \(CrossRef Link\)](#).
- [21] Michal Aharon, Michael Elad and Alfred Bruckstein, "K-svd: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311-4322, October, 2006. [Article \(CrossRef Link\)](#).
- [22] Tanaya Guha and Rabab K. Ward, "Learning sparse representations for human action recognition," *IEEE Transactions on Software Engineering*, vol. 34, no. 8, pp. 1576-1588, December, 2011. [Article \(CrossRef Link\)](#).
- [23] Xin. Wang, Guofang Lv and Lizhong Xu, "Infrared dim target detection based on visual attention," *Infrared Physics & Technology*, vol. 55, no. 6, pp. 513-521, June 2012. [Article \(CrossRef Link\)](#).
- [24] Wenguang Wang, Chenming Li and Jianing Shi, "A robust infrared dim target detection method based on template filtering and saliency extraction," *Infrared Physics & Technology*, vol. 73, pp. 19-28, November 2015. [Article \(CrossRef Link\)](#).
- [25] Xin Wang, Chen Ning and Lizhong Xu, "Spatiotemporal saliency model for small moving object detection in infrared videos," *Infrared Physics & Technology*, vol. 69, pp. 111-117, January 2015. [Article \(CrossRef Link\)](#).
- [26] J. W. Davis, OTCBVS Benchmark Dataset Collection, <http://vcipl-okstate.org/pbvs/bench/>
- [27] Yansheng Li, Yihua Tan, Hang Li and Jinwen Tian, "Biologically inspired multilevel approach for multiple moving targets detection from airborne forward-looking infrared sequences," *Journal of the Optical Society of America A*, vol. 31, no. 4, pp. 734-44, April, 2014. [Article \(CrossRef Link\)](#).
- [28] Chen Zheng, Hong Zheng, Qiangqiang Sui and Xiangzhen Li, "An efficient background prediction algorithm for infrared dim target detection," in *Proc. of IEEE International Conference on Natural Computation*, pp. 89-91, August 10-12, 2010. [Article \(CrossRef Link\)](#).
- [29] Xiangzhi Bai, Fugen Zhou, "Infrared small target enhancement and detection based on modified top-hat transformations," *Computers & Electrical Engineering*, vol. 36, no. 6, pp. 1193-1201, November, 2010. [Article \(CrossRef Link\)](#).
- [30] Feng Jiang, Minhui, Jin and Liquan Song, "Small infrared target detection algorithm based on mathematical morphology," in *Proc. of SPIE*, pp. 81934L-81934L-7, September 8, 2011. [Article \(CrossRef Link\)](#).
- [31] C. L. Philip Chen, Hong Li, Yantao Wei, Tian Xia and Yuan Yan Tang, "A local contrast method for small infrared target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp. 574-581, January 2014. [Article \(CrossRef Link\)](#).
- [32] Xiaoshan Yang, Tianzhu Zhang, and Changsheng Xu, "Cross-Domain Feature Learning in Multimedia," *IEEE Transactions on Multimedia*, vol.17, no.1, pp. 64-78, December 2014. [Article \(CrossRef Link\)](#).
- [33] Jianbing Shen, Xiaoshan Yang, Yunde Jia, and Xuelong Li, "Intrinsic images using optimization," in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 3481-3487, August, 2011. [Article \(CrossRef Link\)](#).



**Xin Wang** received the Ph.D. degree in Computer Application Technology from Nanjing University of Science and Technology, Nanjing, China, in 2010. She is currently an Associate Professor with the College of Computer and Information, Hohai University, Nanjing, China. She has published more than 50 papers in journals and referred conferences. Her current research interests include image processing, computer vision, and pattern recognition.



**Lingling Xu** received the B.S. degree in Communication Engineering from Nanjing University of Science and Technology Zijin College, Nanjing, China, in 2015. Now she is working toward the M.S. degree in the College of Computer and Information, Hohai University. Her current research interests include image processing, target detection and tracking.



**Yuzhen Zhang** received the Ph.D. degree in Optical Engineering from Nanjing University of Science and Technology, Nanjing, China. She is currently an Associate Professor in the university. She has published more than 30 papers in journals and referred conferences. Her current research interests include signal processing, image processing and analysis.



**Chen Ning** received the M.S. degree in Signal and Information Processing from University of Science and Technology of China, Hefei, China, in 2003. He is currently working in the School of Physics and Technology, Nanjing Normal University, Nanjing, China. His current research interests include signal and image processing, and computer vision.