Infrared moving small target detection based on saliency extraction and image sparse representation

Conference Paper · October 2016

DOI: 10.1117/12.2247309

Minjie Wan
Nanjing University of Science and Technology

Infrared moving small target detection based on saliency extraction and image sparse representation

Xiaomin Zhang, Kan Ren, Jin Gao, Chaowei Li, Guohua Gu, et al.

Xiaomin Zhang, Kan Ren, Jin Gao, Chaowei Li, Guohua Gu, Minjie Wan, "Infrared moving small target detection based on saliency extraction and image sparse representation," Proc. SPIE 10157, Infrared Technology and Applications, and Robot Sensing and Advanced Control, 1015734 (1 November 2016); doi: 10.1117/12.2247309

Event: International Symposium on Optoelectronic Technology and Application 2016, 2016, Beijing, China

Infrared moving small target detection based on saliency extraction and image sparse representation

Xiaomin Zhang [1,2], Kan Ren [1, 2*], Jin Gao[3], Chaowei Li[3], Guohua Gu[1,2], Minjie Wan[1,2]

[1]School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Jiangsu Nanjing 210094, China

[2]Collaborative Innovation Center of Social Safety Science and Technology,

Jiangsu Nanjing 210094, China

[3]The Research Institute 202 of China North Industries Group Corporation,

Shanxi Xianyang 712099, China

[*]Corresponding Author: k.ren@njust.edu.cn

Abstract

Moving small target detection in infrared image is a crucial technique of infrared search and tracking system. This paper present a novel small target detection technique based on frequency-domain saliency extraction and image sparse representation. Firstly, we exploit the features of Fourier spectrum image and magnitude spectrum of Fourier Transform to make a rough extract of saliency regions and use a threshold segmentation system to classify the regions which look salient from the background, it gives us a binary image as result. Secondly, a new patch-image model and over-complete dictionary were introduced to the detection system, then the infrared small target detection was converted into a problem-solving and optimization process of patch-image information reconstruction based on sparse representation. More specifically, the test image and binary image can be decomposed into some image patches follow certain rules. We select the target potential area according to the binary patch-image which contains salient region information, then exploit the over-complete infrared small target dictionary to reconstruct the test image blocks which may contain targets. The coefficients of target image patch satisfy sparse features. Finally, for image sequence, Euclidean distance was used to reduce false alarm ratio and increase the detection accuracy of moving small targets in infrared images due to the target position correlation between frames.

Keywords: infrared small target detection, saliency extraction, patch-image, sparse representation

1 Introduction

Infrared moving small target detection is a hot research task of infrared detection technology. It plays an important role in infrared early warning and precise guidance. Generally, dim target in infrared image is fairly small and lack of texture information, with weak brightness and fewer pixels. When there exists heavy noise and clutter, small targets are usually immersed in a complex background with low signal-to-clutter ratio (SCR), which results in the considerable difficulties and challenges of detection. For some time, many kinds of algorithms have been applied to infrared small target detection under various complex backgrounds. They can be roughly classified into the following several categories:

Infrared Technology and Applications, and Robot Sensing and Advanced Control,
edited by Haimei Gong and Aiguo Song, Proc. of SPIE Vol. 10157, 1015734
 · doi: 10.1117/12.2247309

methods based on filtering, feature extraction, mathematical morphology theory and so on. Filtering is the most widely used detection method. For example, Peng et al. proposed a real-time target detection method using a high-pass filtering template [1]; L. Yang et al. proposed the adaptive Butterworth high pass filtering (BHPF) method [2]; Cao et al. introduced two-dimensional least mean square (TDLMS) filter into small target detection [3]. These methods have been widely used due to the smaller amount of calculation and good performance in real-time detection processing. They usually have specific filtering templates, and the key step of them is to effectively remove the background of the original image. However, the detection performance of filtering algorithm will deteriorate rapidly when the small targets were immersed in backgrounds, there will be a large number of false alarm under the disturbance of complex background. In the feature extraction algorithm, different characteristics were extracted from the target and background to distinguish the target from the background. Such as: local contrast method [4] and principal curvature method [5]. Mathematical morphology theory is another important bunch in small target detection domain. For instance, all kinds of top-hat transformations [6,7] can enhance regions of interest in infrared images efficiently, but are sensitive to dim target and clutter background.

2 The proposed method

Many existing methods do not play well when there exists heavy noise and clutter, In order to overcome the drawbacks of the existing methods, this method was proposed and it outperforms several existing methods in adaptability and robustness. The process of this method includes four sections, whose flow chart is presented in Fig.1.

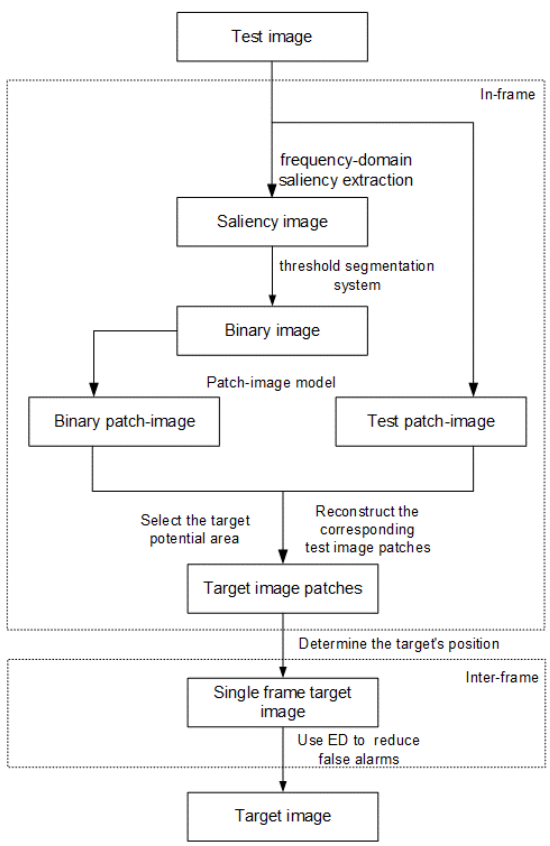


Fig.1. Flow chart of our algorithm

2.1 Saliency extraction

Human can distinguish small targets in infrared image because our human eye is sensitive to high-frequency image components, which can be regarded as saliency region in pictures. It is reasonable to make magnitude transformations of test image so that the high frequency proportion of real target can be improved greatly .We exploit features of Fourier spectrum image and magnitude spectrum of Fourier Transform to make a rough extract of saliency regions and get a saliency map. The procedures of computing saliency map are presented as follows:

Firstly, do Fourier transform on the test image $f(x,y)$ and obtain the magnitude spectrum $A(u,v)$ as Eq.2, $F(f)(u,v)$ in Eq.1 means Fourier spectrum image.

$$F(f)(u,v)=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)e^{-j2\pi(ux/M+vy/N)} \tag{1}$$

$$A(u,v) = |F(u,v)| \tag{2}$$

Secondly, calculate the saliency map as Eq.3. For image signal, spatial frequency means the number of times the intensity changes within the unit length. The physical meaning of Fourier transform is transform image gray distribution function to frequency distribution function. Fourier spectrum image can represent the gradient distribution of space image, it can be understood as low frequency components in spectrum image correspond to the part with small gray level gradient in space image, on the contrary, high frequency components in spectrum image correspond to regions with grayscale change violently, which are usually cause by the small target, noise point and image edge. In addition, the image amplitude spectrum attenuate quickly as the frequency increases, so Eq.3 can be considered as frequency domain high-pass filtering operation, where F^{-1} denotes the inverse Fourier Transform, $g_n(x,y)$ is a 2D Gaussian filter.

$$S(x,y)=g_n(x,y)*\left|F^{-1}\left[F(u,v)./A(u,v)\right]\right|^2 \tag{3}$$

However, it is easily to find that in addition to the candidate target areas, the background areas have a small value, so we need some further processing operation. For the saliency region location, a convolution template is introduced to do binary segmentation according to the gradient changes of pixel values in the saliency map. More specifically, the convolution template is divided into nine squares with the same size, which is presented in Fig.2, the central one is called square under test where the target may appear, and other ones are called background sub squares. We calculate the mean

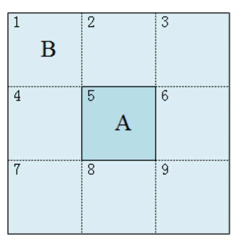


Fig.2. convolution template

value and standard deviation of pixel values in central area and background area separately, then use a threshold segmentation system to classify the interest region which looks salient from the background. We obtained the whole binary image by the convolution of the template and saliency map. The formula of the threshold segmentation system is defined as Eq.4. k is a coefficient, generally belongs to 3-7. A is the target potential area, correspond to square 5 which is under test, B represent the background squares 1-9 except 5. And the saliency extraction pictures were show in Fig.3.

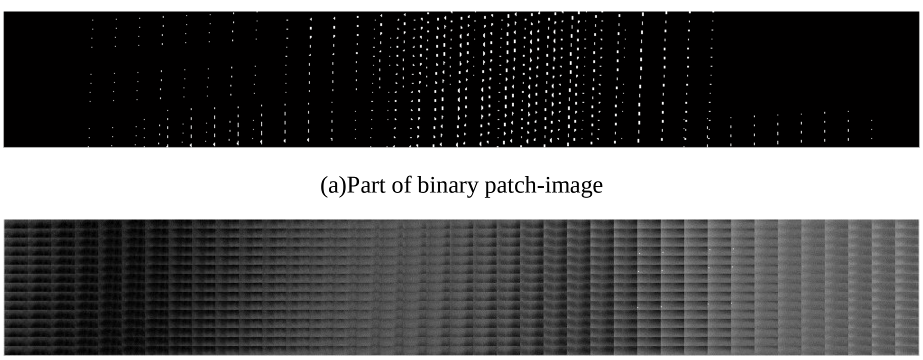
$$BW(x,y)=\begin{cases}1 & meanA-meanB>k*stdB\\ 0 & meanA-meanB\le k*stdB\end{cases} \tag{4}$$

(a)Original infrared (Test) image (b)Saliency image (c)Binary image

Fig.3. Saliency extraction

2.2 Patch-image model

We introduce a new infrared patch-image model and over-complete dictionary to the detection system, then converted the infrared small target detection into a problem-solving and optimization process of patch-image information reconstruction based on sparse representation. Briefly speaking, we obtain a series of local image patches by using a sliding window from top to bottom and left to right in a test image. Then vectorize each patch as a column of a new matrix which is called infrared test patch-image. However, it will bring about a considerable time consumption if we reconstruct each patch of the test image. So we do the same operation on binary image and get a binary patch-image which contains the location information of our interest areas. The same column in infrared patch-image and binary patch-image correspond to the same image sub-block of the test image. Hence, when the column in the binary patch-image is a full zero column vector, we believe that the corresponding position of the test image have no targets. This will greatly reduce our computational burden. The construct of patch-image was shown in Fig.4.



(a)Part of binary patch-image

(b)Part of test patch-image

Fig.4. Patch-image model

2.3 image sparse representation

Image sparse representation based on over-complete dictionary was newly proposed recent years, and is still in the stage of development. Its basic idea is to approximate and approach the target image with the least coefficients, with all this at the expense of an expanding size of the dictionary. When we use the atoms in over-complete dictionary to decompose the target images, the decomposed results are usually more sparse than using complete orthogonal basis, because the number of atoms in the dictionary is far more than the number of complete orthogonal basis. We exploit an Evaluation function to deal with the decomposition coefficient and get an evaluation value, which can be used to judge whether the image contains a target. More specifically, since the image patches cropped from the test image can be linearly represented by several elements of the over-complete small target dictionary and an error term, significant difference between the coefficients of target and background are used to distinguish targets and backgrounds. If the image sub-block contains small target, the coefficients we get from optimization function have some sparse peaks, otherwise, each item of the coefficients tends to a relatively small value and did not satisfy the sparse index. The sparse representation of images can be divided into 3 steps.

Firstly, generate a large number of sample targets by two-dimensional Gaussian model [8, 9] and construct over-complete infrared small target dictionary via model parameter adjustment as Eq.5. Process of building the dictionary has six layers of circulation, from outside to inside in order is: the center position of target x and y; the horizontal and vertical attenuation parameter σ_x and σ_y, the pixel coordinates of dictionary image i and j. Then normalize the data and reshape each atoms as one column of the dictionary data. Fig.5 show some target samples.

$$I_D(i,j) = I_{\max}\exp(-\frac{1}{2}[\frac{(i-x_0)^2}{\sigma_x^2} + \frac{(j-y_0)^2}{\sigma_y^2}])\ s.t.\ th_1 < \frac{I_{\max}}{\sigma_x}, \frac{I_{\max}}{\sigma_y} < th_2 \tag{5}$$

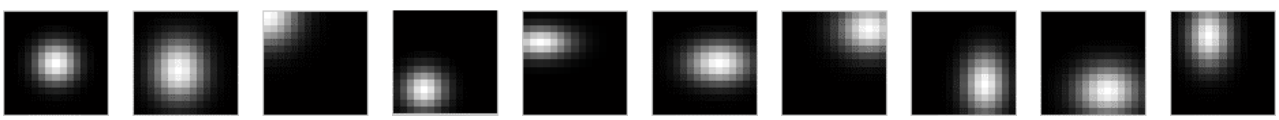


Fig.5. Part of over-complete dictionary

Secondly, reconstruct the candidate target image patch with over-complete dictionary, which can be treated as a sparse coding and optimization process. It can be modeled as Eq.6, y is the normalized test image patch, it is correlated with the dictionary atoms, and it approximately lies in the linear span of D. For entries of the dictionary are similar, the vector x can be treated as a sparse signal to be recovered. In other words, if the test image patch contains target, y can be linearly represented by a few atoms of the normalized over-complete dictionary D.

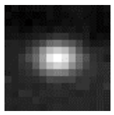
$$\underset{y\in\mathbb{R}}{\left(y\right)} \approx \underbrace{\left(d_1 \middle| d_2 \middle| \dots \middle| d_p\right)}_{D\in\mathbb{R}^{m\times p}} \underset{x\in\mathbb{R}^p,\, sparse}{\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_4 \end{pmatrix}} \quad s.t.\ p > m \tag{6}$$

For the above formula, we know that the sparse coefficient has an infinite set of solutions according to the linear algebra knowledge, but we can pick out a solution with the minimum non-zero elements from all the feasible solutions. With considering the noise, the solving process of sparse representation can be converted into the next mathematical model, represent as Eq.7.

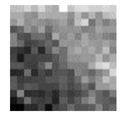
$$\min\|x\|_0\ s.t.\|\boldsymbol{D}x - \boldsymbol{y}\|_2^2 \le \varepsilon \tag{7}$$

However, the objective function of Eq.7 is a NP-hard problem base on L0-norm constraint, we can resolve this difficulty by replacing the non-convex norm $\|\bullet\|_0$ by its convex relaxation $\|\bullet\|_1$, and convert it into a constrained convex optimization problem. We use L1APG to solve it, and the approximate solution can be described as Eq.8. See appendix or literature [10, 11] for details. And the reconstruct coefficients results were show in Fig.6.

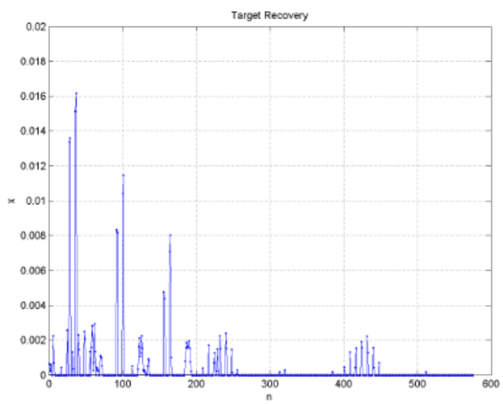
$$\min_x \|\boldsymbol{y} - \boldsymbol{D}x\|_2^2 + \lambda\|x\|_1 \quad s.t. x \geq 0 \tag{8}$$



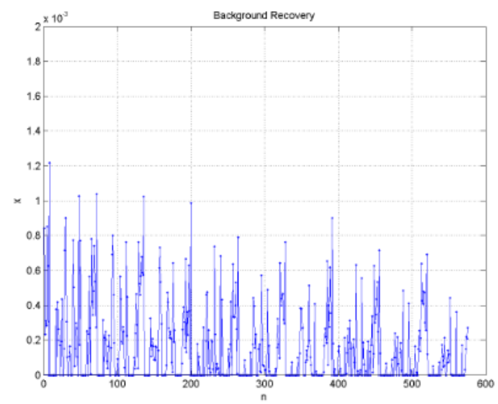
(a)Target image patch



(b)Background image patch



(c)Target recovery(S_{index}=8.0833)



(d)Background recovery(S_{index}=2.2555)

Fig.6. Reconstruct coefficients results

Thirdly, an Evaluation function was built to deal with the decomposition coefficient and get an evaluation value, which can be used to judge whether the image contains a target. k is a coefficient belongs to 0-1.

$$\underset{i=1,2,\cdots,p}{S_{index}} = \frac{num(x_i > 0)}{num\left(\left(x - k * \max\left(x_i\right)\right) > 0\right)} \tag{9}$$

2.4 False alarm rejection via Euclidean distance

For image sequence, Euclidean distance is used to reduce false alarm ratio and increase the detection accuracy of moving small targets in infrared images due to the target position correlation between frames. For the fact that noise appears as random speckles in each frame, may cause some false alarms with steep edges and some dark points. Although the moving trajectory of targets is also unknown, we can predict that targets in next frame will appear around the targets of current frame, which can be considered as a kind of position correlation of space. In other words, we can measure the Euclidean distance [12, 13] of suspected targets in adjacent frames and judge whether it is a real target or false alarm. The position discrimination has 2 steps.

First of all, for the image frame I_{j-m}, I_j, I_{j+m} calculate the correspond Euclidean distance of suspected targets as

Eq.10, of which T_j means the target in j_{th} frame, $\|\bullet\|_2$ denotes 2-norm. Dis_1 and Dis_2 are set to calculate the Euclidean distance of targets in $(j-m)_{th}$ and, j_{th} , j_{th} and $(j+m)_{th}$ frame separately.

$$\begin{cases} Dis_1 = \|T_j - T_{j-m}\|_2 \\ Dis_2 = \|T_{j+m} - T_j\|_2 \end{cases} \tag{10}$$

Next, Introduce two thresholds Th_1 and Th_2 as judgment basis. Obtain the addition and subtraction results of Dis_1 and Dis_2 , then compare them with the two thresholds. If they satisfy Eq.11, the candidate target at position T_j is a false alarm, otherwise, it is a real target. Fig.7 shows the detection result.

$$\begin{cases} |Dis_1 - Dis_2| < Th_1 \\ |Dis_1 + Dis_2| > Th_2 \end{cases} \tag{11}$$

3 Experiments

We can see that Experiment results based on simulation and real image sequences both have a considerable improve in detection rate and detection efficiency. Compared with baseline methods, the presented method has the minimum false alarm rate and the best receiver operating characteristic (Fig.8).

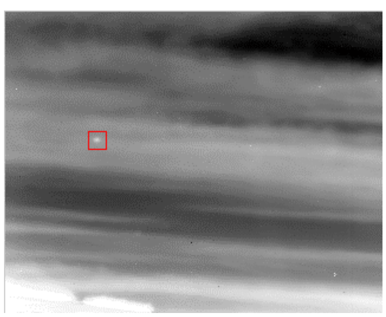


Fig.7. Detection results

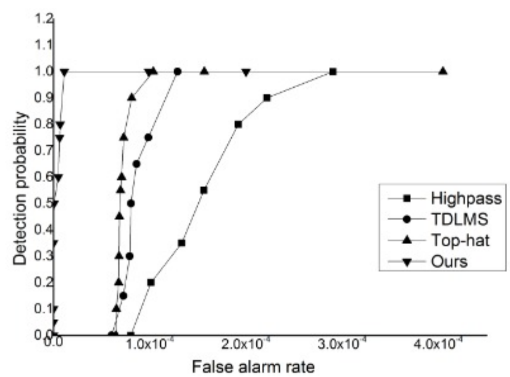


Fig8. ROC of baseline methods and ours

4 Conclusion

A novel small target detection technique based on frequency-domain saliency extraction and image sparse representation was presented. For the image patches cropped from the test image can be linearly represented by several elements of the over-complete small target dictionary and an error term, significant difference between the coefficients of target and background is used to distinguish targets and backgrounds. A saliency detection process was introduced to reduce calculation amount and improve detection efficiency. Qualitative and quantitative evaluations demonstrate that the proposed technique achieves an accurate detection performance. However, Infrared small target detection still remains a challenging task due to various conditions, we will continue our research on searching the differences between target and background characters in the process of Infrared small target detection.

5 Appendix

Algorithm of L1APG

Object: Approximate the solution of : $\arg\min_{x} \|\boldsymbol{y} - \boldsymbol{D}x\|_2^2 + \lambda\|x\|_1 \quad s.t.?x \geq 0$.

Input arguments: We are given the matrix $\boldsymbol{D}$, the vector $\boldsymbol{y}$, and the para data

b ------- $\boldsymbol{D}' * y$ transformed object vector; $\boldsymbol{A}$ ------- $\boldsymbol{D}' * \boldsymbol{D}$ transformed dictionary

para ------Lambda: Sparsity Level; Maxit: Maximal Iteration number; tol: convergence precision

Initialize: $N = \dim(y)$ **,** $\alpha_0 = \alpha_{-1} = \boldsymbol{0} \in \boldsymbol{R}^N$ **,** $\boldsymbol{t}_0 = \boldsymbol{t}_{-1} = 1$

Main Iteration: For $k = 0,1,\cdots,$ iterate until convergence

$$i \;\; \beta_{k+1} = \alpha_k + \frac{t_{k-1} - 1}{t_k}\left(\alpha_k - \alpha_{k-1}\right) \; ;ii \;\; \alpha_{k+1} = \max\left(0, \beta_{k+1} - \lambda / norm(A)\right) \; ;iii \;\; t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$$

Output: The proposed solution is $x = \alpha$

References

[1] Peng J, Zhou W. Infrared background suppression for segmenting and detecting small target[J]. Acta Electronica Sinica, 1999, 12: 012.

[2] L. Yang, J. Yang, K. Yang, Adaptive detection for infrared small target under seasky complex background, Electron. Lett. 40 (17) (2004) 1083–1085.

[3]Gu Y, Wang C, Liu B X, et al. A kernel-based nonparametric regression method for clutter removal in infrared small-target detection applications[J]. Geoscience and Remote Sensing Letters, IEEE, 2010, 7(3): 469-473.

[4] C.L.P. Chen et al., A local contrast method for small infrared target detection, Geosci. Remote Sens., IEEE Trans. 52 (1) (2014) 574–581.

[5] Y. Zhao et al., Principal curvature for infrared small target detection, Infrared Phys. Technol. 69 (2015) 36–43.

[6]JiCheng L, ZhengKang S, Tao L. Detection of spot target in infrared clutter with morphological filter[C]//Aerospace and Electronics Conference, 1996. NAECON 1996., Proceedings of the IEEE 1996 National. IEEE, 1996, 1: 168-172.

[7] YE Bin, PENG Jia-xiong. Small Target Detection Method Based on Morphology Top-Hat Operator[J]. JOURNAL OF IMAGE AND GRAPHICS, 2002,7(7):638-642.DOI:10.3969/j.issn.1006-8961.2002.07.003.

[8] ZHAO Jia-Jia, TANG Zheng-Yuan, YANG Jie, LIU Er-Qi, ZHOU Yue. Infrared small target detection based on image sparse representation [J]. JOURNAL OF INFRARED AND MILLIMETER WAVES,2011,30(2):156-161,166.

[9]Bruckstein A M, Elad M. From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images[J]. Siam Review, 2009, 51(1):págs. 34-81.

[10] Ji H. Real time robust L1 tracker using accelerated proximal gradient approach[J]. 2012, 157(10):1830-1837.

[11] Wang P, Qian W, Chen Q. Robust visual tracking with contiguous occlusion constraint[J]. Optical Review, 2015, 23:1-13.

[12]Wan, Minjie, et al. "In-frame and inter-frame information based infrared moving small target detection under complex cloud backgrounds." Infrared Physics & Technology 76 (2016): 455-467.

[13]Wang W, Li C, Shi J. A robust infrared dim target detection method based on template filtering and saliency extraction[J]. Infrared Physics & Technology, 2015, 73: 19-28.