A Robust Directional Saliency-Based Method for Infrared Small-Target Detection Under Various Complex Backgrounds

Shengxiang Qi, Jie Ma, Chao Tao, Changcai Yang, and Jinwen Tian

***Abstract*—Infrared small-target detection plays an important role in image processing for infrared remote sensing. In this letter, different from traditional algorithms, we formulate this problem as salient region detection, which is inspired by the fact that a small target can often attract attention of human eyes in infrared images. This visual effect arises from the discrepancy that a small target resembles isotropic Gaussian-like shape due to the optics point spread function of the thermal imaging system at a long distance, whereas background clutters are generally local orientational. Based on this observation, a new robust directional saliency-based method is proposed incorporating with visual attention theory for infrared small-target detection. Experimental results demonstrate that the proposed algorithm outperforms the state-of-the-art methods for real infrared images with various typical complex backgrounds.**

***Index Terms*— Infrared image, saliency detection, target detection, visual attention.**

I. Introduction

INFRARED small-target detection plays a critical role in large amounts of practical projects such as infrared warning and defense alertness, in which not only accuracy is needed but also robustness is required. Generally, since a small target just occupies very few pixels in infrared images [1], detection results are majorly affected by random noise and nonstationary clutters. Although numerous methods have been proposed aiming at specific situations, e.g., sea–sky background [2], many of them may fail in other different circumstances. Because of this, small-target detection in infrared images is still a rather difficult and challenging problem. In a general way, the task is often performed as seeking an effective way to enhance targets relative to backgrounds to amplify the signal-to-clutter ratio (SCR) of images. For this purpose, many existing algorithms are presented and can be broadly divided into two categories: the filtering method and the predicting method. In the filtering method, targets are extracted from processed images, which are obtained by filtering original images directly [2]–[4]. Alternatively, the predicting method computes the residual map between the original image and its background predicted image and then extracts targets from this map [5], [6].

Different from traditional algorithms, we formulate the problem of infrared small-target detection as salient region detection. The innovation is inspired by the fact that a small target can often attract attention of human eyes in images. This visual effect arises from the discrepancy that a small target resembles isotropic Gaussian-like shape due to the optics point spread function (PSF) of the thermal imaging system at a long distance [7], [8], whereas background clutters are generally local orientational. In other words, the "isotropic" targets are salient relative to the "orientational" backgrounds. Based on this observation, a new robust directional saliency-based method (DSBM) is proposed incorporating with visual attention theory [9]–[11] for infrared small-target detection. First, we decompose the original image into different directional channels using a designed second-order directional derivative (SODD) filter based on a facet model [12]. Second, we employ a saliency detection method, i.e., phase spectrum of Fourier transform (PFT) [10], to calculate saliency maps for each channel to enhance targets relative to backgrounds. Third, we compute the high SCR "target-saliency" map by a designed saliency fusing method that can fuse all saliency maps from each channel. Finally, we use an empirical threshold to extract targets from the target-saliency map.

II. Theory and Methodology

A. Computing the SODD Map Using the Facet Model

In the proposed algorithm, we first decompose the original image into different directional channels. Due to the capability of detecting ramps and isolated uplift of signals, the SODD filter is used as the decomposition tool to compute SODD maps for each channel. It can transform clutters and targets into fixed orientational striplike texture and Gaussian-like spots (although it is not rigorous Gaussian), respectively. A procedure of processing a 2-D signal using our designed SODD filter is displayed in Fig. 1. The original map [see Fig. 1(a)] contains a ramp and a Gaussian uplift representing clutters and a small target, respectively. Fig. 1(b) is the primitive SODD map in Fig. 1(a) in an assigned direction (channel), and Fig. 1(c) is the amended SODD map in which the ramp and the

Manuscript received March 27, 2012; accepted July 26, 2012. Date of publication September 10, 2012; date of current version November 24, 2012. This work was supported in part by the National Natural Science Foundation of China under Grant 61004111, Grant 61074156, and Grant 61104191 and in part by the National 973 Plan of China under Grant 2012CB719900. (*Corresponding author: J. Ma.*)

S. Qi, J. Ma, C. Yang, and J. Tian are with the Institute for Pattern Recognition and Artificial Intelligence, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: shengxiang.qi@gmail.com; majie@mail.hust.edu.cn).

C. Tao is with the School of Geosciences and Info-Physics, Central South University, Changsha 410083, China.

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

Digital Object Identifier 10.1109/LGRS.2012.2211094

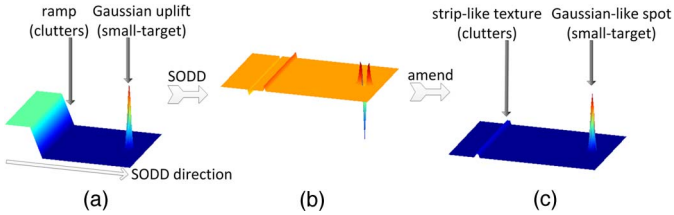

Fig. 1. Procedure of processing a 2-D signal using our designed SODD filter.

Gaussian uplift are transformed into shape of striplike texture and Gaussian-like spot, respectively. Specifically, the designed filter is described below in details.

Because the Laplacian operator is susceptible to noise and lack of smoothness, we use a facet model [12] to design the SODD filter. This model assumes that the underlying gray-level intensity surface for each pixel can be approximated by a bivariate cubic function in its neighborhood. Assume that $R = \{-2 \quad -1 \quad 0 \quad 1 \quad 2\}$ and $C = \{-2 \quad -1 \quad 0 \quad 1 \quad 2\}$, the 5×5 neighborhood window for a pixel can be denoted by $R \times C$, in which the pixel is at the center (0, 0). Then, the intensity value of the pixel expressed by the bivariate cubic function defined over $R \times C$ is calculated as follows:

$$f(r,c) = \sum_{i=1}^{10} K_i \cdot P_i(r,c) \tag{1}$$

where $\{P_i(r,c)\} = \{1, r, c, r^2 - 2, rc, c^2 - 2, r^3 - (17/5)r, (r^2-2)c, r(c^2-2), c^3 - (17/5)c\}$ is the set of discrete orthogonal polynomials in which $(r,c) \in R \times C$; K_i are coefficients for the bivariate cubic function. Given (1), the second-order partial derivatives evaluated along the row and column at the center pixel (0, 0) in $R \times C$ are

$$\left.\frac{\partial^2 f(r,c)}{\partial r^2}\right|_{(0,0)} = 2K_4 \qquad \left.\frac{\partial^2 f(r,c)}{\partial r \partial c}\right|_{(0,0)} = K_5$$
$$\left.\frac{\partial^2 f(r,c)}{\partial c^2}\right|_{(0,0)} = 2K_6. \tag{2}$$

Note that the values of coefficients K_i $(i = 4, 5, 6)$ are related to different pixels (x, y) in the image; we denote these coefficients as $K_i(x, y)$ $(i = 4, 5, 6)$ to avoid confusion. According to [12], these coefficients are determined by least squares surface fitting and orthogonal property of polynomials with the result $K_i(x,y) = \sum_r \sum_c (I(x+r, y+c) P_i(r,c) / \sum_r \sum_c P_i^2(r,c))$ $(i = 4, 5, 6)$, in which $I(x, y)$ is the image intensity of (x, y). For actual calculation, they can be individually computed by a linear combination of the intensity values in the symmetric $R \times C$ neighborhood of pixels, each of which has a weight $w_i(r,c) = (P_i(r,c) / \sum_r \sum_c P_i^2(r,c))$ $(i = 4, 5, 6)$. Therefore, given the set of discrete orthogonal polynomials, i.e., $\{P_i(r,c)\}$, the three weight kernels are obtained as follows:

$$W_4 = \frac{1}{70}\begin{bmatrix} 2 & 2 & 2 & 2 & 2 \\ -1 & -1 & -1 & -1 & -1 \\ -2 & -2 & -2 & -2 & -2 \\ -1 & -1 & -1 & -1 & -1 \\ 2 & 2 & 2 & 2 & 2 \end{bmatrix}$$

$$W_5 = \frac{1}{100}\begin{bmatrix} 4 & 2 & 0 & -2 & -4 \\ 2 & 1 & 0 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \\ -2 & -1 & 0 & 1 & 2 \\ -4 & -2 & 0 & 2 & 4 \end{bmatrix}$$

$$W_6 = W_4^T. \tag{3}$$

Given (2), we then deduce the formula of SODD along direction vector l at pixel (x_0, y_0) as follows:

$$\begin{aligned}\left.\frac{\partial^2 f(x,y)}{\partial l^2}\right|_{(x_0,y_0)} &= \left[f_{xx}(x,y)\cos^2\alpha + 2f_{xy}(x,y)\right. \\ &\quad \left.\times \cos\alpha\cos\beta + f_{yy}(x,y)\cos^2\beta\right]\big|_{(x_0,y_0)} \\ &= 2K_4(x_0,y_0)\cos^2\alpha + 2K_5(x_0,y_0) \\ &\quad \times \cos\alpha\cos\beta + 2K_6(x_0,y_0)\cos^2\beta \end{aligned} \tag{4}$$

where α is the angle between l and the x-axis (row of image); β is the angle between l and the y-axis (column of image).

However, note that the target values of the primitive SODD map [see Fig. 1(b)] acquired by (4) are less than zero, whereas the neighborhood values of the target along the assigned direction are larger than zero. This effect gives rise to deviation of the target from Gaussian-like shape, which will lead to disturbance for saliency detection. Accordingly, we need to amend the SODD map, which consists of: 1) set map values larger than zero to be zero; 2) inverse the whole map after all values are normalized to a fixed range [0, 1]; and 3) filter the map using a 3×3 smoothing mask to smooth edges of targets and strips. After this, the shape characteristics of the target and clutters are highlighted in the amended map [see Fig. 1(c)].

B. Computing the Directional Saliency Map Using PFT

After computing SODD maps for each directional channel, characteristics of targets and clutters are separated into two types, i.e., the former are Gaussian-like and the latter are strip-like. In general as a fact, the "conspicuous" small target with its particular shape and brightness often attracts attention of human eyes relative to the "ordinary" clutters in SODD maps. Based on this observation, we use the saliency detection method [9]–[11] derived from visual attention theory to highlight the Gaussian-like targets. This method often simulates the brain and vision system to identify the relevant salient regions, which accords with the mechanism of the human visual system. Additionally, the PFT method [10] has very low computational complexity among most saliency detection approaches; it is suitable to compute saliency maps for SODD maps. For explanation, two instances of PFT are displayed in Fig. 2. Each shows an original map at the left containing a Gaussian spot (small target) and orientational strips (clutters) and its PFT saliency map at the right. It is legible that Gaussian spots are highlighted as salient regions while strips are inhibited, which accords with our visual effects.

PFT [10] is an improved method from [9]. It proves that the location information of salient regions can be acquired from the image's PFT. More specifically, detailed description on PFT can

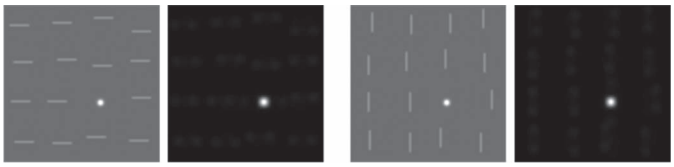

Fig. 2. Two instances of PFT, each of which shows an original map at the left containing a Gaussian spot (small target) and orientational strips (clutters) and its PFT saliency map at the right.

be referred to [9] and [10]. In this letter, we briefly express the procedure of computing saliency maps using PFT as follows:

$$f(u,v) = F\left[I(x,y)\right] \tag{5}$$

$$p(u,v) = P\left[f(u,v)\right] \tag{6}$$

$$S(x,y) = g(x,y) * \left\| F^{-1}\left[e^{i\cdot p(u,v)}\right]\right\|^2 \tag{7}$$

where $I(x,y)$ denotes the image, which is namely each SODD map in DSBM; $F[.]$ and $F^{-1}[.]$ denote the Fourier transform and the inverse Fourier transform, respectively; $P[.]$ represents the phase spectrum of the image; $g(x,y)$ is a 2-D Gaussian filter $(\sigma = 2.5)$; and $S(x,y)$ is the saliency map. These expressions indicate that the saliency map can be simply obtained by computing the phase spectrum of the original image and then calculating the inverse Fourier transform using them alone, in which the amplitude spectrum is set to be 1. The square and convolution in (7) are just applied to standout relevant salient regions.

C. Computing the Target-Saliency Map Using a Designed Saliency Fusing Method

Directional saliency maps of different channels have different magnitudes. It is infeasible to fuse these maps directly since they represent *a priori* not comparable modalities with different dynamic ranges. To solve this problem, we employ a normalization operator $\mathbf{N}(.)$ [11], which can globally promote maps in which a small number of strong peaks of activity (conspicuous locations) are present, while globally suppressing maps that contain numerous comparable peak responses. It consists of: 1) normalize the values in the map to a fixed range $[0,\ldots,M]$ in order to eliminate modality-dependent amplitude differences; 2) find the location of the map's global maximum M and computing the average $\bar{m}$ of all its other local maxima; and 3) globally multiply the map by $(M-\bar{m})^2$.

Then, we design an effective saliency fusing method to fuse all the directional saliency maps into a final "target-saliency" map, which is expressed as follows:

$$\Re = \sum_{g} \mathbf{N}(S_g) \cdot \mathbf{N}(\perp S_g) \tag{8}$$

where g represents the number of orthogonal direction groups; S_g is defined as the directional saliency map and $\perp S_g$ is the orthogonal directional saliency map of S_g in group g; and $\Re$ is the fused map, i.e., the "target-saliency" map. By means of multiplication between orthogonal directions, this expression can inhibit the orientational striplike background clutters and amplify the isotropic Gaussian-like targets.

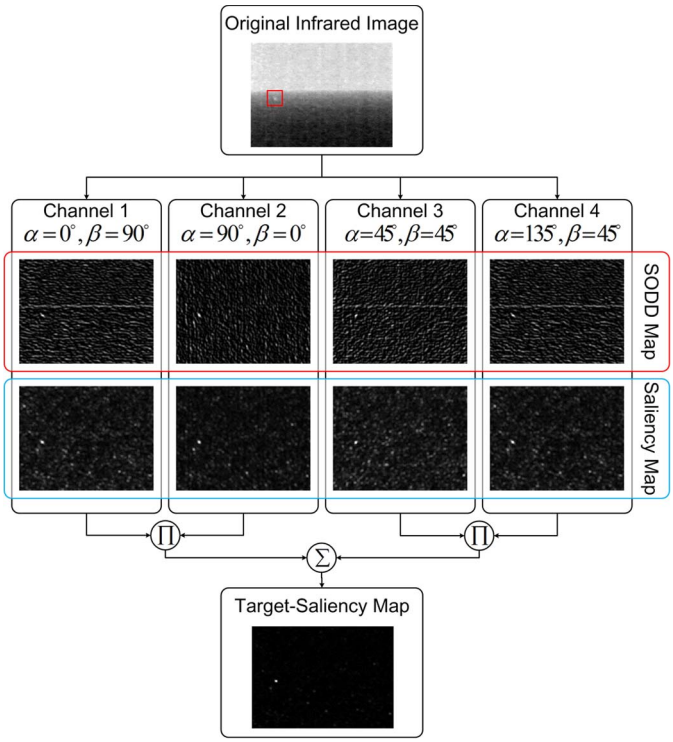

Fig. 3. Flow diagram of acquiring the high SCR target-saliency map by DSBM.

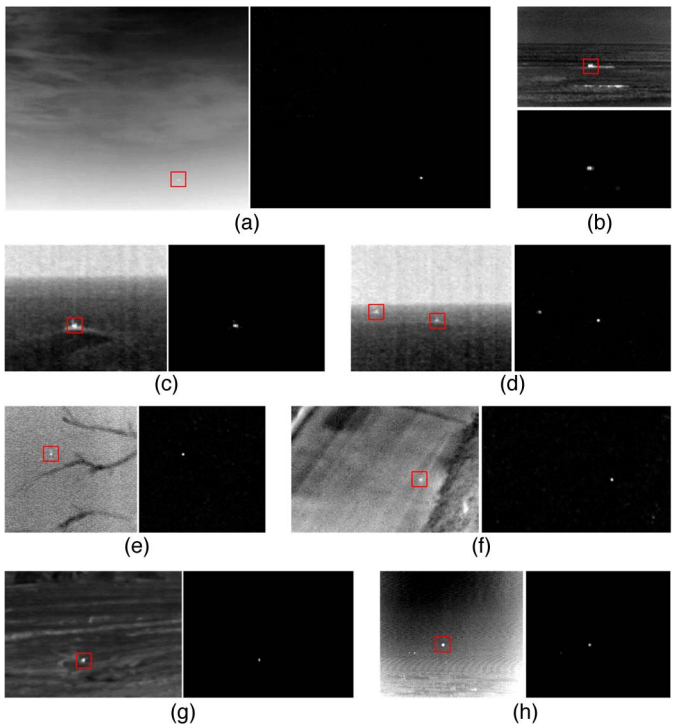

Fig. 4. (Left) Eight groups of real infrared images and (right) their target-saliency maps processed by DSBM.

III. DSBM for Infrared Small-Target Detection

Fig. 3 shows a flow diagram of acquiring the target-saliency map by DSBM. The original image is decomposed into two orthogonal direction groups (four directional channels) by an

TABLE I
EVALUATION INDICATORS OF FIVE DETECTION METHODS

Detective Methods	Evaluation Indicators	Image (a)	Image (b)	Image (c)	Image (d)	Image (e)	Image (f)	Image (g)	Image (h)
Top-hat	SCR Gain	7.7131	1.4145	11.8914	17.4571	1.5965	1.9766	1.6311	7.3117
	BSF	5.7620	1.1836	7.0213	5.0163	0.9009	1.3560	1.8976	4.5776
	Time(s)	**0.0424**	**0.0073**	**0.0142**	**0.0151**	**0.0103**	**0.0145**	0.0713	**0.0094**
BHPT	SCR Gain	4.6845	1.2268	9.0171	15.0385	1.5115	1.3018	1.9172	6.4509
	BSF	5.0238	1.1882	6.4297	5.3925	1.2506	1.5877	2.5415	6.8426
	Time(s)	1.4586	0.1295	0.3198	0.3089	0.2948	0.3947	0.6848	0.3791
Facet-model	SCR Gain	8.6069	1.9799	15.5852	28.4302	3.2212	3.1803	6.2232	15.8132
	BSF	20.8286	6.0264	30.8193	22.0367	4.6585	4.4876	14.6685	30.8406
	Time(s)	0.1217	0.0515	0.0702	0.0608	0.0562	0.0702	0.0936	0.0655
LS-SVM	SCR Gain	8.2377	**4.6283**	42.4939	55.2414	7.7211	5.4459	16.6291	28.5137
	BSF	9.8500	**8.1279**	**42.5577**	26.2740	6.4869	5.2214	**23.8054**	42.0196
	Time(s)	0.0666	0.0103	0.0175	0.0170	0.0176	0.0204	**0.0398**	0.0234
DSBM	SCR Gain	**131.2548**	4.1978	**46.8349**	**173.6772**	**30.0032**	**21.3833**	**19.4015**	**62.4788**
	BSF	**71.9620**	4.6840	34.6033	**52.2253**	**11.4939**	**13.7805**	23.6038	**69.4824**
	Time(s)	1.2664	0.1834	0.3181	0.3303	0.2636	0.3504	0.6992	0.4438

SODD filter, and then directional saliency maps are computed by PFT; finally, the target-saliency map is obtained using the saliency fusing method. Detailed process steps of DSBM for infrared small-target detection are described as follows.

Step 1) Select appropriate orthogonal direction groups according to actual requirements. Generally, two groups, one for channel 1 ($\alpha = 0^\circ$ $\beta = 90^\circ$) and channel 2 ($\alpha = 90^\circ$ $\beta = 0^\circ$), another for channel 3 ($\alpha = 45^\circ$ $\beta = 45^\circ$) and channel 4 ($\alpha = 135^\circ$ $\beta = 45^\circ$), are sufficient for detection.

Step 2) Do convolutions between weight kernels (3) and the original infrared image to compute fitting coefficients $K_i(x, y)$ $(i = 4, 5, 6)$ for each pixel.

Step 3) Compute SODD maps for each directional channel using designed SODD filter (4), and then amend these maps.

Step 4) Compute directional saliency maps for each channel using the PFT method (5), (6), (7).

Step 5) Fuse directional saliency maps of all channels using the designed saliency fusing method (8) to acquire the target-saliency map.

Step 6) Choose the segmentation threshold, which is 0.35 times of the maximal value (an empirical value from experiments), to extract a small target from the target-saliency map. Note that although approaches of threshold selection are various, such a simple pattern is adequate in most cases according to experiments.

IV. EXPERIMENTS AND RESULTS

In order to validate the performance and robustness of our proposed algorithm for infrared small-target detection, experiments are performed on eight typical real infrared images with sky-cloud [see Fig. 4(a)], sea-wave [see Fig. 4(b)–(d)], tree-fork [see Fig. 4(e)], road-side [see Fig. 4(f)], and continent-ground [see Fig. 4(g) and (h)] clutters representing five kinds of complex backgrounds. These images and their corresponding target-saliency maps processed by DSBM are shown in Fig. 4. Meanwhile, we compare our algorithm with four state-of-the-art approaches, including morphological detection method (Top-hat) [5], adaptive Butterworth high-pass filter method (BHPT) [2], facet-based method (facet model) [3], and least squares support vector machine-based method (LS-SVM) [4]. For fair comparison, we adopt two common evaluation indicators [13], i.e., SCR Gain and background suppression factor (BSF) defined as follows:

$$\text{SCR Gain} = \frac{(S/C)_{\text{out}}}{(S/C)_{\text{in}}} \qquad \text{BSF} = \frac{C_{\text{in}}}{C_{\text{out}}} \tag{9}$$

where S and C represent the signal amplitude and clutter standard deviation, respectively. In experiments, S is calculated as the difference between the mean of target and background values, and C is calculated as the standard deviation of background. The subscripts *in* and *out* express the images before and after the detection. Performed on MATLAB R2010 software in a 3.10-GHz PC, experimental results, including SCR Gain, BSF, and run time, are listed in Table I.

Data in Table I show that DSBM has good performance in SCR Gain and BSF, particularly the former. Specifically, SCR Gain reflects the amplification of target signals relative to backgrounds after and before processing, whereas BSF only expresses the suppression level of backgrounds without any target information. Thus, compared with BSF, SCR Gain is more significant in target detection. Considering this fact, DSBM actually outperforms other methods in most cases, including Fig. 4(a) and (c)–(h). The reason for a little lower BSF of DSBM in some cases is that saliency detection does not totally suppress backgrounds in essence but only to amplify the salient targets relative to clutters to make signals' conspicuousness. In addition, in images such as those in Fig. 4(b), LS-SVM has a little better performance than DSBM. This is because the sea-wave clutters close to the target result in deviation from Gaussian-like shape of the target, whereas LS-SVM is almost unaffected since the position of the target's extreme point almost does not change by this deformation. However, in situations such as Fig. 4(e), (f), and (h), which have strong clutters and noise, DSBM still performs better than LS-SVM. This fact can also prove that the proposed method is more robust than other methods in such harsh circumstances. In spite that

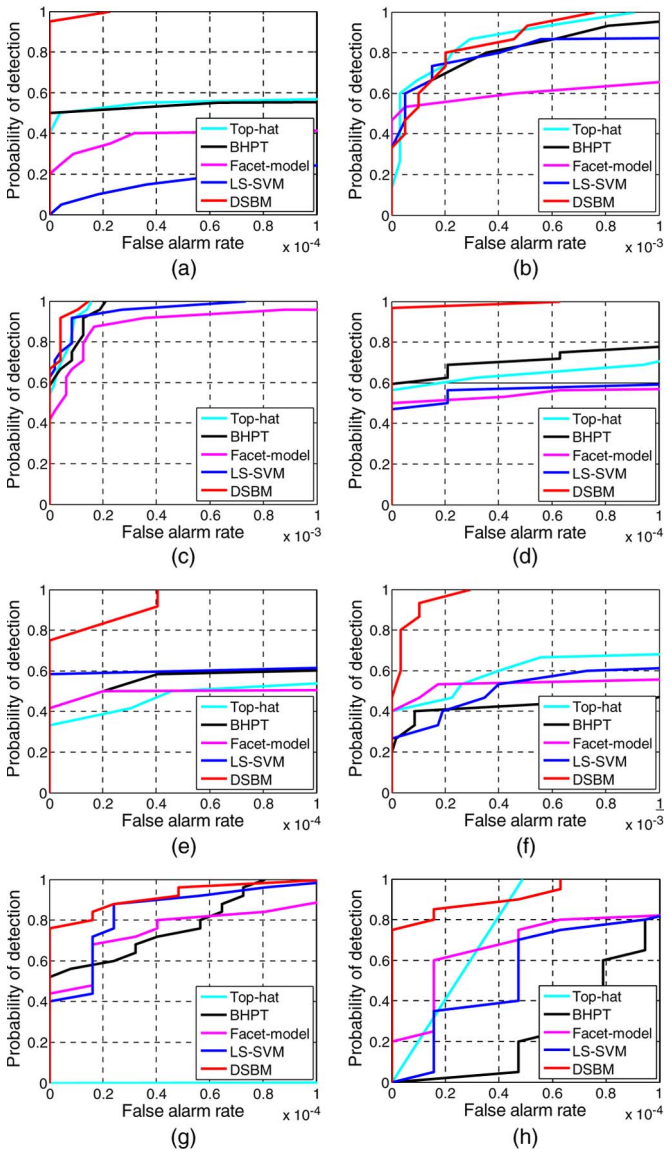

Fig. 5. Comparison of ROC curves obtained by five methods for each test image in Fig. 4.

DSBM is more time consuming than other methods, it has an acceptable order of magnitude just as traditional method BHPT and can be improved by parallel computing or using more efficient saliency detection methods.

Furthermore, in Fig. 5, we draw receiver operation characteristic (ROC) curves of the detection results for each method to provide a quantitative comparison of detection performance. As well as the description in [6], the ROC curves represent the varying relationship of detection probability P_d and false alarm rate P_f. P_d is defined as the ratio of the number of detected pixels to the number of real target pixels, and P_f is the ratio of the number of false alarms to the total number of pixels in the whole image. Compared with "general" evaluation terms such as SCR Gain and BSF, ROC curves are more "specific" in describing the performance of target extraction using a threshold on processed high SCR maps. In this sense, they are more significant than SCR Gain and BSF. In Fig. 5, the results show that DSBM has better ROC curves than other methods, which indicates that the detection probability of the proposed method is always higher than others' under the same false alarm rate. In other words, DSBM is more accurate than other methods for target detection. Note that the morphological detection algorithm can also get good ROC curves for several images such as those in Fig. 4(b) and (h). However, this method largely depends on the suitable structural elements according to priori knowledge of targets. Once there is no suitable structural element to the image, the results may become poor such as Fig. 5(g). In summary, DSBM actually shows its great performance and robustness for infrared small-target detection under various complex backgrounds.

V. Conclusion

In this letter, we have presented a robust DSBM for infrared small-target detection inspired by visual effect of human eyes between a small target and backgrounds. Saliency detection is used to enhance target signals in this method. For fair and comprehensive evaluation, real infrared images with different typical complex backgrounds and common evaluation indicators, including SCR Gain, BSF, run time, and ROC curves, are used in the experiments. The results indicate that the proposed algorithm has good performance and robustness and performs better than the state-of-the-art methods.

References

[1] F. Zhang, C. F. Li, and L. Shi, "Detecting and tracking dim moving point target in IR image sequence," *Infrared. Phys. Technol.*, vol. 46, no. 4, pp. 323–328, Apr. 2005.

[2] L. Yang, J. Yang, and K. Yang, "Adaptive detection for infrared small target under sea–sky complex background," *Electron. Lett.*, vol. 40, no. 17, pp. 1083–1085, Aug. 2004.

[3] G.-D. Wang, C.-Y. Chen, and X.-B. Shen, "Facet-based infrared small target detection method," *Electron. Lett.*, vol. 41, no. 22, pp. 1244–1246, Oct. 2005.

[4] P. Wang, J. W. Tian, and C. Q. Gao, "Infrared small target detection using directional highpass filters based on LS-SVM," *Electron. Lett.*, vol. 45, no. 3, pp. 156–158, Jan. 2009.

[5] V. T. Tom, T. Peli, M. Leung, and J. E. Bondaryk, "Morphology-based algorithm for point target detection in infrared backgrounds," in *Proc. Signal Data Process. Small Targets*, Orlando, FL, 1993, vol. 1954, pp. 2–11.

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

[7] T. Soni, J. R. Zeidler, and W. H. Ku, "Performance evaluation of 2-D adaptive prediction filters for detection of small objects in image data," *IEEE Trans. Image Process.*, vol. 2, no. 3, pp. 327–340, Jul. 1993.

[8] D. S. K. Chan, D. A. Langan, and D. A. Staver, "Spatial processing techniques for the detection of small targets in IR clutter," in *Proc. SPIE*, Apr. 1990, vol. 1305, pp. 53–62.

[9] X. D. Hou and L. Q. Zhang, "Saliency detection: A spectral residual approach," in *Proc. CVPR*, 2007, pp. 1–8.

[10] C. L. Guo, Q. Ma, and L. M. Zhang, "Spatio-temporal saliency detection using phase spectrum of quaternion Fourier transform," in *Proc. CVPR*, 2008, pp. 1–8.

[11] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[12] R. Haralick, "Digital step edges from zero crossing of second directional derivatives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 1, pp. 58–68, Jan. 1984.

[13] C. I. Hilliard, "Selection of a clutter rejection algorithm for real-time target detection from an airborne platform," in *Proc. SPIE*, Apr. 2000, vol. 4048, pp. 74–84.