Infrared small moving target detection method based on graph matching

Xiabin Dong (董夏斌)*, Yongbin Zheng (郑永斌), Shengjian Bai (白圣建), Wanying Xu (徐婉莹), and Xinsheng Huang (黄新生)

College of Mechatronic Engineering and Automation, National University of Defense Technology, Changsha 410073, China

**Corresponding author: dxb_nudt@163.com*

Received June 23, 2014; accepted October 11, 2014; posted online November 14, 2014

We apply graph matching method to detect infrared small moving targets using image sequences. Candidates (interest points) detected in the first frame form one graph and the same candidates in the last frame form another one. The real moving targets are extracted by matching these two graphs. Experimental results demonstrate that the proposed method is robust and efficient to the translation and rotation of the background.

OCIS codes: 100.4999, 100.2000, 110.3080, 040.2480.
doi: 10.3788/COL201412.121002.

The ability to detect moving targets in infrared images or videos has a major impact on the areas of prewarning, precision guide, and so on[1–6]. Yang *et al.*[1] used an adaptive Butterworth high-pass filter to detect small target. Wang *et al.*[2] proposed a small target detection method based on the cubic facet model. Deshpande *et al.*[3] provided a moving target detection method based on max-mean and max-median filters. Zhang *et al.*[4] presented an algorithm for detecting dim moving point targets under the condition of constant false-alarm ratio. Bae[5] used spatial bilateral filter and temporal cross product of temporal pixels to detect moving targets. The performances of the abovementioned methods are good when the background is static. To enhance the efficiency of small moving target detection, we propose a new method based on graph matching which is robust to the translation and rotation of the background. The experiments reflect the method as being efficient and robust.

Infrared small targets look like small bright dots, so they can be detected by difference of Gaussian (DOG) filters. The candidates are called as interest points in this letter. The difference of two Gaussians with different standard deviations forms a DOG filter.

$$\begin{aligned}\mathrm{DOG}(x,y,\sigma_1,\sigma_2) &= G(x,y,\sigma_1)-G(x,y,\sigma_2)\\ &= \frac{1}{2\pi}\left(\frac{1}{\sigma_1^2}\exp\left(-\frac{x^2+y^2}{2\sigma_1^2}\right)-\frac{1}{\sigma_2^2}\exp\left(-\frac{x^2+y^2}{2\sigma_2^2}\right)\right),\end{aligned} \quad (1)$$

where $\sigma_1 \in \{2,3,4\}$, $\sigma_2 \in \{3,4,5\}$, and $\sigma_1 < \sigma_2$, which is similar to the center surround mechanism of human visual system.

A group of feature maps with different scales are obtained after filtering the first frame of image sequence using DOG filters. All of these feature maps are combined together to get the saliency map. The rule of the combination of feature maps is given as

$$\mathrm{Sali}(x,y) = \max_{\sigma_1,\sigma_2}\{I(x,y) * \mathrm{DOG}(x,y,\sigma_1,\sigma_2)\} \quad \sigma_1 < \sigma_2, \quad (2)$$

where $I(x, y)$ stands for the input image and the symbol "*" denotes the operation of convolution. Graph *b* in Fig. 1 shows the result of this step, that is, the saliency map.

Point (x, y) is an interest point if $\mathrm{Sali}(x, y)$ is local maximum and greater than threshold th1. The total number of interest points is N. The detected interest points are marked by rectangles in graph *b* of Fig. 1.

Tracking method[7–9] can be used to track these interest points detected in the first frame to find the corresponding interest points in the last frame. In this letter, we use an interesting tracking method[7]. The process of tracking the nth interest point is described as follows, and the method of tracking the others is same.

Step 1: Estimating the nth interest point's velocity v_{i-1}^n and acceleration a_{i-1}^n in the $(i-1)$th frame using

$$\begin{aligned} v_{i-1}^n &= (IP_{i-1}^n - IP_{i-2}^n)\big/t,\\ a_{i-1}^n &= (v_{i-1}^n - v_{i-2}^n)\big/t. \end{aligned} \quad (3)$$

Then the estimated position of the nth interest point in the ith frame $I\tilde{P}_i^n$ can be expressed

$$I\tilde{P}_i^n = IP_{i-1}^n + v_{i-1}^n \cdot t + 0.5 \cdot a_{i-1}^n \cdot t^2 + \tilde{\varepsilon}, \quad (4)$$

where IP_{i-1}^n is the detected position of the nth interest point in the $(i-1)$th frame and $\tilde{\varepsilon}$ is the estimated error.

Step 2: A local region (the area is $w \times w$) centered at $I\tilde{P}_i^n$ in the ith frame of the original image is filtered using DOG filters. Then the local saliency map of the nth interest point is obtained.

Step 3: A Gaussian window centered at $I\tilde{P}_i^n$ is added in the local saliency map to reduce the influences of points away from the center. The point with maximum saliency value in the local region is treated as the real interest point in the ith frame, the detected position is IP_i^n.

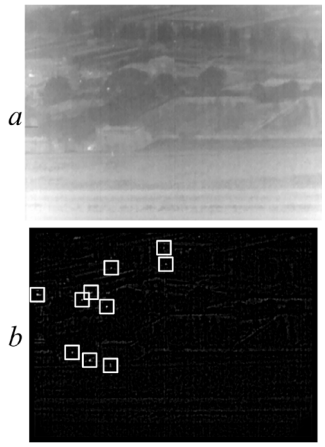

Fig. 1. Graphs a and b are input image and saliency map using DOG filters.

In this process, these interest points are tracked for L frames, and an interest point is eliminated if it deviates from the field. After this process, two graphs are formed. Interest points in the first frame, being treated as dots, and the distances between each other, being treated as the edges, form the first graph. Those of the last frame form the second graph. The moving targets are detected using the graph matching method.

Suppose that the rotation angle and the translation between the backgrounds of the first and the last frames are $\Delta\theta$ and ΔX, respectively. The rotation center is X. Then the nth background point X^n without considering the scale changes is established as

$$\begin{aligned} X_l^n - X &= W_f^l(\Delta\theta)(X_f^n - X - \Delta X), \\ W_f^l(\Delta\theta) &= \begin{bmatrix} \cos\Delta\theta & -\sin\Delta\theta \\ \sin\Delta\theta & \cos\Delta\theta \end{bmatrix}, \end{aligned} \tag{5}$$

$$\begin{aligned} \left\|X_l^n - X_l^m\right\|_2 &= \left\| \begin{matrix} W_f^l(\Delta\theta)(X_f^n - X - \Delta X) - W_f^l(\Delta\theta) \\ (X_f^m - X - \Delta X) \end{matrix} \right\|_2 \\ &= \left\|W_f^l(\Delta\theta)(X_f^n - X_f^m)\right\|_2 \\ &= \left\|X_f^n - X_f^m\right\|_2, \end{aligned} \tag{6}$$

where W_f^l is the transformation matrix of the coordinate systems from the first frame to the last frame, and X_f^n and X_l^n are the positions of the nth background points of the first and the last frames, respectively.

Equation (6) explains that the values of 2-norms of the difference of two background points are equal, that is, the length between two background points does not change. Let us consider the graphs shown in Fig. 2.

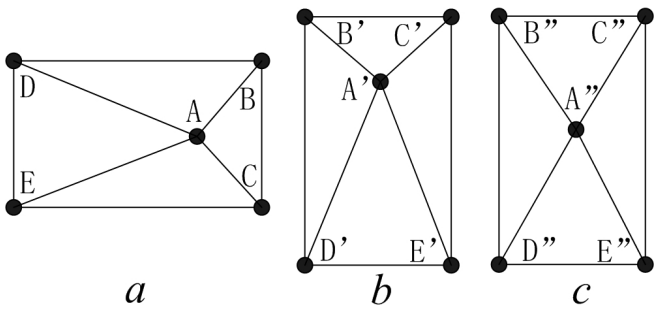

Fig. 2. Graph a, graph b is formed by rotating graph a, and graph c, point A'' has a positional change.

The dots in these graphs stand for interest points. The distance between two interest points forms the edge of the graph. Graphs b and c are formed through rotating graph a by 90° counterclockwise. Otherwise, the interest point A'' in graph c has a motion of translation. Figure 2 illustrates that the relative positions of two points in graph b do not change comparing with those in graph a. If there is a moving target, such as point A'' in graph c, the relative positions between point A'' and others are changed. In light of this regulation, a novel small moving target detection method based on graph matching is proposed. We use the difference of relative positions in the construction of the metric which is formulated as

$$E(n) = \sum_{m=1}^{N} \left(\left\|X_f^n - X_f^m\right\|_2 - \left\|X_l^n - X_l^m\right\|_2 \right)^2, \quad n = 1, 2, ..., N \tag{7}$$

where n stands for the nth interest point. The larger the $E(n)$, the more likely the nth interest point is the moving target. The criterion of the moving targets is given as

$$\text{The } n\text{th point is} \begin{cases} \text{the target point} & E(n) \geq s \cdot \max(E), \\ \text{the background point} & \text{else}, \end{cases} \tag{8}$$

where s is a parameter. This factor makes this method detect not only one moving target but also more targets.

Figure 3 gives the detection result of moving targets in Fig. 1 using graph matching. Figure 3(a) shows the E values of different points and Fig. 3(b) shows the detected moving targets.

To evaluate the target detection performance, the proposed methods are compared with other five sophisticated methods: high-pass filter method[1], max-median method and max-mean method[3], Zhang's method[4], and Bae's method[5] by using a mass of image sequences.

In this letter, we present only the results of four sequences (Seq1–Seq4): not only one moving target exists in Seq1, the background of Seq2 has a rotational motion, the background of Seq3 has only a translation motion, and Seq4 has a static background. Three parameters are needed in our algorithms. Threshold th1 is needed when detecting the interest points. The total number of the frames L is needed in tracking process. Different sequences have different thresholds in our method. Table 1 gives the best threshold of each sequence. The last parameter s is an important parameter of graph matching. Results of many experiments

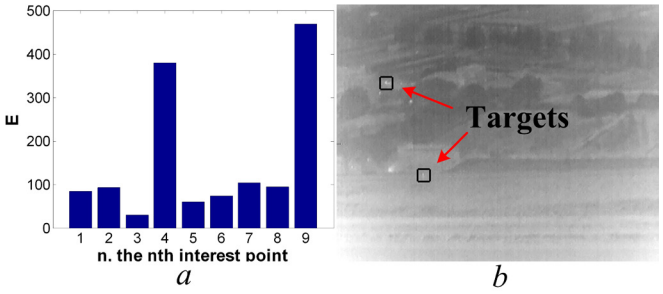

Fig. 3. Results of moving target detection using graph matching: (a) E values of different points and (b) detected moving targets.

indicate that $s = 0.75$ is the best choice. So we choose $s = 0.75$ in our experiments.

Figure 4 shows the subjective comparison of different methods. The result of each method is its best performance which is obtained by adjusting the parameters of this method. Figure 4(a) shows the original images of each sequence and all moving targets are marked by circles. The results of these five sophisticated methods have many false targets using Seq1–Seq4 (Figs. 4(b)–(f)). However, the results are much better using Seq5 and Seq6. That is because these methods do not take the dynamic backgrounds into consideration. The results of the proposed method which are shown in Fig. 4(g) are much better than those of the compared methods.

In order to discuss the detection performance of these methods further, the objective comparison by target-to-clutter ratio (TCR) is presented as

$$\mathrm{TCR} = \frac{N_{\mathrm{TT}}}{N_{\mathrm{TT}} + N_{\mathrm{FC}} + N_{\mathrm{FT}}}, \tag{9}$$

where N_{TT} and N_{FT} are the quantities of true targets and false targets detected in images, respectively, and N_{TC} and N_{FC} are the quantities of true clutters and false clutters detected in images, respectively. The TCR values of these sequences obtained by different methods are shown in Table 2 where symbols (b)–(g) stand the same as Fig. 4. These results are still their best performances. From the data in Table 2, we have the same conclusion.

Sometimes, it is not enough to mention that the method with better best performance is more efficient and more accuracy than the method with worse best performance. We need to compare the results under different conditions such as different false-alarm rates. So the performance of these algorithms is evaluated by a receiver operating characteristic (ROC) using Seq1 further.

Table 1. Parameters Used in Experiments

	Seq1	Seq2	Seq3	Seq4
th1	0.059	0.035	0.035	0.200
L	10	20	10	10

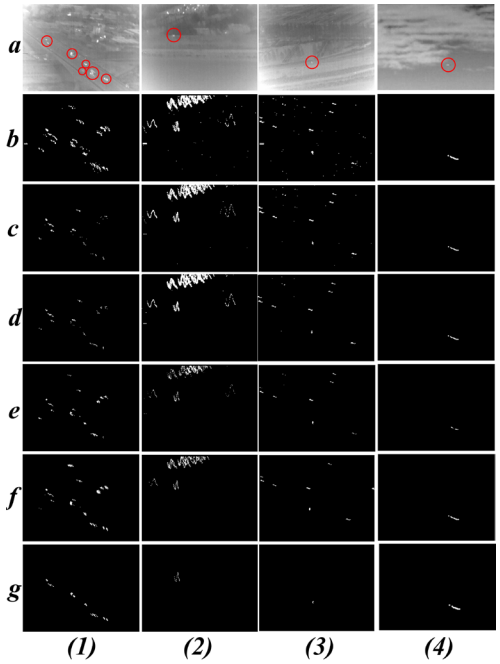

Fig. 4. Experimental results of test infrared sequences: (a) original images, and detection results of (b) high-pass filter, (c) max-median, (d) max-mean, (e) Zhang's method, (f) Bae's method, and (g) graph matching.

A ROC curve plots the true positive ratio, TPR as a function of FPR, the false positive ratio in sequence, given by

$$\mathrm{TPR} = \frac{N_{\mathrm{TT}}}{N_{\mathrm{TT}} + N_{\mathrm{FC}}}, \mathrm{FPR} = \frac{N_{\mathrm{FT}}}{N_{\mathrm{FT}} + N_{\mathrm{TC}}}. \tag{10}$$

These values are obtained by adjusting the thresholds of these methods. Figure 5 shows the ROC curves. According to the properties of ROC curve, the proposed method is much better than others.

In conclusion, sophisticated methods are sensitive to the translation and rotation of the background. So we propose an infrared small target detection method based on graph matching to conquer this shortcoming. Experimental results verify that the proposed method

Table 2. Comparison Results of TCR of Six Image Sequences Obtained by Different Methods

	Seq1	Seq2	Seq3	Seq4
	TCR	TCR	TCR	TCR
(b)	0.315	0.090	0.087	1.000
(c)	0.332	0.078	0.085	1.000
(d)	0.338	0.084	0.096	1.000
(e)	0.407	0.093	0.159	1.000
(f)	0.431	0.085	0.141	1.000
(g)	0.910	1.000	1.000	1.000

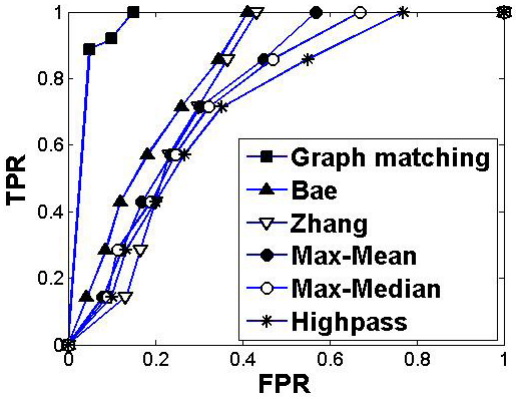

Fig. 5. ROC curve obtained by different methods using Seq1.

can efficiently detect the dim and small moving targets under complicated background.

This work was supported by the National Natural Science Foundation of China under Grant No. 61403412.

References

1. L. Yang, J. Yang, and K. Yang, Electron. Lett. **40,** 17 (2004).
2. G. D. Wang, C. Y. Chen, and X. B. Shen, Electron. Lett. **41,** 22 (2005).
3. S. D. Deshpande, M. H. Er, V. Ronda, and P. Chan, Proc. SPIE **3809,** 74 (1999).
4. F. Zhang, C. Li, and L. Shi, Infrared Phys. Technol. **46,** 323 (2005).
5. T. W. Bae, Infrared Phys. Technol. **54,** 403 (2011).
6. H. Deng, J. Liu, and Z. Chen, Chin. Opt. Lett. **8,** 010024 (2010).
7. X. Dong, X. Huang, Y. Zheng, L. Shen, and S. Bai, Infrared Phys. Technol. **62,** 36 (2014).
8. W. Hao, K. Zhang, and Z. Zhan, Chin. Opt. Lett. **10,** S21005 (2012).
9. R. Liu and Z. Jing, Chin. Opt. Lett. **10,** 021001 (2012).