

Article

Infrared Small Moving Target Detection via Saliency Histogram and Geometrical Invariability

Minjie Wan ¹, Kan Ren ^{1,2,*}, Guohua Gu ¹, Xiaomin Zhang ¹, Weixian Qian ¹, Qian Chen ¹ and Shuai Yu ³

¹ School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China; MinjieWan.1992@yahoo.com (M.W.); gghnjust@mail.njust.edu.cn (G.G.); zxm_1993@hotmail.com (X.Z.); qianweixian_njust@yahoo.com (W.Q.); chenq@njust.edu.cn (Q.C.)

² State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China

³ Xi'an Institute of Applied Optics, Xi'an 710065, China; yushuai111@yahoo.com

* Correspondence: k.ren@njust.edu.cn; Tel.: +86-258-431-0993

Academic Editors: Stefano Sfarra and Dario Ambrosini

Received: 20 April 2017; Accepted: 27 May 2017; Published: 1 June 2017

Abstract: In order to detect both bright and dark small moving targets effectively in infrared (IR) video sequences, a saliency histogram and geometrical invariability based method is presented in this paper. First, a saliency map that roughly highlights the salient regions of the original image is obtained by tuning its amplitude spectrum in the frequency domain. Then, a saliency histogram is constructed by means of averaging the accumulated saliency value of each gray level in the map, through which bins corresponding to bright target and dark target are assigned with large values in the histogram. Next, single-frame detection of candidate targets is accomplished by a binarized segmentation using an adaptive threshold, and their centroid coordinates with sub-pixel accuracy are calculated through a connected components labeling method as well as a gray-weighted criterion. Finally, considering the motion characteristics in consecutive frames, an inter-frame false alarm suppression method based on geometrical invariability is developed to improve the precision rate further. Quantitative analyses demonstrate the detecting precision of this proposed approach can be up to 97% and Receiver Operating Characteristic (ROC) curves further verify our method outperforms other state-of-the-arts methods in both detection rate and false alarm rate.

Keywords: IR small moving target; saliency map; saliency histogram; geometrical invariability

1. Introduction

IR imaging based small moving target detection is one of the most significant techniques for military, astronautics and aeronautics applications [1]. The performance of an infrared search and track (IRST) system highly relies on the precision of small target detection. A high-performance IR small target detection algorithm should remove the background clutters effectively and examine the real targets not only in a single frame, but also in consecutive frames. Moreover, motion trajectories need to be delineated, which makes it easier to monitor and capture the targets of interest. Although modern IR detectors possess the advantages of fast detection, cheap equipment and simple setup [2], the specific imaging mechanism and detection condition still result in the following inherent properties of IR images which cause much inconvenience for target detection [3–5]. On the one hand, IR imaging is based on IR radiation, so the target/background contrast may be weak if their radiometric quantities are similar. Furthermore, it is an inevitable fact that pixel size of the existing IR cameras cannot be small enough to generate high-resolution images, which means the images to be processed are always

blurred. In addition, the target size becomes comparatively small (fewer than 20×20 pixels) on account of the long military observation distance.

To efficiently examine small moving targets and remove various sorts of background clutters in IR images simultaneously, numerous algorithms have been developed so far, including filter based methods, mathematical morphology based methods, wavelet based methods, and so on. Filter based methods, the representatives of which are max-mean/max-median filter [6], high-pass filter [7] as well as two-dimensional least mean square (TDLMS) filter [8], utilize fixed templates to suppress clutters according to intensity difference. Although they can meet the requirement of real-time processing, the results are always inaccurate [9,10]. Mathematical morphology theory, like Hat transformation (including top-hat and bottom-hat transformation) [11], is another embranchment in the field of small target detection. These kinds of methods are aimed at enhancing regions of interest via morphological operations, but usually fails when the target is dim or the clutter is heavy [12,13]. Wavelet based algorithms [14] design a group of filters which are matched to a point spread function (PSF) at different scales by choosing a mother wavelet similar to PSF. Unfortunately, they are quite time-consuming and the false alarm rates are always high [15].

Although much progress has been achieved in the past decades, there are some significant problems that remain to be worthy of further investigation: on the one hand, dark target whose IR radiation is lower than the surroundings is seldom covered in previous algorithms; furthermore, using motion features to eliminate false alarms created in the single frame and forming complete trajectories are still tough tasks.

In this paper, we first present a new saliency histogram on the basis of a saliency map to distinguish visually salient regions from the background. Based on the fact that both bright and dark IR small targets have relatively different grayness with the background, proving IR small targets can be seen as salient regions in IR images, the gray levels correspond to targets would be assigned with large bin values in the saliency histogram. Then, an adaptive threshold is calculated via Otsu's method [16–18] to roughly extract IR targets according to the above-constructed histogram, and sub-pixel-accuracy centroid coordinates of all of the candidate target regions are obtained through a connected components labeling algorithm and an intensity-weighted criterion. For consecutive frames, we apply a uniformly accelerated motion model to make track correlations [19] and form completed motion trails for each candidate point. Then, the real small moving targets can be picked out from all the correlated points by use of the potential geometrical invariability existing in the sequences. Figure 1 gives an illustration of the framework of our method.

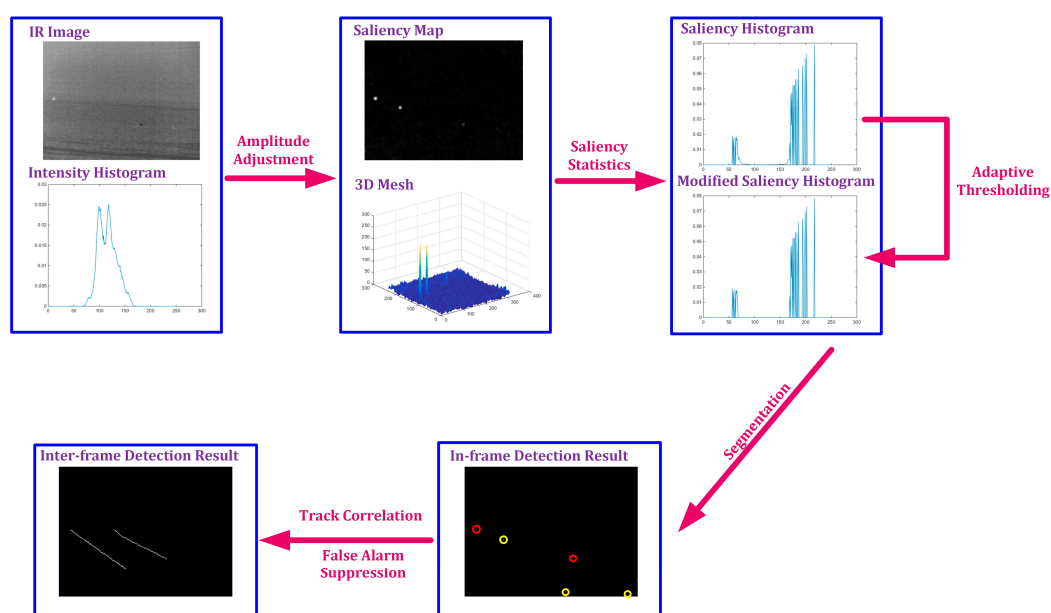


Figure 1. Flow chart of the presented method.

In conclusion, we argue that the main contribution of our work is to come up with an IR small moving target detection method that is suitable for both bright and dark targets and has a high detection accuracy under different conditions.

2. In-Frame Detection Based on Saliency Histogram

2.1. Saliency Map

Saliency detection is a popular and efficient tool that roughly presents the saliency distribution and locates the visually salient regions [20,21]. An IR small target, regardless of whether it is bright or dark, is able to attract much more attention from human eyes because of the distinct differences of brightness and shape when compared to its surrounding background. That is to say, IR small targets can be seen as salient regions [22].

In this section, a saliency map is constructed based on the amplitude spectrum in the frequency domain. Let us suppose that $I(x,y)$ is the input IR image and its Fourier frequency spectrum $F(u,v)$ is obtained by two-dimensional discrete Fourier transform (DFT) as Equation (1):

$$F(u,v) = \sum_{x=1}^W \sum_{y=1}^H I(x,y) \cdot \exp\left(-j \cdot 2\pi \left(\frac{ux}{W} + \frac{vy}{H}\right)\right) = \text{Re}(u,v) + j \cdot \text{Im}(u,v) \quad (1)$$

where W and H are the width and height of $I(x,y)$, respectively; $\text{Re}(u,v)$ and $\text{Im}(u,v)$ are the real part and imaginary part of $F(u,v)$, respectively; j is an imaginary unit. Then, we can extract the amplitude spectrum $A(u,v)$ and phase spectrum $P(u,v)$ as Equations (2) and (3):

$$A(u,v) = |F(u,v)| = \sqrt{\text{Re}^2(u,v) + \text{Im}^2(u,v)} \quad (2)$$

$$P(u,v) = \Theta(F(u,v)) = \text{arctg}\left(\frac{\text{Im}(u,v)}{\text{Re}(u,v)}\right) \quad (3)$$

where $|\cdot|$ and $\Theta|\cdot|$ denote calculating the norm and phase angle of an arbitrary complex number, and $\text{arctg}(\cdot)$ is an arctan function.

From the perspective of frequency domain, high-frequency components correspond to the salient target while low-frequency components represent the smooth background. However, the proportion of high-frequency components is much larger than that of low-frequency components, which is reflected by the fact that amplitude values of high frequencies are much smaller.

Based on the above-mentioned analysis, an exponential function is designed to make an reversal of the amplitude distribution, but the phase spectrum $P(u,v)$ remains unchanged. Equation (4) shows the way of amplitude adjustment:

$$A'(u,v) = \left(\frac{1}{\gamma}\right)^{A(u,v)}, \quad \gamma > 1 \quad (4)$$

where $A'(u,v)$ is the adjusted amplitude spectrum and γ is a fixed parameter (γ is set as 100 uniformly). As the illustration given by Figure 2, Equation (4) is a decreasing function, which ensures that a small $A(u,v)$ becomes larger while a large one turns to be smaller. That is to say, high frequencies (IR target) are enhanced in the spatial domain meanwhile low frequencies (background) are suppressed. Finally, the saliency map $S(x,y)$ can be calculated by inverse discrete Fourier transform (IDFT) and smoothed by Gaussian filtering as follows:

$$S(x,y) = \text{Ga}(x,y) * \left| F^{-1}(A'(u,v) \cdot \exp(j \cdot P(u,v))) \right|^2 \quad (5)$$

where $Ga(x,y)$ is a two-dimensional Gaussian filter; $*$ is a convolution operator; and $F^{-1}(\cdot)$ is an IDFT operator.

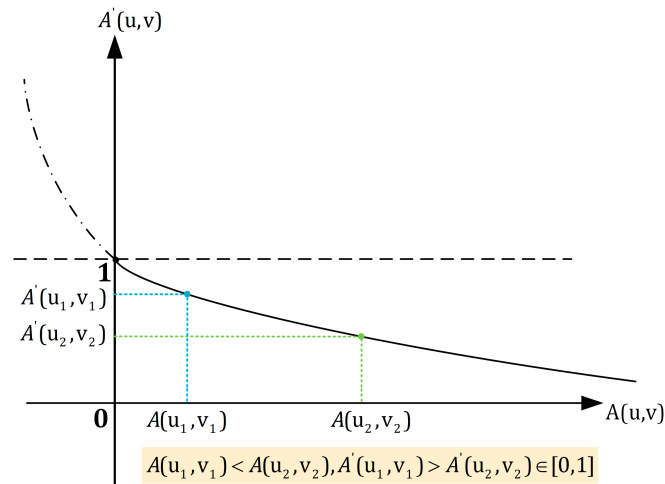


Figure 2. Schematic diagram of amplitude transformation.

2.2. Saliency Histogram

Considering the saliency map S and the input IR image I share the same size, it is obvious that there is a pixel-wise correspondence between S and I in the spatial domain. Based on this, a normalized saliency histogram χ is developed via averaging the accumulating saliency value of each gray level, which is illustrated by Equation (6):

$$\chi(k) = \begin{cases} \frac{S_a(k)/N(k)}{\text{sum}(\chi)} & \text{if } S_a(k) \neq 0 \\ 0 & \text{others} \end{cases} \quad (6)$$

where $\chi(k) \in [0,1]$ is the bin value for an arbitrary gray level k ($k = 1, 2, 3, \dots, L$ and L is the maximum gray level value) in the saliency histogram; $\text{sum}(\chi)$ means the sum of all the bin values which is used to make a normalization; $N(k) = \text{NUM}\{(x,y)|I(x,y) = k\}$ refers to the total number of the pixels whose gray levels are k in I , where $\text{NUM}\{\cdot\}$ means counting the number of elements in a set; $S_a(k)$ is defined as the cumulative sum of a set of saliency values in S , and

$$S_a(k) = \sum_x \sum_y S(x,y), (x,y) \in \{(x,y)|I(x,y) = k\} \quad (7)$$

Now, let us have a further discussion for the design of our saliency histogram. First of all, the ideal result of saliency histogram should meet the request that only those bins whose gray levels correspond to IR target are assigned with large saliency values. Enlightened by this idea, we calculate the saliency map S , which provides a quantitative measurement of saliency for each pixel and develop Equation (7) to measure the degree of saliency for each gray level. However, $S_a(k)$ is not a rigorous metric because there is a special case where the gray level k' in the background area has a low saliency value but quite a large quantity of pixels, and the $S_a(k')$ is still large. To address this problem, we take the number of pixels with different gray levels into consideration and use the averaging procedure to propose the final saliency histogram as Equation (6).

2.3. Adaptive Segmentation and Centroid Localization

The above-introduced saliency histogram χ reveals that the gray levels with large bin values correspond to the salient regions. For the purpose of extracting the target regions in a single frame as

correctly as possible, we need to remove those bins with remarkably small values through a simple threshold and generate a modified saliency histogram χ' as follows:

$$\chi'(k) = \begin{cases} \chi(k) & \text{if } \chi(k) > \chi_* \\ 0 & \text{others} \end{cases} \quad (8)$$

where χ_* is an adaptive threshold calculated by Otsu's method.

According to this modification, these removed bins stand for the smooth background and parts of the clutters. As a result, we can make a binarization for the input IR image on the basis of χ' directly, and a binarized image BI can be obtained from Equation (9):

$$BI(x, y) = \begin{cases} 0 & \text{if } \chi'(I(x, y)) = 0 \\ 1 & \text{others} \end{cases} \quad (9)$$

After binarization, it is highly possible that some of the extracted regions are incomplete due to the inhomogeneity of intensity. For $BI(x, y)$, close operation [23] is utilized to smooth the contours and fill in the holes existing in target regions as

$$BI_c = BI \blacksquare M = (BI \oplus M) \ominus M \quad (10)$$

where BI_c is the resulting image after close operation; M is a 5×5 sized structuring element; \blacksquare is a close operator; \oplus is a dilation operator, and \ominus is an erosion operator.

Intended to avoid forming repeated trajectories of the same target, every region only needs to be represented by a single centroid precisely. A typical fast connected components labeling algorithm [24] is applied to label all of the candidate regions. For all of the connected regions with the same label, we use a gray-weighted criterion to calculate its centroid position as follows:

$$P = \frac{\sum_{i=1}^N p_i \cdot I(p_i)}{\sum_{i=1}^N I(p_i)} \quad (11)$$

where P denotes the space coordinate of the centroid; p_i denotes the i -th pixel in each connected region; and N is the total pixel number of the region. By this procedure, every candidate target region is represented by an isolated point, which is called candidate point in the following contents.

3. Inter-Frame Detection Based on Geometrical Invariability

3.1. Track Correlation

Within a short time, the motion of a small moving target can be described with a uniformly accelerated motion model [25], which is applied to implement track correlation and form completed trajectories in our inter-frame detection.

Assume the space coordinates of all the candidate points in an arbitrary frame compose a set $P^\xi = \{P_1^\xi, P_2^\xi, P_3^\xi, \dots, P_n^\xi\}$, where ξ represents the frame number and n is the total number of candidate points in that frame. Here, we just take the m -th candidate point P_m^ξ as an example to explain the process of track correlation.

First, the velocity v_m^ξ and acceleration a_m^ξ of P_m^ξ are written as

$$v_m^\xi = (P_m^\xi - P_m^{\xi-1}) / \Delta t \quad (12)$$

$$a_m^\xi = (v_m^\xi - v_m^{\xi-1}) / \Delta t \quad (13)$$

where Δt means the time interval. For two adjacent frames, $\Delta t = 1$.

Next, based on the uniformly accelerated motion model, the estimated position $\widetilde{P}_m^{\xi+1}$ of P_m^ξ in the $(\xi + 1)$ -th frame is predicted as

$$\widetilde{P}_m^{\xi+1} = P_m^\xi + v_m^\xi \Delta t + \frac{1}{2} a_m^\xi \Delta t^2 + \Delta P^{\xi+1} \quad (14)$$

where $\Delta P^{\xi+1}$ denotes the displacement of background in the two adjacent frames. In this paper, we only consider the rotation and translation displacement caused by camera motion, and $\Delta P^{\xi+1}$ is calculated by an automatic registration method introduced in Ref. [26].

Furthermore, as is shown in Figure 3, a circular gate whose radius is \mathfrak{R} (we take $\mathfrak{R} = 5$ empirically) is set at $\widetilde{P}_m^{\xi+1}$. In the $(\xi + 1)$ -th frame, all of the candidate points (represented by the light green dots) located inside the gate make up of another set $Q_m = \{Q_1, Q_2, Q_3, \dots, Q_t\}$ where t is the quantity of candidate points inside the gate and the point Q_* is the maximum bin value in the saliency histogram is selected as the correlated one for P_m^ξ (if there is more than one point located inside the gate). Thus, a complete trajectory of P_m^ξ can be established if the above-discussed track correlation is repeated in every frame.

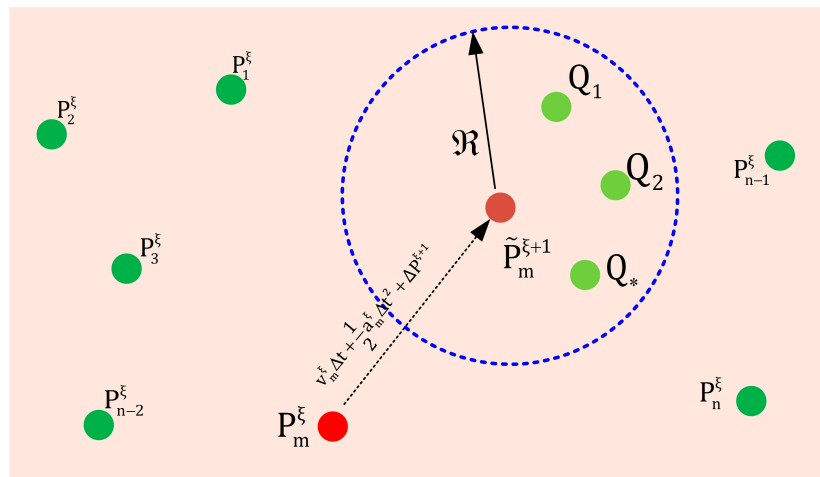


Figure 3. Schematic diagram of track correlation.

3.2. False Alarm Suppression

The procedure of track correlation presented above is implemented for each candidate point detected in the first frame and lasts for L frames. However, a formed trajectory would be eliminated on the condition that we cannot find the next correlated point for it in the following ε frames continuously. In this paper, we uniformly set $L = 20$ and $\varepsilon = 3$.

For each candidate point in the L -th frame, two maps can be constructed: points in the L' -th frame, where this candidate point is correlated successfully for the first time, being seen as dots, and the Euclidean distances between each other, being seen as edges, compose the first map; those of the L -th frame (these points should also exist in the L' -th frame) compose the second map.

Here, we demonstrate a fact that the distance between two background points remains constant regardless of the rotation and translation of background. An assumption is made that the rotation angle and the translation distance for the two maps defined above are $\Delta\Phi$ and $\Delta\rho$, and the rotation center is P_* . As a result, for the n -th point P_n^f in the first map and P_n^s in the second map, there is an equation existing obviously:

$$P_n^s - P_* = \aleph(\Delta\Phi) \cdot (P_n^f - P_* + \Delta\rho) \quad (15)$$

where $\aleph(\Delta\Phi)$ means the rotation matrix and

$$\mathbb{R}(\Delta\Phi) = \begin{pmatrix} \cos \Delta\Phi & -\sin \Delta\Phi \\ \sin \Delta\Phi & \cos \Delta\Phi \end{pmatrix} \quad (16)$$

Based on Equations (15) and (16), the Euclidean distance between two background points, P_m^s and P_n^s in the second map, can be expressed as

$$\|P_n^s - P_m^s\|_2 = \|\mathbb{R}(\Delta\Phi) \cdot (P_n^f - P_* + \Delta\rho) - \mathbb{R}(\Delta\Phi) \cdot (P_m^f - P_* + \Delta\rho)\|_2 \quad (17)$$

where $\|\cdot\|_2$ stands for 2-norm. Under the circumstance that P_m^s and P_n^s are two background points, there are no other displacements for them except for the translation of background $\Delta\rho$. It is absolutely true that rotation cannot change the length of a vector, so Equation (17) can be further written as

$$\|P_n^s - P_m^s\|_2 = \|\mathbb{R}(\Delta\Phi) \cdot (P_n^f - P_m^f)\|_2 = \|P_n^f - P_m^f\|_2 \quad (18)$$

In contrast, if P_m^s is a moving target point while P_n^s is a background point, P_m^s may have an extra displacement $\Delta\rho_m$ caused by self-motion, which means Equation (18) needs to be modified as

$$\begin{aligned} \|P_n^s - P_m^s\|_2 &= \|\mathbb{R}(\Delta\Phi) \cdot (P_n^f - P_* + \Delta\rho) - \mathbb{R}(\Delta\Phi) \cdot (P_m^f - P_* + \Delta\rho + \Delta\rho_m)\|_2 \\ &= \|\mathbb{R}(\Delta\Phi) \cdot (P_n^f - P_m^f - \Delta\rho_m)\|_2 \\ &= \|P_n^f - P_m^f + \Delta\rho_m\|_2 \\ &> \|P_n^f - P_m^f\|_2 \end{aligned} \quad (19)$$

Therefore, a conclusion can be summarized that the distance between two background points is constant, i.e., there is a geometrical invariability between background points. However, if there is at least a moving target point among them, the distance is changed. Motivated by this regulation, we propose a geometrical invariability based false alarm suppression method below. For the n -th candidate point in the second map, an index ψ denoting the difference of relative position is developed to judge whether P_n^s is a real moving target or a false alarm:

$$\psi(P_n^s) = \sum_{m=1}^N \left| \|P_n^s - P_m^s\|_2 - \|P_n^f - P_m^f\|_2 \right| \quad (20)$$

where N is the number of trajectories in the second map.

Lastly, the judge criterion is defined as follows:

$$P_n^s = \begin{cases} \text{target} & \text{if } \psi(P_n^s) > \text{thr} \\ \text{false alarm} & \text{others} \end{cases} \quad (21)$$

where thr is a threshold and it is set as $\text{thr} = \eta \cdot \max(\psi)$. We take $\eta = 0.7$ uniformly in our method.

4. Experiments and Analyses

We briefly introduce the test sequences used in our simulations, and then make comparisons of the results of both in-frame detection and inter-frame detection with several state-of-the-art algorithms to prove the robustness and precision of our algorithm under different natural backgrounds.

4.1. Introduction of Datasets

Four IR video sequences, captured by mid-wave infrared (MWIR) refrigerant imagers at a frame rate of 25 fps, are selected as our datasets for analysis. Table 1 shows the detailed information of these sequences and the corresponding first frames are displayed in Figure 4, where real moving targets that

need to be detected are marked with red circles, and the fake targets, as well as the regions that may possibly generate false alarms, are marked with yellow circles.

Table 1. Detailed information of our datasets.

Dataset	Background	Image Size	Frame Number	Number of Bright Targets	Number of Dark Targets
Seq.1	sky-cloud	512 × 512	100	192	0
Seq.2	ground	493 × 309	50	0	47
Seq.3	sea	313 × 236	150	150	145
Seq.4	cloud	320 × 256	200	192	0

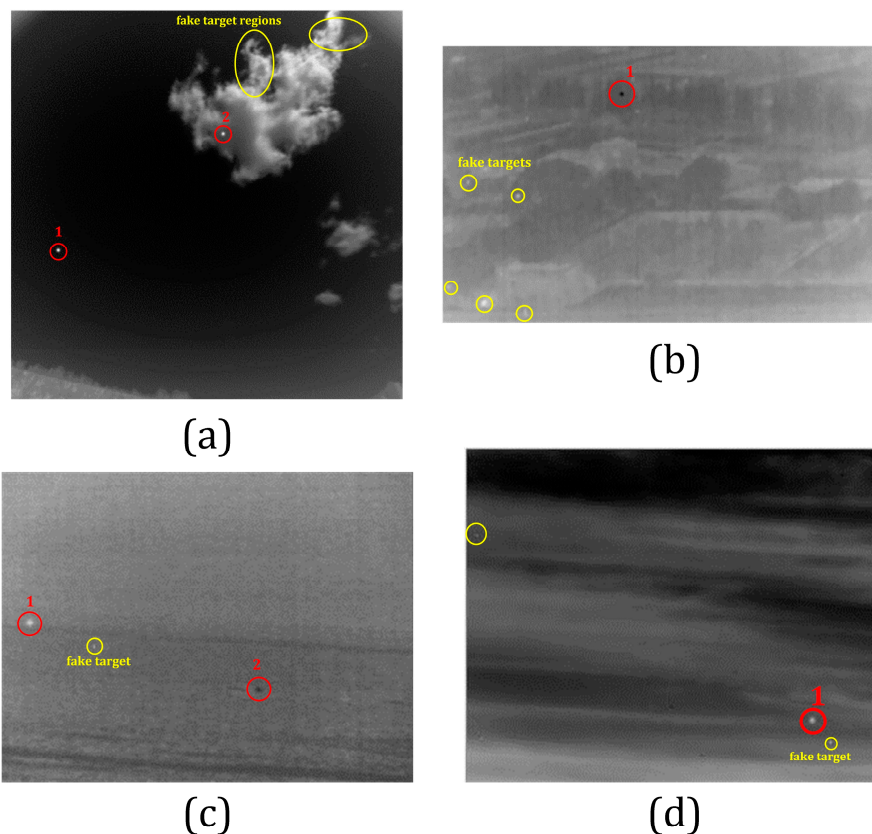


Figure 4. First frames of the tested infrared (IR) sequences: (a) Seq.1; (b) Seq.2; (c) Seq.3; (d) Seq.4.

The sky background of Seq.1 is clear, but the cloud regions are of great inhomogeneity of intensity, which means interference of cloud edge is the main source of false alarms. Among these four groups, Seq.2 and Seq.3 contain dark IR small targets, improving the difficulty and complexity of detection to some extent. Those bright but still points existing around the hills and houses in Seq.2 are fake targets, and it is quite hard to distinguish these kinds of points in a single frame. The sea waves are also very challenging for small target detection because the edges of waves are difficult to be suppressed by filtering based algorithms, and the motion of waves also cause much difficulty for the inter-frame detection. In addition, a dead point caused by the detector itself always exists during the whole sequence, which is a tough task for us to address. Seq.4 has a moving background and the cloud layer is quite dense. In addition, there are a certain number of dead pixels due to the poor quality of imaging.

4.2. Experimental Results of In-Frame Detection

First of all, this section is organized to show the processing results by our algorithm and the four other conventional algorithms. Then, several metrics are applied to make a quantitative comparison.

We choose the four groups of IR images shown in Figure 4 as the tested samples to perform the experiment. Meanwhile, four state-of-the-art algorithms: Max-Mean, Butterworth high-pass (BHP), Hat transformation (including Top-hat and Bottom-hat transformations) and two-dimensional least mean square (TDLMS) are selected to compare with our method. For Max-Mean, the raw IR image is filtered by a max-mean filter, and the filtered output is subtracted from the original image to enhance the IR small target. Furthermore, specific frequency components belonging to IR targets are extracted by setting a specific cut-off frequency of the Butterworth high-pass filter in the BHP method. Hat transformation denotes the pixel-wise difference between the raw image and the resulting image processed by morphological opening or closing operation. It should be noted that Top-hat transformation is designed for detecting bright targets while Bottom-hat transformation is designed for dark targets. As a result, Bottom-hat transformation is implemented in Seq.2 while Top-hat transformation is implemented in other sequences. Lastly, TDLMS detects small targets in the way of calculating the difference between the original image and the background estimation.

Figures 5–8 show the significant intermediate results and the output results of Seqs.(1–4) produced by our method, respectively. Note that real targets are marked with red circles and false alarms are marked with yellow ones. These figures demonstrate the following five points clearly:

- (1) While intensity histogram only discloses the pixel number of each gray level, the peaks of the proposed saliency histogram are only located in gray levels corresponding to salient regions, i.e., saliency histogram reveals the saliency distribution of each gray level;
- (2) The thresholds calculated via Otsu's method can suppress the bins with small bin values in the raw saliency histogram well and only few bins are preserved in the modified saliency histograms;
- (3) The saliency maps obtained by tuning amplitude spectra can roughly highlight the IR target no matter whether it is bright or dark, but the edges of sea waves and clouds with large gradient values are also preserved in the map;
- (4) Each candidate region in the binarized image is represented by a single centroid point in the final output, although few false alarms still exist;
- (5) The main sources of our false alarms are edges, dead pixels and still bright points.

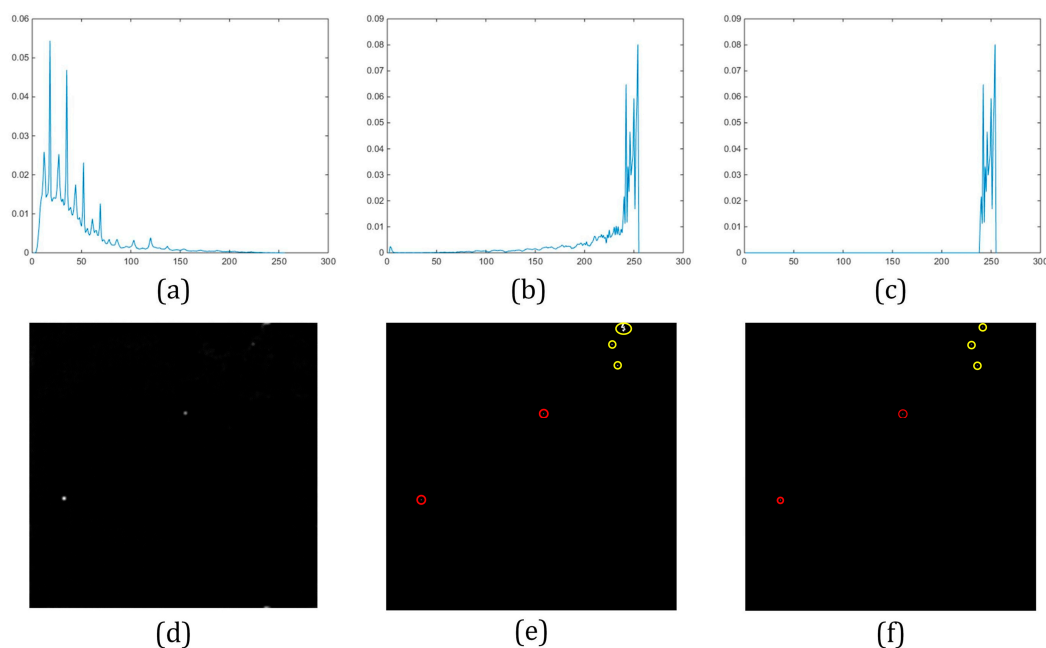


Figure 5. In-frame detection results of Seq.1 generated by our method: (a) intensity histogram; (b) original saliency histogram; (c) modified saliency histogram; (d) saliency map; (e) binarized image; (f) in-frame detection result.

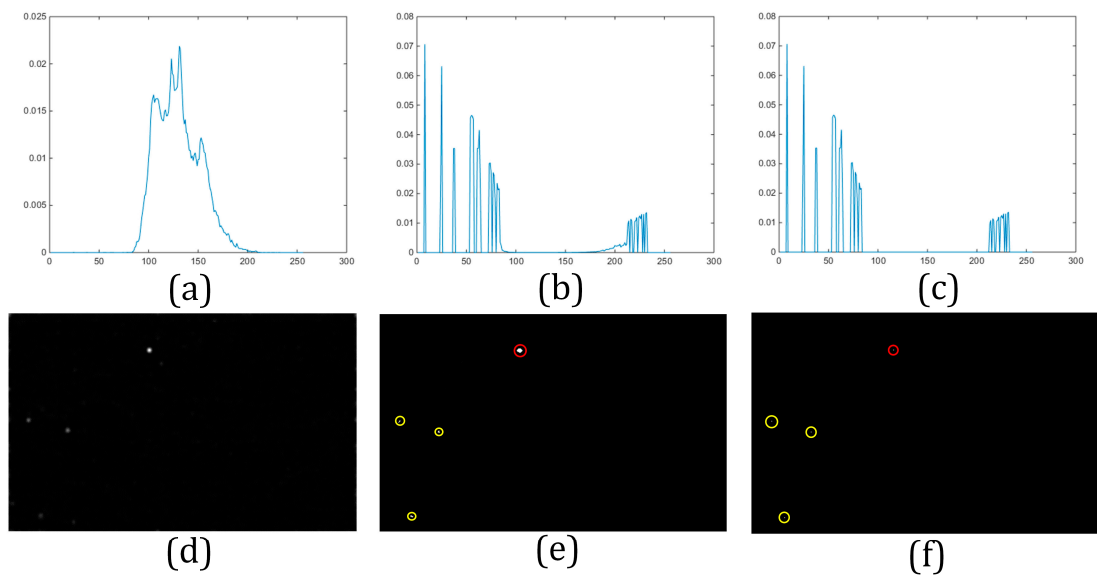


Figure 6. In-frame detection results of Seq.2 generated by our method: (a) intensity histogram; (b) original saliency histogram; (c) modified saliency histogram; (d) saliency map; (e) binarized image; (f) in-frame detection result.

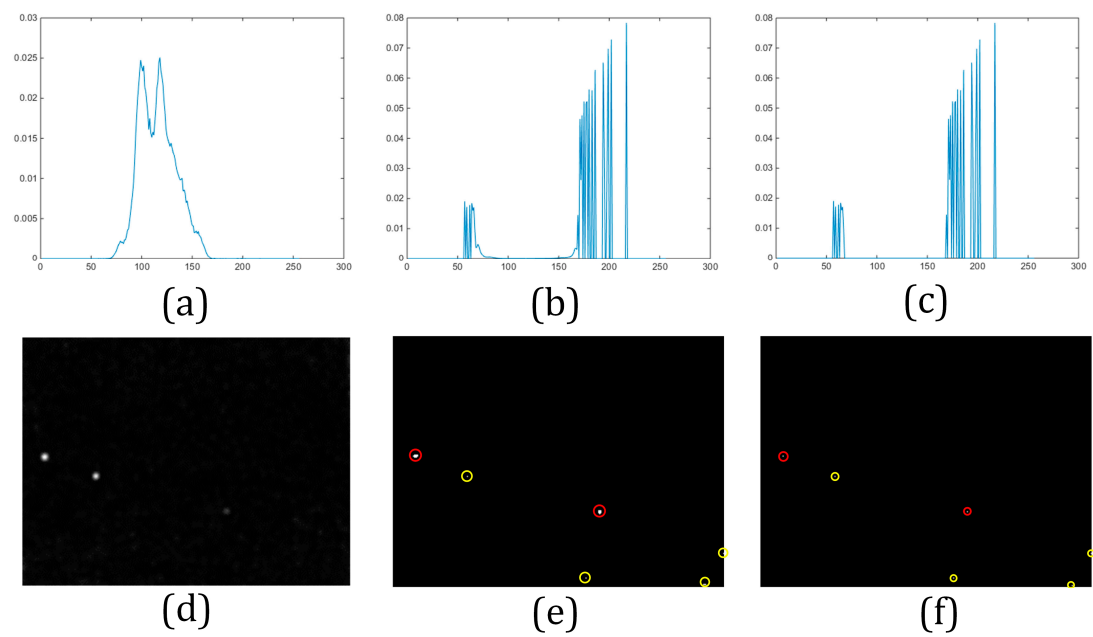


Figure 7. In-frame detection results of Seq.3 generated by our method: (a) intensity histogram; (b) original saliency histogram; (c) modified saliency histogram; (d) saliency map; (e) binarized image; (f) in-frame detection result.

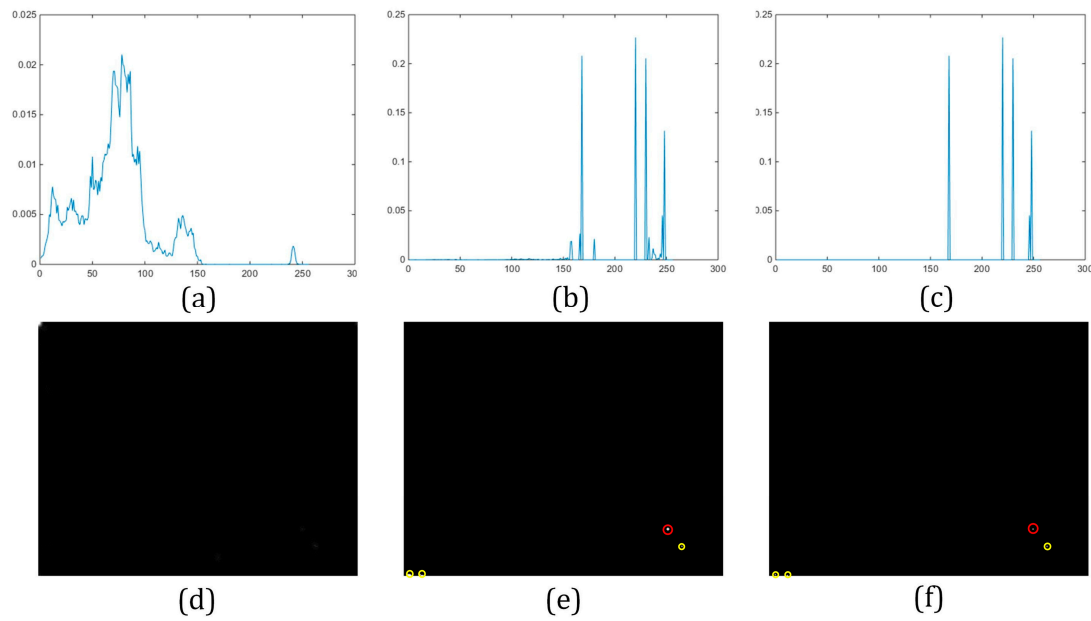


Figure 8. In-frame detection results of Seq.4 generated by our method: (a) intensity histogram; (b) original saliency histogram; (c) modified saliency histogram; (d) saliency map; (e) binarized image; (f) in-frame detection result.

Figure 9 presents the original processing results using the state-of-the-art algorithms, and Figure 10 further shows their binarized results via thresholds calculated by Otsu's method. For Max-Mean, it has poor performances when dealing with the cloud and sea backgrounds, indicating that it cannot remove the strong interferences caused by edges. BHP has the same drawbacks when addressing the edges, which can be seen from Figure 10b, because this algorithm is sensitive to frequency characteristics and large numbers of false alarms would appear, if the cutting-off frequency is chosen inappropriately. In addition, according to Figure 10c, an obvious phenomenon of Hat transformation is found that if there are both bright and dark targets in the image, it is inevitable that at least one kind of target would be lost because we cannot use Top-hat transformation and Bottom-hat transformation simultaneously. Finally, TDLMS has a stronger ability to suppress edges when compared to others, but it is easy to generate large quantities of candidate points for the same target region and this would cause repeated trajectories in inter-frame detections remarkably.

In order to discuss the performances more convincingly, two widely-accepted metrics called precision rate P and recall rate R [27] are selected to measure the detection results quantitatively. As is illustrated in Figure 11, an assumption is made that N_T is the pixel number of true targets existing in the current frame; N_D is the pixel number of targets detected by the tested algorithm, and $N_C = N_T \cap N_D$ is the pixel number of targets detected correctly. In this case, precision rate P and recall rate R can be thus defined as

$$P = \frac{N_C}{N_D} \times 100\% \quad (22)$$

$$R = \frac{N_C}{N_T} \times 100\% \quad (23)$$

Furthermore, a comprehensive evaluation index η representing the detecting precision of each algorithm [28] can be expressed as

$$\eta = \frac{(1 + \lambda^2) \cdot P \cdot R}{\lambda^2 \cdot P + R} \times 100\% \quad (24)$$

where λ is harmonic coefficient, and we set this coefficient as $\lambda = 1$ in this paper. $\eta = 1$ indicates that there are no false alarms and all of the real targets are discovered, while $\eta = 0$ means that none of the real targets are found out. Hence, the larger η is, the more satisfactory the result is. Table 2 reveals the statistical data of η for the in-frame detections at length. Clearly, the result presented by Table 2 matches the qualitative analyses made above. Our method achieves the largest η in all sequences, but the accuracies are not at a high level as a whole.

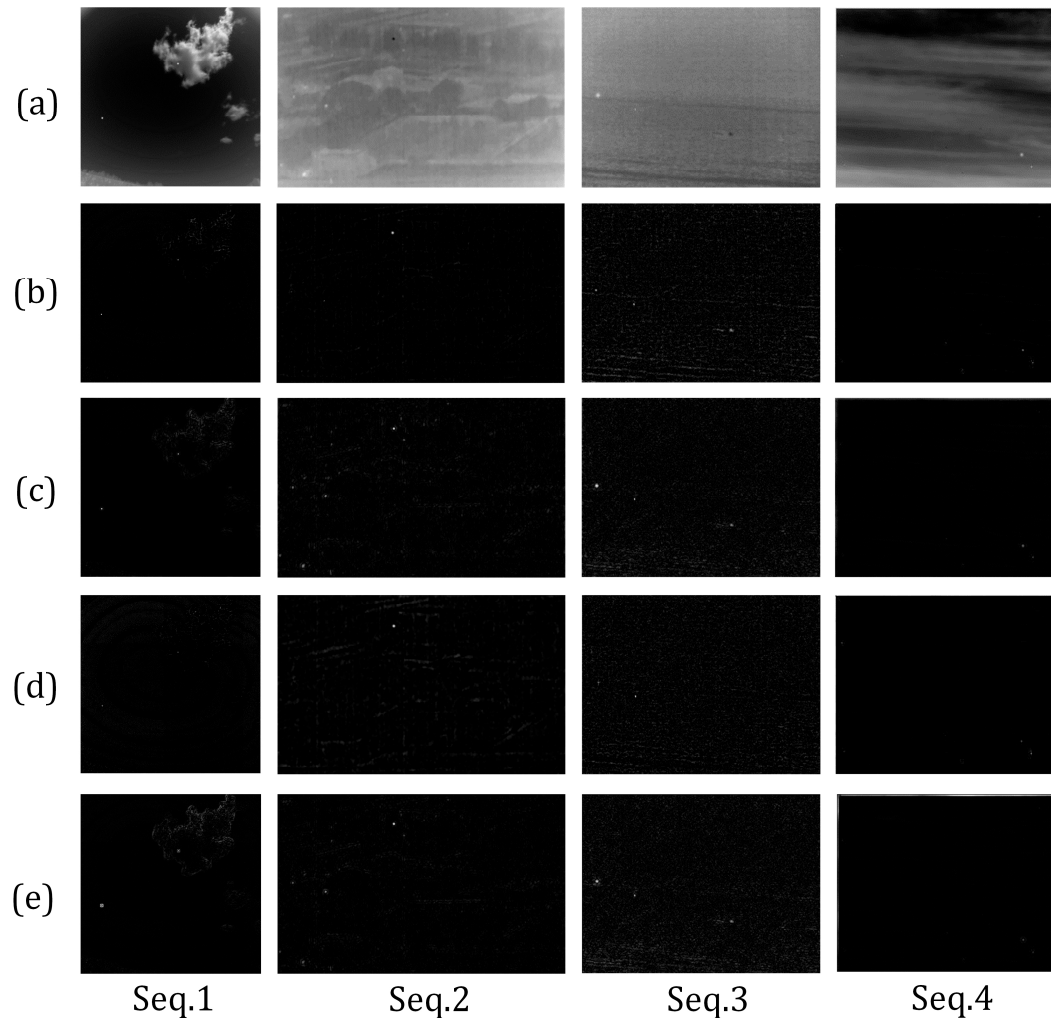


Figure 9. Original results of in-frame detection generated by four state-of-the-art algorithms: (a) Raw IR images; (b) Max-mean; (c) Butterworth high-pass (BHP) (d) Hat; (e) Two-dimensional least mean square (TDLMS).

Table 2. Statistical results of $\eta(\%)$ for the in-frame detection.

Methods	Seq.1	Seq.2	Seq.3	Seq.4
Max-Mean	1.05	5.00	12.90	2.23
BHP	4.24	3.39	9.52	10.00
Hat	13.33	0	26.67	0.31
TDLMS	4.44	3.08	14.29	0.30
Ours	57.14	40.00	50.00	40.00

BHP means Butterworth high-pass and TDLMS means two-dimensional least mean square.

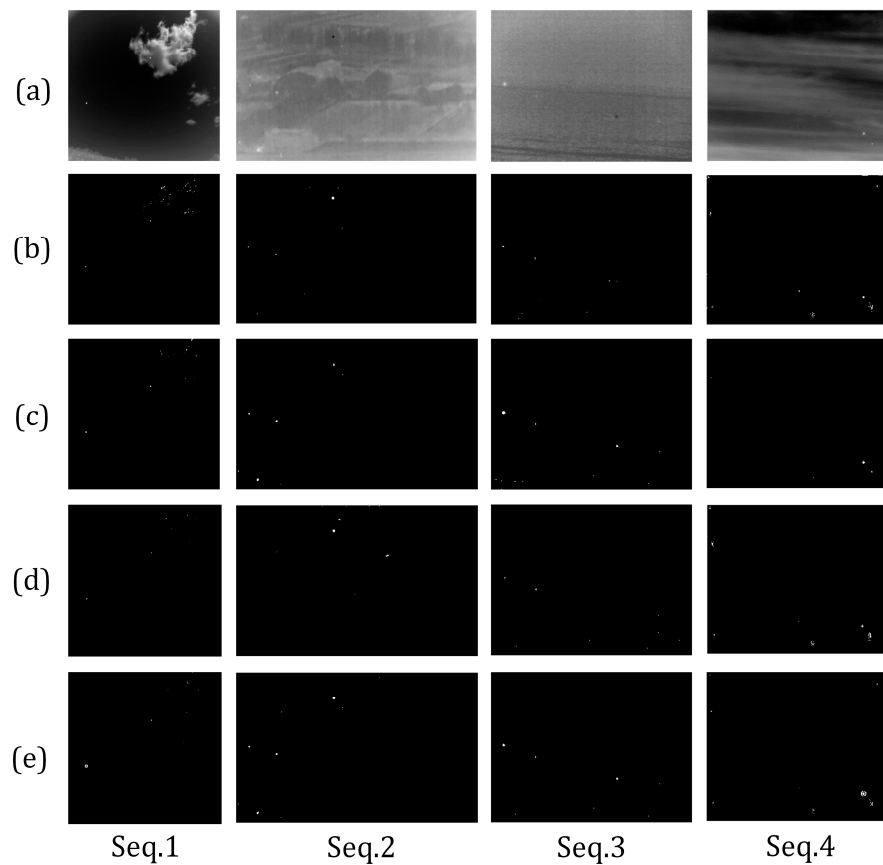


Figure 10. Binarized results of in-frame detection via Otsu's method: (a) Raw IR images; (b) Max-Mean; (c) BHP; (d) Hat; (e) TDLMS.

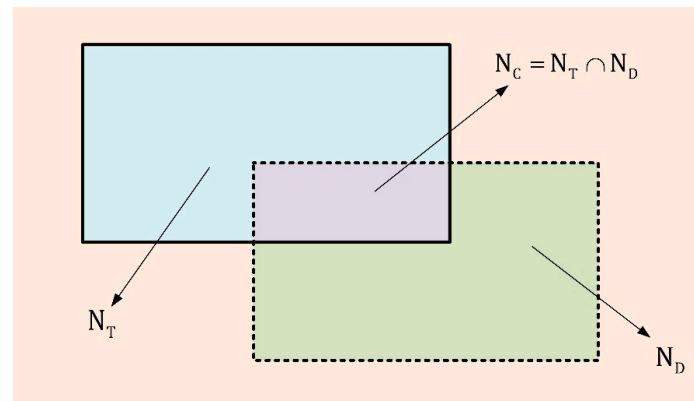


Figure 11. Schematic diagram of the parameters used for the measurement of detection accuracy.

4.3. Results of Inter-Frame Detection

In this part, the binarized detection results of the consecutive frames are accumulated in the final frame. Figure 12 presents the cumulative results for each sequence in detail. Overall, our algorithm has the best performances in all sequences and nearly all of the false alarms remaining in the single frame detection are removed by the subsequent inter-frame detection effectively. Moreover, complete trajectories are drawn in the final resulting images, and the phenomenon of repeated trajectories for the same target is successfully avoided. By comparison, other methods have less satisfactory performances in this experiment. As is shown in the resulting images of Seq.1, Seq.3 and Seq.4, Max-Mean is easy to

produce discontinuous trajectories, meaning that real targets are lost in certain frames. The trajectories generated by BHP are relatively complete, but regions with clutter edges contain large numbers of false alarms, which is especially outstanding in Seq.1 and Seq.3. Hat transformation suffers from missing detection seriously, and the trail of the dark IR small target is completely missed in Seq.3. However, this method can reduce the quantity of repeated trails to some extent when compared with the other three algorithms. Furthermore, it has a good anti-inference ability to clutter edges. In contrast, TDLMS tends to severely suffer from the repeated trails, and it is obvious that the trajectories produced by this method are thicker than others.

Table 3 presents the η values of the five groups in detail. From the data provided by this table, we can clearly find that η values of our method are higher than 97% and are at least two times larger than other methods in all of the four sequences. Through further analyses for the source of false alarms produced by contrastive algorithms, we argue that the major sources contain two aspects: (1) the repeated trajectories increase the quantities of redundant and useless points to a great extent; (2) the false detection results of edges have been accumulated.

The ROC (Receiving Operating Characteristic) curve [29] is an effective tool to describe the quality of detection methods. For an ROC curve, the abscissa and ordinate stand for the probability of false alarm rate (P_f) and detection rate (P_d), which are expressed as Equations (25) and (26), respectively:

$$P_f = \frac{N_F}{N_0} \quad (25)$$

$$P_d = \frac{N_C}{N_D} \quad (26)$$

where N_0 denotes the total pixel number of the current frame. A good ROC means that the tested method is able to highlight the target and suppress the clutters at the same time. The ROC curves of the four experiments are drawn as Figure 13.

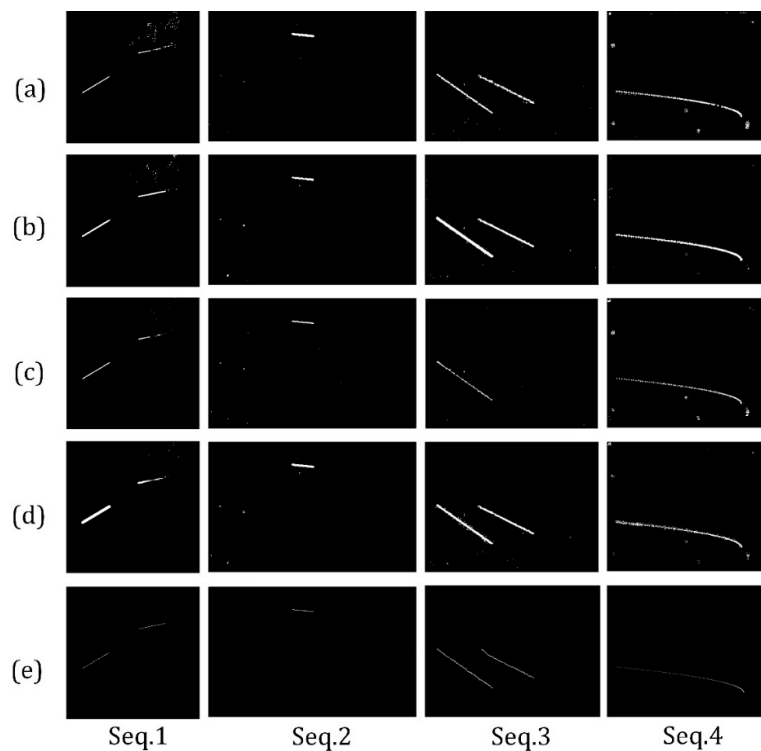


Figure 12. Cumulative results of inter-frame detection.

In light of the four groups of ROC curves, it is apparent that the area under the ROC curve belonging to our method is always far larger than the contrastive methods. However, the performance of each contrastive method varies a lot under different backgrounds, further demonstrating that the robustness of the four methods is weaker than our method.

Table 3. Statistical results of $\eta(\%)$ for the inter-frame detection.

Methods	Seq.1	Seq.2	Seq.3	Seq.4
Max-Mean	18.49	28.99	35.54	17.33
BHP	19.26	29.59	30.60	20.09
Hat	29.37	26.11	16.12	17.94
TDLMS	27.04	27.10	34.66	28.12
Ours	97.32	98.62	98.04	97.47

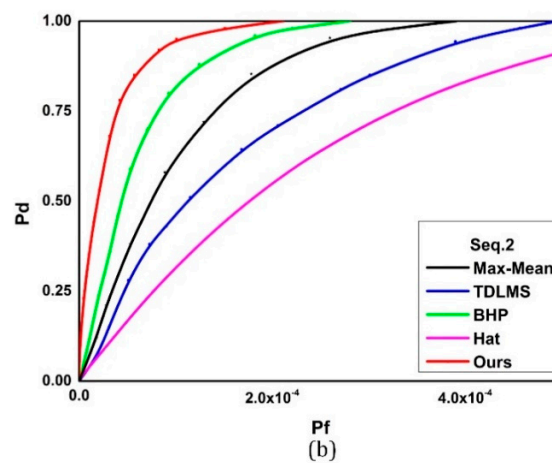
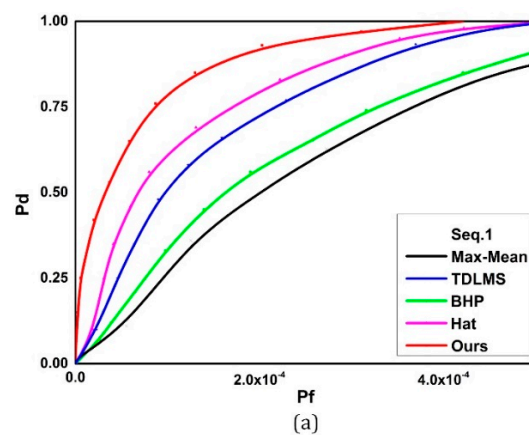


Figure 13. Cont.

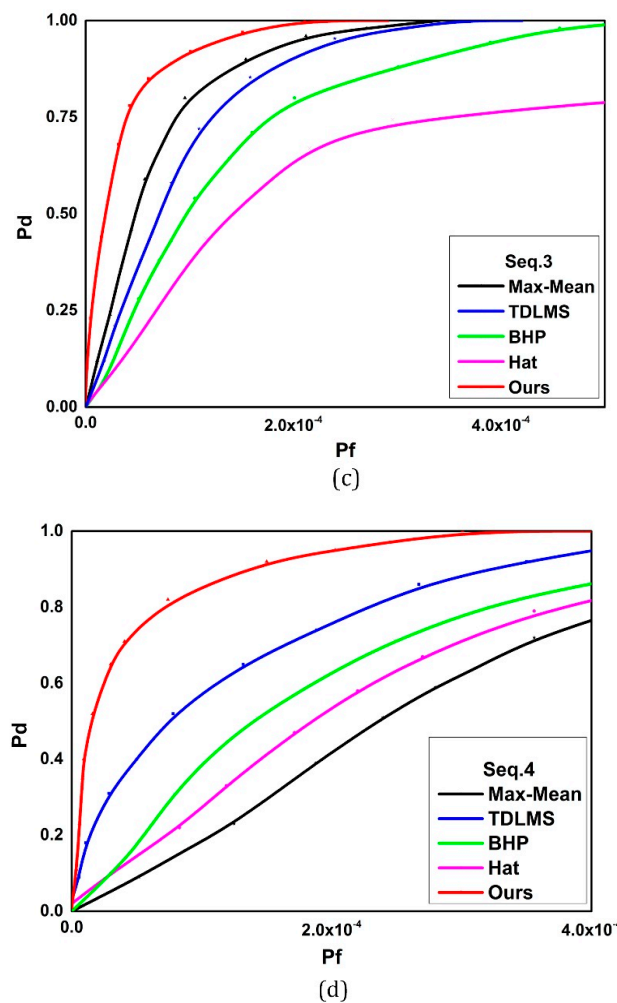


Figure 13. Receiving Operating Characteristic (ROC) curves of the four tested sequences: (a) Seq.1, (b) Seq.2, (c) Seq.3, (d) Seq.4.

5. Conclusions

It is easy for conventional IR small target detection methods to generate large quantities of false alarms due to the small size of target, the lack of color or texture information and the interference of clutters. Furthermore, the existing algorithms scarcely have the ability to detect both bright and dark IR small targets accurately at the same time, and the inter-frame motion information is also ignored by most of the researchers.

In this paper, an IR small moving target detection method using saliency histogram and geometrical invariability is proposed. For the in-frame detection part, a saliency histogram is established by averaging the cumulative saliency value of each gray level so that a single-frame segmentation can be made via an adaptive threshold of the histogram, and the centroid position of candidate targets is calculated via a connected components labeling algorithm and a gray-weighted criterion. For the inter-frame detection part, false alarms are further removed according to the geometrical invariability existing between two relatively still points. Large numbers of experiments convincingly prove that our method has robustness and satisfactory precision under various natural backgrounds compared with other state-of-the-art methods.

In our future work, we plan to concentrate on investigating more well-performed features of the IR small target in single-frame detection so as to reduce the computing quantity of the inter-frame detection and further improve the final detection precision.

Acknowledgments: This paper is financially supported by the National Natural Science Foundation of China (61675099) and the Open Foundation of State Key Laboratory of Networking and Switching Technology (Beijing University of Posts and Telecommunications) (SKLNST-2016-2-07).

Author Contributions: Xiaomin Zhang conceived and designed the experiments; Guohua Gu performed the experiments; Weixian Qian analyzed the data; Qian Chen and Shuai Yu contributed reagents/materials/analysis tools; and Minjie Wan and Kan Ren wrote the paper.

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

References

1. Bai, X.; Zhou, F. Infrared small target enhancement and detection based on modified top-hat transformations. *Comput. Electr. Eng.* **2010**, *36*, 1193–1201. [[CrossRef](#)]
2. Zhang, H.; Fernandes, H.; Dizeu, F.B.D.; Hassler, U.; Fleuret, J.; Genest, M.; Castanedo, C.I.; Robitaille, F.; Joncas, S.; Maldague, X. Pulsed micro-laser line thermography on submillimeter porosity in carbon fiber reinforced polymer composites: experimental and numerical analyses for the capability of detection. *Appl. Opt.* **2016**, *55*, D1–D10. [[CrossRef](#)] [[PubMed](#)]
3. Wan, M.; Gu, G.; Qian, W.; Ren, K.; Chen, Q. Hybrid active contour model based on edge gradients and regional multi-features for infrared image segmentation. *Opt. Int. J. Light Electron Opt.* **2017**, *140*, 833–842. [[CrossRef](#)]
4. Kim, S.; Lee, J. Scale invariant small target detection by optimizing signal-to-clutter ratio in heterogeneous background for infrared search and track. *Pattern Recognit.* **2012**, *45*, 393–406. [[CrossRef](#)]
5. Zhang, F.; Li, C.; Shi, L. Detecting and tracking dim moving point target in IR image sequence. *Infrared Phys. Technol.* **2005**, *46*, 323–328. [[CrossRef](#)]
6. Deshpande, S.D.; Er, M.H.; Ronda, V.; Chan, P. Max-mean and max-median filters for detection of small targets. In Proceedings of the SPIE's International Symposium on Optical Science, Engineering, and Instrumentation, International Society for Optics and Photonics, Denver, CO, USA, 4 October 1999; pp. 74–83.
7. Yang, L.; Yang, J.; Yang, K. Adaptive detection for infrared small target under sea-sky complex background. *Electron. Lett.* **2004**, *40*, 1083–1085. [[CrossRef](#)]
8. Bae, T.W.; Kim, Y.C.; Ahn, S.H.; Sohng, K.I. An efficient two dimensional least mean square (TDLMS) based on block statistics for small target detection. *J. Infrared Millim. Terahertz Waves* **2009**, *30*, 1092–1101. [[CrossRef](#)]
9. Wan, M.; Gu, G.; Cao, E.; Hu, X.; Qian, W.; Ren, K. In-frame and inter-frame information based infrared moving small target detection under complex cloud backgrounds. *Infrared Phys. Technol.* **2016**, *76*, 455–467. [[CrossRef](#)]
10. Sun, X.; Hou, W.; Yu, Q.; Liu, X.; Shang, Y. Small infrared target detection using frequency-spatial cues in a single image. *J. Electron. Imaging* **2014**, *23*, 043003. [[CrossRef](#)]
11. Zeng, M.; Li, J.; Peng, Z. The design of top-hat morphological filter and application to infrared target detection. *Infrared Phys. Technol.* **2006**, *48*, 67–76. [[CrossRef](#)]
12. Wang, W.; Li, C.; Shi, J. A robust infrared dim target detection method based on template filtering and saliency extraction. *Infrared Phys. Technol.* **2015**, *73*, 19–28. [[CrossRef](#)]
13. Yang, C.; Ma, J.; Qi, S.; Tian, J.; Zheng, S.; Tian, X. Directional support value of Gaussian transformation for infrared small target detection. *Appl. Opt.* **2015**, *54*, 2255–2265. [[CrossRef](#)] [[PubMed](#)]
14. Ye, Z.; Ruan, Y.; Wang, J.; Zou, Y. Detection algorithm of weak infrared point targets under complicated background of sea and sky. *J. Infrared Millim. Terahertz Waves* **2000**, *19*, 121–124.
15. Bae, T. Spatial and temporal bilateral filter for infrared small target enhancement. *Infrared Phys. Technol.* **2014**, *63*, 42–53. [[CrossRef](#)]
16. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [[CrossRef](#)]
17. Zuo, C.; Chen, Q.; Sui, X. Range limited bi-histogram equalization for image contrast enhancement. *Opt. Int. J. Light Electron Opt.* **2013**, *124*, 425–431. [[CrossRef](#)]
18. Xu, H.; Chen, Q.; Zuo, C.; Yang, C.; Liu, N. Range limited double-thresholds multi-histogram equalization for image contrast enhancement. *Opt. Rev.* **2015**, *22*, 246–255. [[CrossRef](#)]
19. Shi, Z.; Fernando, W.A.C.; Kondoz, A. Adaptive direction search algorithms based on motion correlation for block motion estimation. *IEEE Trans. Consum. Electron.* **2011**, *57*, 1354–1361. [[CrossRef](#)]

20. Fuchida, M.; Pathmakumar, T.; Mohan, R.E.; Tan, N. Vision-Based Perception and Classification of Mosquitoes Using Support Vector Machine. *Appl. Sci.* **2017**, *7*, 51. [[CrossRef](#)]
21. Zhao, J.; Gao, X.; Chen, Y.; Feng, H.; Wang, D. Multi-window visual saliency extraction for fusion of visible and infrared images. *Infrared Phys. Technol.* **2016**, *76*, 295–302. [[CrossRef](#)]
22. Ma, J.; Zhou, Z.; Wang, B.; Zong, H. Infrared and visible image fusion based on visual saliency map and weighted least square optimization. *Infrared Phys. Technol.* **2017**, *82*, 8–17. [[CrossRef](#)]
23. Gonzalez, R.S.; Wintz, P. Image enhancement. In *Digital Image Processing*; Addison Wesley Publishing Co., Inc.: Reading, MA, USA, 1977.
24. He, L.; Chao, Y.; Suzuki, K.; Wu, K. Fast connected-component labeling. *Pattern Recognit.* **2009**, *42*, 1977–1987. [[CrossRef](#)]
25. Dong, X.; Huang, X.; Zheng, Y.; Shen, L.; Bai, S. Infrared dim and small target detecting and tracking method inspired by human visual system. *Infrared Phys. Technol.* **2014**, *62*, 100–109. [[CrossRef](#)]
26. Gong, M.; Zhao, S.; Jiao, L.; Tian, D.; Wang, S. A novel coarse-to-fine scheme for automatic image registration based on SIFT and mutual information. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 4328–4338. [[CrossRef](#)]
27. Alazrai, R.; Momani, M.; Daoud, M.I. Fall Detection for Elderly from Partially Observed Depth-Map Video Sequences Based on View-Invariant Human Activity Representation. *Appl. Sci.* **2017**, *7*, 316. [[CrossRef](#)]
28. Zhang, T.; Han, J.; Zhang, Y.; Bai, L. An adaptive multi-feature segmentation model for infrared image. *Opt. Rev.* **2016**, *23*, 220–230. [[CrossRef](#)]
29. Qi, S.; Ming, D.; Ma, J.; Sun, X.; Tian, J. Robust method for infrared small-target detection based on Boolean map visual theory. *Appl. Opt.* **2014**, *53*, 3929–3940. [[CrossRef](#)] [[PubMed](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).