

Article

Local Spatial–Temporal Matching Method for Space-Based Infrared Aerial Target Detection

Lue Chen ^{1,2,3} , Peng Rao ^{1,2,*}, Xin Chen ^{1,2} and Maotong Huang ^{1,2,*}¹ Key Laboratory of Intelligent Infrared Perception, Chinese Academy of Sciences, Shanghai 200083, China; chenlue@mail.sitp.ac.cn (L.C.); chenxin@mail.sitp.ac.cn (X.C.)² Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China³ University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: peng_rao@mail.sitp.ac.cn (P.R.); mt_huang@mail.sitp.ac.cn (M.H.)

Abstract: The feature of a space-based infrared signal is that the intensity of clutter is much stronger than that of an aerial target. Such a feature poses a great challenge to aerial target detection since the existing infrared target detection methods are prone to enhance clutter but ignore the real target, which results in missed detection and false alarms. To tackle the challenge, we propose a concise method based on local spatial–temporal matching (LSM). Specifically, LSM mainly consists of local normalization, local direction matching, spatial–temporal joint model, and inverse matching. Local normalization aims to enhance the target to the same strength as the clutter, so that the weak target will not be ignored. After normalization, a direction-matching step is applied to estimate the moving direction of the background between the basic frame and referenced frame. Then the spatial–temporal joint model is constructed to enhance the target and suppress strong clutter. Similarly, inverse matching is conducted to further enhance the target. Finally, a salience map is obtained, on which the aerial target is extracted by the adaptive threshold segmentation. Experiments conducted on four space-based infrared datasets indicate that LSM handles the above challenge and outperforms seven state-of-the-art methods in space-based infrared aerial target detection.



Citation: Chen, L.; Rao, P.; Chen, X.; Huang, M. Local Spatial–Temporal Matching Method for Space-Based Infrared Aerial Target Detection. *Sensors* **2022**, *22*, 1707. <https://doi.org/10.3390/s22051707>

Academic Editor: Saeed Moradi

Received: 13 January 2022

Accepted: 19 February 2022

Published: 22 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The task of aerial target detection is of great importance in many fields, including air traffic surveillance [1] and intelligence reconnaissance [2]. Space-based infrared (IR) imaging technology has the advantages of all-day and wide-area imaging, while both the on-orbit experiment [3] and ground theoretical research [4] have certified that an aerial target can be detected by space-based IR detectors. Therefore, research on space-based IR aerial target detection continues to attract much attention.

The sizes of the aerial targets on the space-based IR images range from 5×5 to 9×9 pixels, which is in accordance with the definition [5] of IR small target. However, space-based IR aerial target detection is considerably different from ground-based IR small target detection. First, a remote imaging distance (>300 km) weakens the intensity of aerial targets. Second, the complex earth background and frequent human activities generate strong clutter whose spatial characters are similar to the small target in space-based images. The above two factors result in a feature in which the aerial target is much weaker than the clutter in the space-based IR image. As shown in Figure 1, both aerial target and clutter cover several pixels in the space-based IR image, while the intensity of clutter is nearly twenty times stronger than that of the target. During detection, the weak target may lead to missed detection, and the strong clutter could yield false alarms. Third, the computational resources are limited on the space-based platform, while they are unlimited

on the ground-based platform, which means the space-based detection method must be resource-friendly. Therefore, how to enhance weak targets and suppress strong clutter efficiently has been a critical issue for the space-based detection method, and it has been the challenge for space-based IR aerial target detection.

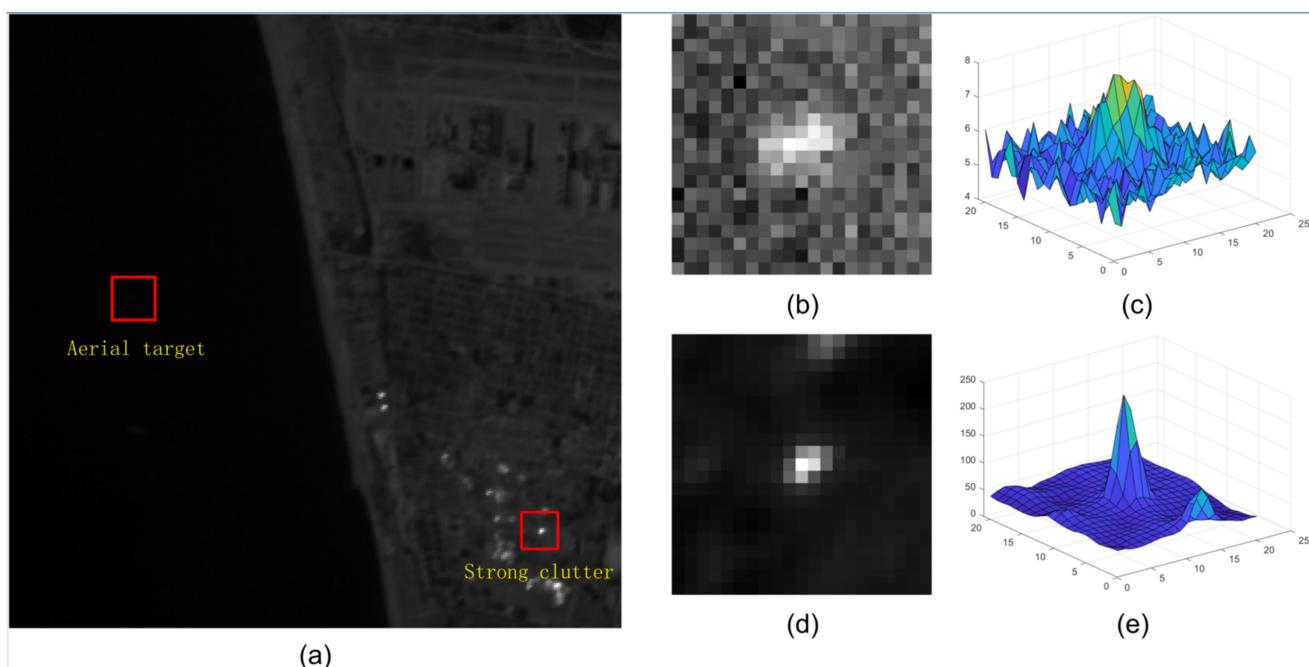


Figure 1. (a) The aerial target and strong clutter in the space-based IR image; their details are magnified in (b–e). (b) The local slice of an aerial target. (c) Three-view of the aerial target. (d) Local slice of strong clutter. (e) Three-view of strong clutter.

In the past decades, researchers continually proposed hundreds of IR small target detection methods, aiming for efficient detection under different scenarios. Nevertheless, most methods are proposed for ground-based detection instead of space-based detection.

The filter-based methods, such as TDLMS [6] and Top-hat [7], are easy to achieve, but they still struggle to enhance weak the target. Recently, some researchers have designed more complicated filters to detect small targets under specific background; for example, Lu et al. proposed a filter-based method for maritime IR small target detection [8]. The local contrast method (LCM) proposed by Chen et al. [9] attracts much attention for the concise structure. Additionally, a great number of LCM-based methods [10–12] working on the ground-based platform have subsequently been proposed and detect small targets under complex backgrounds. Moreover, most space-based detection methods, such as local blob-like contrast map and local gradient map (LBCM-LGM) [13], neighborhood saliency map (NSM) [14], spatial-temporal local contrast method (STLCM) [2], and spatial-temporal local contrast filter (STLCF) [15], are LCM based. Though these methods perform well on weak target enhancement, they suppress the strong clutter inefficiently, leading to false alarms. In recent years, the mainstream detection methods are mostly based on IR image patch (IPI), low-rank representation (LLR), and deep learning (DL). According to the different spatial correlations of target and background, some methods based on IPI [16–19] or LLR [20–22] have been proposed to extract the small target from the IR image. However, these methods cannot distinguish the real target from the background when the target has close intensity to its neighboring region. In addition, they show poor real-time performance since they require a great number of iterations during optimization. Other mainstream methods based on DL [23–26] show effective performance in complex backgrounds. However, the performance of DL-based methods severely relies on their

datasets; thus, they are not suitable for space-based detection because the space-based dataset is scarce in its current state.

As far as we know, although thousands of infrared small target detection methods have been proposed, most are ground-based detection methods; the number of space-based detection methods is much fewer than ground-based methods. Most existing space-based detection methods are LCM-based since computational resources on space platforms are limited while LCM-based methods are easy to implement on hardware and consume fewer computational resources than IPI- or DL-based methods. In 2018, a single-frame method called neighboring saliency map (NSM) for space-based detection was proposed and detected a dim target with a signal-to-clutter (*SCR*) less than 1. The space-based detection methods based on spatial–temporal local contrast maps (STLCF) [15] and spatial–temporal local contrast maps (STLCM) [2] are both LCM-based methods. Lv et al. proposed a method that detects the space-based weak moving target with an $SCR \approx 1$ or even <1 [27]. They further proposed a dim small moving target detection and tracking method based on a spatial–temporal joint processing model (STJP) [28], which also performed well on space-based dim target detection. However, the existing space-based detection methods mainly focus on dim target enhancement but ignore interference resulting from the strong clutter.

Although current methods, both space-based and ground-based methods, achieve detection in complex backgrounds, they are only for conditions where the target strength is close to or stronger than the clutter or highly light background. Therefore, it is significant to overcome the space-based detection challenge posed by a feature in which the clutter is much stronger than the aerial target.

To conquer the above challenge, we propose a space-based IR aerial target detection method based on local spatial–temporal matching (LSM), which has a concise structure. The contributions of LSM are given as follows.

- (1) Local normalization is proposed to shorten the difference between aerial target and strong clutter, which ensures that the weak target and strong clutter will be processed in subsequent steps within the same value domain.
- (2) Local direction matching and spatial–temporal joint model are constructed to suppress the strong clutter and enhance aerial target by considering the spatial–temporal difference between aerial target and background.
- (3) A reverse matching step is leveraged to further enhance the target and eliminate the residual clutter.
- (4) Experiments conducted on the space-based IR datasets demonstrate that LSM can enhance the weak target and suppress the strong clutter simultaneously and effectively and that it performs better than the existing methods on space-based IR aerial target detection.

2. Proposed Methods

The local spatial–temporal matching detection method (LSM) is suitable for the IR image sequence obtained by a space-based platform under staring imaging mode. LSM consists of five steps: local normalization, local direction matching, spatial–temporal joint model, reverse matching, and adaptive threshold segmentation. The details of LSM are elaborated in this section, and an overview is given in Figure 2.

2.1. Local Slices Extraction and Normalization

The first step in the proposed methods is local normalization, which reduces the sensitivity of the subsequent steps to the strong clutter. As shown in Figure 2, at the local normalization step, a local slice named \mathbf{R}_{11} is extracted at the point (x, y) in the base frame \mathbf{I}_b (b represents the frame number in the sequence). The neighboring region of \mathbf{R}_{11} is defined in Equation (1):

$$\Omega_{\mathbf{R}_{11}} = \{(i, j) | \max(|i - x|, |j - y|) \leq 3s + 4\}, s = 1, 2, 3, 4, \quad (1)$$

where $\Omega_{R_{11}}$ is the neighboring region, and s is the radius of the target. In reference frame I_{b+l} , the local region of the pixel $I_{b+l}(x, y)$ represented by Ω_{local} is extracted and defined as:

$$\Omega_{\text{local}} = \{(p, q) | \max(|p - x|, |q - y|) \leq r_{\text{match}}\}, \quad (2)$$

where l represents frame interval, and r_{match} represents the matching radius determined by practical engineering tasks. In our work, r_{match} is set to 1; the range of is illustrated in Figures 2 and 3. Then nine slices with the same dimension as R_{11} are extracted and named $R_{2m}, m = 1, 2, 3, \dots, 9$. The positions of R_{2m} s are further illustrated in Figure 4, and the yellow points are the centers of R_{2m} s.

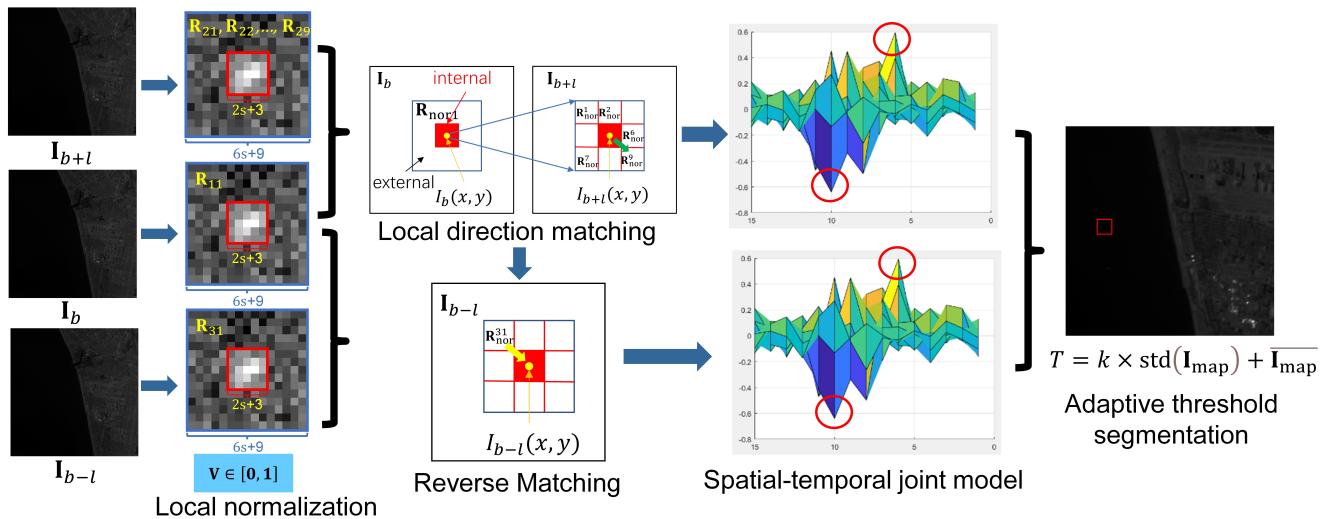


Figure 2. Overview of LSM.

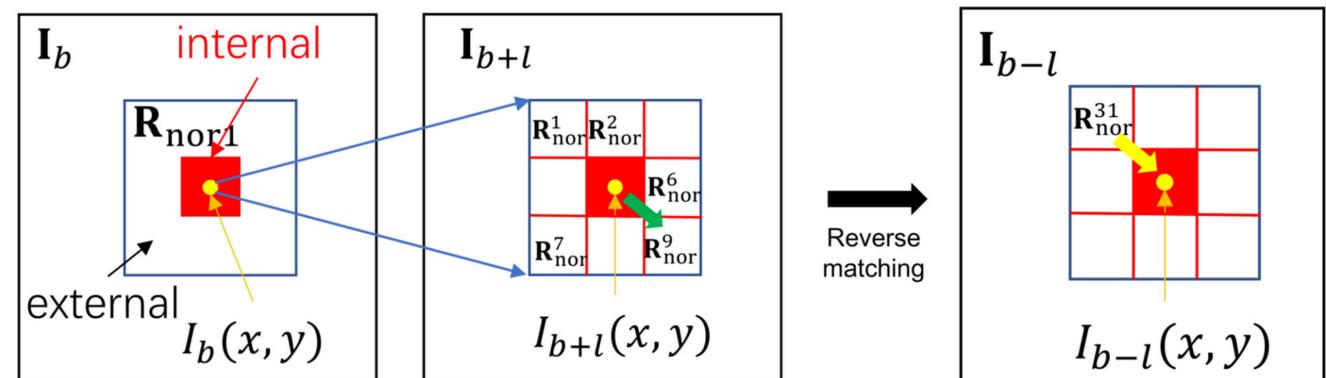


Figure 3. Local spatial-temporal matching and reverse matching.

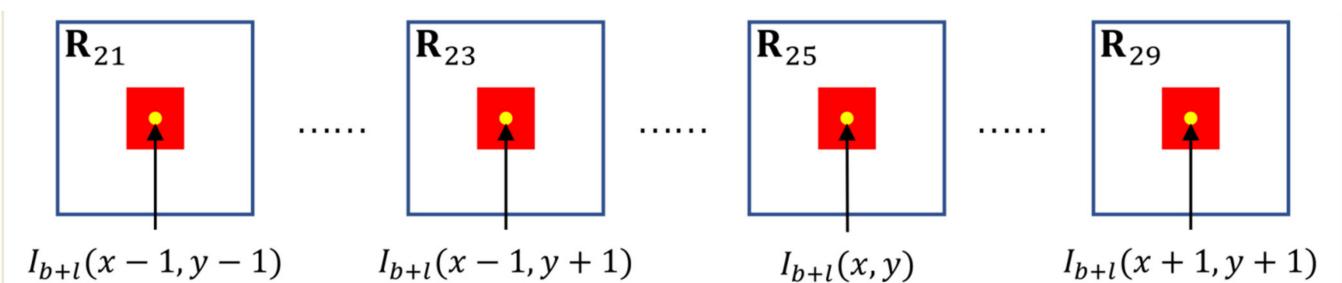


Figure 4. Positions of the local slices in the reference frame I_{b+l} .

The considerable difference between target and strong clutter causes missed detection and false alarm. Thus, the local normalization is designed to transfer the intensity into range [0, 1]. The definition of local normalization is as follows:

$$R_{\text{nor}1}(g, h) = \{[R_{11}(g, h) - \min(\mathbf{R}_{11})]/[\max(\mathbf{R}_{11}) - \min(\mathbf{R}_{11})]\}, \quad (3)$$

$$R_{\text{nor}}^m(g, h) = \{[R_{2m}(g, h) - \min(\mathbf{R}_{2m})]/[\max(\mathbf{R}_{2m}) - \min(\mathbf{R}_{2m})]\}, \quad (4)$$

where (g, h) represents the position in the \mathbf{R}_{11} and \mathbf{R}_{2m} s, $R_{\text{nor}1}(g, h)$ means the normalized value at the point (g, h) within \mathbf{R}_{11} , $R_{\text{nor}}^m(g, h)$ does the same. After the local normalization, both the target and clutter are processed within the value domain [0, 1] during the subsequent steps.

2.2. Local Direction Matching

When the space-based imaging system works under the staring mode, the backgrounds including the strong clutter in the IR sequence are moving within a tiny area. Therefore, the background can be supposed to move straightly in a short frame interval. Local direction matching is designed to determine which local slice of \mathbf{R}_{2m} in \mathbf{I}_{b+l} is the most similar to \mathbf{R}_{11} in the \mathbf{I}_b . In this paper, the local matching function is designed to measure the matching degree. The matching coefficient at point (x, y) is also determined. The functions are given as follows:

$$r_m = \frac{2 \times \sum_{g,h} \left[|R_{\text{nor}1}(g, h) - \overline{R}_{\text{nor}1}| \times |R_{\text{nor}}^m(g, h) - \overline{R}_{\text{nor}}^m| \right]}{\sum_{g,h} (R_{\text{nor}1}(g, h) - \overline{R}_{\text{nor}1})^2 + \sum_{g,h} (R_{\text{nor}}^m(g, h) - \overline{R}_{\text{nor}}^m)^2}, \quad (5)$$

$$r_1(x, y) = \max(r_m), \quad (6)$$

$$m_{\max} = \arg \max_m r_m, \quad (7)$$

where r_m s represents the matching degree between $\mathbf{R}_{\text{nor}1}$ and $\mathbf{R}_{\text{nor}}^m$ s, r_1 is the matching coefficient, and m_{\max} determines the local slice in $\mathbf{R}_{\text{nor}}^m$ s that is most similar to $\mathbf{R}_{\text{nor}1}$. As shown in Figure 3, if $m_{\max} = 9$, $\mathbf{R}_{\text{nor}}^9$ is the slice matching to $\mathbf{R}_{\text{nor}1}$, which means the background moves from (x, y) to $(x + 1, y + 1)$ during $[b, b + l]$, as illustrated by the green arrow at the local direction matching step.

2.3. Spatial–Temporal Joint Model

Once local direction matching is performed, suppression of strong clutter and aerial target enhancement can be achieved by the spatial–temporal joint model. First, the difference slice \mathbf{R}_{dif} is obtained by local slice difference:

$$\mathbf{R}_{\text{dif}} = \mathbf{R}_{\text{nor}1} - \mathbf{R}_{\text{nor}}^{m_{\max}}, \quad (8)$$

after which most backgrounds, including the clutter in \mathbf{I}_b , are suppressed initially, even if the clutter is much stronger than the aerial target.

The neighboring region of \mathbf{R}_{dif} is divided into internal and external regions, and their relationships are given as follows:

$$\Omega_{\text{int}} = \{(g, h) | \max(|g - x|, |h - j|) \leq s + 1\}, s = 1, 2, 3, \quad (9)$$

$$\Omega_{\text{int}} \cup \Omega_{\text{ext}} = \Omega_{\mathbf{R}_{\text{dif}}}, \quad (10)$$

$$\Omega_{\text{int}} \cap \Omega_{\text{ext}} = \emptyset, \quad (11)$$

where Ω_{int} and Ω_{ext} are the internal and external regions, respectively, and $\Omega_{\mathbf{R}_{\text{dif}}}$ represents the neighboring region of \mathbf{R}_{dif} , which has the same range as $\Omega_{\mathbf{R}_{11}}$; \emptyset is the null set. The

relationship between Ω_{int} and Ω_{ext} is illustrated in Figure 3, where the red region represents Ω_{int} , and the rest of the blue rectangle represents the range of Ω_{ext} .

Then, the nonuniformity stripes resulting from the inadequate preprocessing can be suppressed by the equation:

$$d_{\text{dif1}}(x, y) = \max(\mathbf{R}_{\text{int}}) - \max(\mathbf{R}_{\text{ext}}), \quad (12)$$

where \mathbf{R}_{int} is the matrix constructed by the pixels in Ω_{int} , and \mathbf{R}_{ext} does the same.

If the target appears, the dipole containing positive and negative peaks are left in \mathbf{R}_{int} . The dipole is highlighted by a pair of red circles in Figure 2. Thus, the dipole value at (x, y) is extracted:

$$d_{\text{dipole1}}(x, y) = [\max(\mathbf{R}_{\text{int}}) - \min(\mathbf{R}_{\text{int}})]^2, \quad (13)$$

where d_{dipole1} is the dipole value. At this step, the clutter can be further suppressed, but the aerial target can be significantly enhanced by quadratic operation.

Finally, the value of local spatial–temporal matching between \mathbf{I}_b and \mathbf{I}_{b+l} is calculated as:

$$I_{v1}(x, y) = [1 - r_1(x, y)] \times d_{\text{dif1}}(x, y) \times d_{\text{dipole1}}(x, y), \quad (14)$$

where $I_{v1}(x, y)$ represents the matching value at the point (x, y) .

2.4. Reverse Matching

After obtaining $I_{v1}(x, y)$, reverse matching is added into LSM. As shown in Figure 3, at (x, y) , if the offset from \mathbf{I}_b and \mathbf{I}_{b+l} is in the direction indicated by the green arrow, the offset from \mathbf{I}_{b-l} and \mathbf{I}_b is in the direction indicated by the yellow arrow in Figure 3, and \mathbf{I}_{b-l} is another reference frame. In this case, the local backgrounds in \mathbf{I}_{b-l} and \mathbf{I}_b can be reverse-matched. At (x, y) in \mathbf{I}_b , the offsets of local background from \mathbf{I}_{b-l} to \mathbf{I}_b can be determined by:

$$dy = \begin{cases} +1, & m_{\max} = 3, 6, 9 \\ 0, & m_{\max} = 2, 5, 8 \\ -1, & m_{\max} = 1, 4, 7 \end{cases}, \quad (15)$$

$$dx = \text{fix}((9 - m_{\max})/3) - 1, \quad (16)$$

where dy and dx denote the offsets in the horizontal and vertical directions, respectively, and $\text{fix}(*)$ represents the operation calculating the nearest integer in the direction to zero.

As shown in Figure 3, the local slice in \mathbf{I}_{b-l} , given by \mathbf{R}_{31} , is determined in the reverse matching step. The neighboring region of \mathbf{R}_{31} is formulated as:

$$\Omega_{\mathbf{R}_{31}} = \{(i, j) | \max(|x + dx - i|, |y + dy - j|) \leq 3 \times s + 4\}, \quad (17)$$

where $\Omega_{\mathbf{R}_{31}}$ denotes the neighboring region of \mathbf{R}_{31} . The normalized slice of \mathbf{R}_{31} and the matching coefficient between \mathbf{R}_{11} and \mathbf{R}_{31} are obtained by Equations (3) and (5), and $\mathbf{R}_{\text{nor}}^{31}$ and $r_2(x, y)$ represent the normalized slice and matching coefficient, respectively.

Finally, the spatial–temporal joint model is constructed. Identically, the value of local spatial–temporal matching between \mathbf{I}_b and \mathbf{I}_{b-l} is calculated by Equations (8)–(14) and represented by $I_{v2}(x, y)$.

2.5. Adaptive Threshold Segmentation

The mean filter is introduced to suppress noise, which is conducted as:

$$d_{\text{dif2}}(x, y) = R_{\text{nor1}}(3 \times s + 5, 3 \times s + 5) - \overline{\mathbf{R}_{\text{nor1}}}, \quad (18)$$

where $d_{\text{dif2}}(x, y)$ denotes the value after mean filtering at point (x, y) . In addition, the saliency map \mathbf{I}_{map} is obtained by:

$$I_{\text{map}}(x, y) = d_{\text{dif2}}(x, y) \times I_{v1}(x, y) \times I_{v2}(x, y), \quad (19)$$

where $I_{\text{map}}(x, y)$ is the map value at (x, y) . The results are shown in Figure 5; even though the target is much weaker than clutter in Figures 1 and 5a, it is enhanced significantly, and the strong clutter is well suppressed.

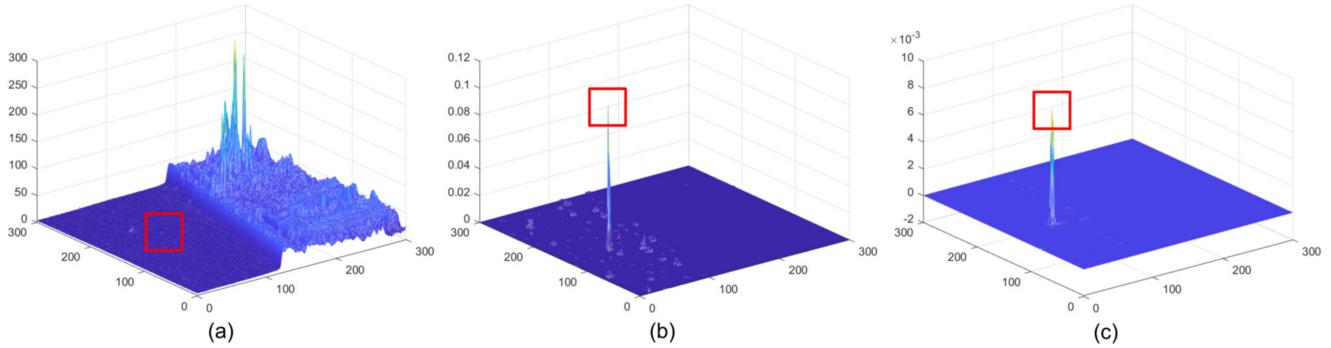


Figure 5. (a) Three-view of raw image; (b) three-view of the result after spatial–temporal matching; (c) three-view of the saliency map.

In I_{map} , clutter and background are suppressed, but the IR aerial target is retained and enhanced. Finally, the aerial target is detected by adaptive threshold segmentation:

$$T = k \times \text{std}(\mathbf{I}_{\text{map}}) + \overline{\mathbf{I}_{\text{map}}}, \quad (20)$$

where $\text{std}(*)$ represents the standard deviation operation, and k is the segmentation parameter. k has been experimentally proved to show that $k \in [10, 30]$ is effective. When the value of the element in \mathbf{I}_{map} is greater than T , it is set to one, and the opposite is set to zero. The point set to one is the aerial target. The entire procedure of LSM is given in Algorithm 1.

Algorithm 1 Procedure of LSM.

Input: Base frame \mathbf{I}_b , reference frames \mathbf{I}_{b-l} and \mathbf{I}_{b+l} .
Output: The position of the aerial target.
(1) Obtain the size $[row, col]$ of \mathbf{I}_b .
(2) **for** $x = 1 : row$ **do**
(3) **for** $y = 1 : col$ **do**
(4) Obtain the local slices \mathbf{R}_{11} and \mathbf{R}_{2m} s by Equations (1) and (2);
(5) Obtain the normalized slices $\mathbf{R}_{\text{nor}1}$ and $\mathbf{R}_{\text{nor}s}^m$ s by Equations (3) and (4);
(6) Calculate the matching coefficient $r_1(x, y)$ and determine the $\mathbf{R}_{2m_{\max}}$ by Equations (5)–(7);
(7) Construct the spatial–temporal joint model between \mathbf{I}_b and \mathbf{I}_{b+l} and calculate $I_{v1}(x, y)$ by Equations (8)–(14);
(8) Conduct reverse matching and obtain \mathbf{R}_{31} by Equations (15)–(17);
(9) Calculate the normalized slice of \mathbf{R}_{31} by Equation (3);
(10) Calculate the matching coefficient $r_2(x, y)$ by Equation (5);
(11) Construct the spatial–temporal joint model between \mathbf{I}_{b-l} and \mathbf{I}_b and calculate $I_{v2}(x, y)$ by Equations (8)–(14);
(12) Calculate the saliency map value $I_{\text{map}}(x, y)$ by Equations (18) and (19);
(13) **end for**
(14) **end for**
(15) Obtain the saliency map \mathbf{I}_{map} ;
(16) Calculate the adaptive threshold T by formula Equation (20);
(17) Output the position of the aerial target.

3. Experiments

3.1. Experimental Condition and Evaluation Index

The datasets used for experiments were four space-based IR sequences with different backgrounds, and the aerial targets were simulated targets with the same intensity distri-

bution proportions as the real targets. The backgrounds and real targets were obtained from the identical space-based system working under staring imaging mode. The details of the four sequences are found in Table 1. The speeds of aerial targets ranged from 1.1 to 2.0 pixel/frame. The value of l was set to two after parameter analysis conducted in Section 4.

Table 1. Details of experimental datasets.

Dataset	Size		Target Size	Details of Backgrounds
	Frame			
Seq.1	300 × 300		7 × 7	Sea and land background, strong clutter, background moving speed is 0.11 pixel/frame;
	100			
Seq.2	200 × 256		7 × 7	Sea background, strong clutter, residual nonuniformity stripe, background moving speed is 0.14 pixel/frame;
	135			
Seq.3	200 × 256		7 × 7	Land background, strong clutter, background moving speed is 0.53 pixel/frame;
	135			
Seq.4	200 × 256		5 × 5	Sea and land background, strong clutter, background moving speed is 0.24 pixel/frame.
	120			

To evaluate the detection effectiveness, LSMs are compared with seven state-of-the-art detection methods, including fusion saliency map (FSM) [10], double-neighborhood gradient method (DNGM) [11], neighborhood saliency map (NSM) [14], spatial-temporal local contrast filter (STLCF) [15], spatial-temporal local contrast method (STLCM) [2], spatial-temporal joint processing model (STJP) [28], and multiscale local target characteristics algorithm (MLTC) [29]. NSM, STLCF, STLCM, and STJP are existing space-based detection methods, FSM is a newly proposed detection method utilized for low-altitude slow target detection that has a similar background to space-based detection, and DNGM and MLTC are new detection methods proposed in 2020 and 2021, respectively.

The evaluation indices are background suppression factor (*BSF*), the gain of signal-to-clutter ratio (*GSCR*), detection rate (P_d), false alarm rate (P_f), and area under the curve (AUC). *BSF* is a global index for evaluating the performance of global background suppression and is defined as:

$$BSF = \sigma_{\text{before}} / \sigma_{\text{after}}, \quad (21)$$

where σ_{after} and σ_{before} are standard deviations of the processed and raw image, respectively. *GSCR* is the index used to evaluate the target enhancement performance. The *GSCR* is calculated by:

$$SCR = |\mu_{\text{tar}} - \mu_{\text{bk}}| / \sigma_{\text{bk}}, \quad (22)$$

$$GSCR = SCR_{\text{after}} / SCR_{\text{before}}, \quad (23)$$

where μ_{tar} represents the mean of the target, and σ_{bk} are the mean and standard deviation of the background, respectively, and SCR_{after} and SCR_{before} are *SCR* values of the processed and primitive targets, respectively.

The indices used to evaluate the detection effectiveness are P_d and P_f , whose formulas are:

$$P_d = N_{\text{detected}} / N_{\text{real}}, \quad (24)$$

$$P_f = N_{\text{false}} / N_{\text{pixel}}, \quad (25)$$

where N_{detected} is the number of real targets detected by the method, N_{real} is the total number of real targets, N_{false} is the number of targets falsely detected, and N_{pixel} is the number of pixels. To visualize the detection effectiveness, a receiver operating characteristic curve (ROC) was drawn according to the relationship between P_d and P_f . The area under the ROC curve is represented by AUC.

3.2. Experimental Results

The three views of the results corresponding to different methods are given in Figures 6–9. The experimental results under two situations are both exhibited. Under the first situation where the aerial target is much weaker than clutter, as shown in images in Seqs.1, 2, and 4, STLCF, STLCM, FSM, NSM, and DNGM suppress most of the background in the images, but the clutter with strong intensities is still retained. STJP is sensitive to clutter. Only MLTC and our method can enhance the weak targets. However, MLTC still enhances the background component, which results in false alarms. The images in Seq.3 show a situation in which the intensity of a small target is close to strong clutter. Most methods, such as STLCF, STLCM, STJP, MLTC, and the proposed method, perform well on target enhancement, but STLCM, STJP, and MLTC also bring a great number of false alarms. According to the above comparison, our method fits the task to suppress the strong clutter and enhance the weak target simultaneously. The quantitative comparison and analysis are given in the next part, for more compelling results.

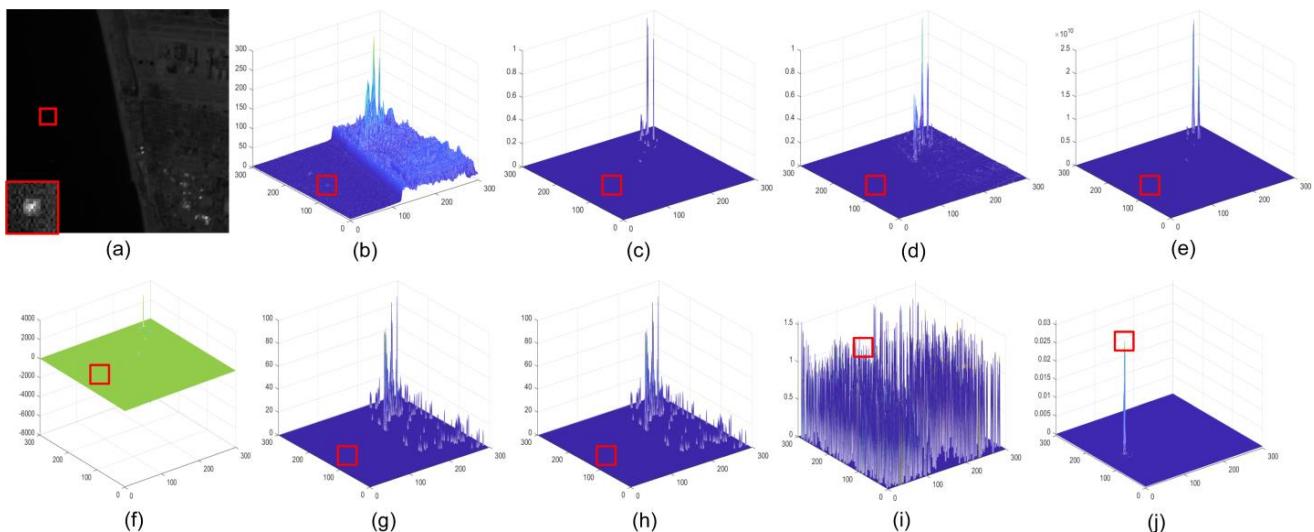


Figure 6. Detection results of different methods in Seq.1. The positions of aerial targets are highlighted by the red rectangles; (a) raw image, the details of target are magnified at the left bottom corner; (b) three-view of raw image; (c) result of STLCM; (d) result of STLCF; (e) result of FSM; (f) result of NSM; (g) result of STJP; (h) result of DNGM; (i) result of MLTC; (j) result of LSM.

The results of *BSF* are listed in Table 2. LSM achieved the highest *BSF* value on Seq.4 but had lower values than FSM or STLCM on the other three sequences because of the zero-setting operation in the two methods. In FSM, when the mean of the variance difference between the internal window and the external window is less than zero, the spatial variance saliency map value of a pixel will be zero, and the output will be zero eventually. In STLCM, the final value of a pixel will be set to zeros if this pixel is not the local maximum point. In an IR image, the local maximum points are usually composed of the small target, clutter, and noise. Therefore, because of the zero-setting operation, those pixels around the local maximum point will be assigned to values of zero. It is clear that the more zero points the final saliency map has, the lower the value of σ_{after} in Equation (21) will be, and the value of *BSF* will consequently increase.

It is worth noting that *BSF* is a global index evaluating the background suppression ability of a method in the whole image, but clutter only accounts for a little proportion in the background. Thus, taking the results in Figures 6–9 and Table 2 into consideration, STLCM and FSM fail to suppress the clutter even if they suppress conventional background suppression better than our method. The other methods received lower values than LSM on all sequences. These results indicate that most methods can suppress most backgrounds but have poor abilities to suppress the strong clutter on the space-based IR images. STLCM and

STLCF suppress background by the direct interframe difference, in which the weak target is suppressed but residual clutter still exists. MLTC, NSM, and DNGM fail to suppress clutter since it has a similar spatial distribution as the aerial target.

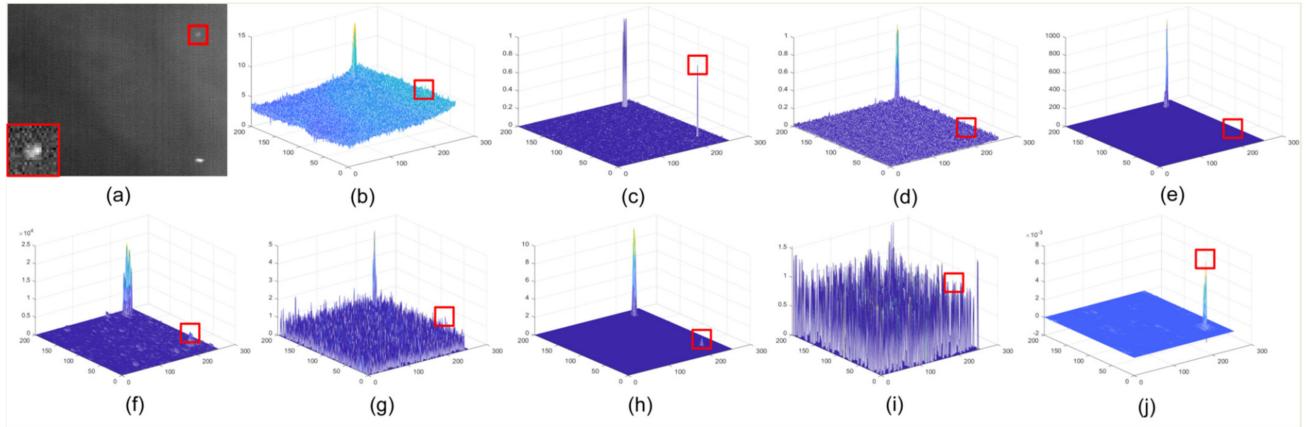


Figure 7. Detection results of different methods in Seq.2. (a) raw image, the details of target are magnified at the left bottom corner; (b) three-view of raw image; (c) result of STLCM; (d) result of STLCF; (e) result of FSM; (f) result of NSM; (g) result of STJP; (h) result of DNGM; (i) result of MLTC; (j) result of LSM.

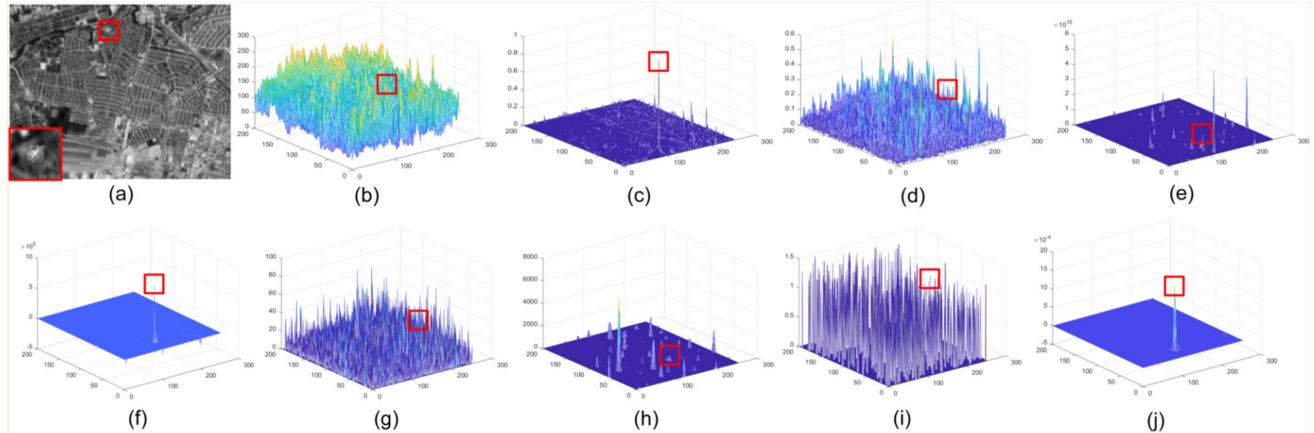


Figure 8. Detection results of different methods in Seq.3. (a) raw image, the details of target are magnified at the left bottom corner; (b) three-view of raw image; (c) result of STLCM; (d) result of STLCF; (e) result of FSM; (f) result of NSM; (g) result of STJP; (h) result of DNGM; (i) result of MLTC; (j) result of LSM.

The average GSCR values are listed in Table 3. LSM receives the best results on Seqs.1, 3, and 4. STJP and STLCF find it hard to enhance the target in the space-based IR image. STLCM, FSM, and NSM receive malfunctions on Seqs.1, 2, and 4 because of the considerable intensity difference between target and clutter. Only DNGM shows a better result than LSM on Seq.3 but enhances clutter better, as shown in Figure 7. MLTC can enhance the aerial targets that are much weaker than clutter in space-based images, but our method performs better than MLTC, as shown in Table 3.

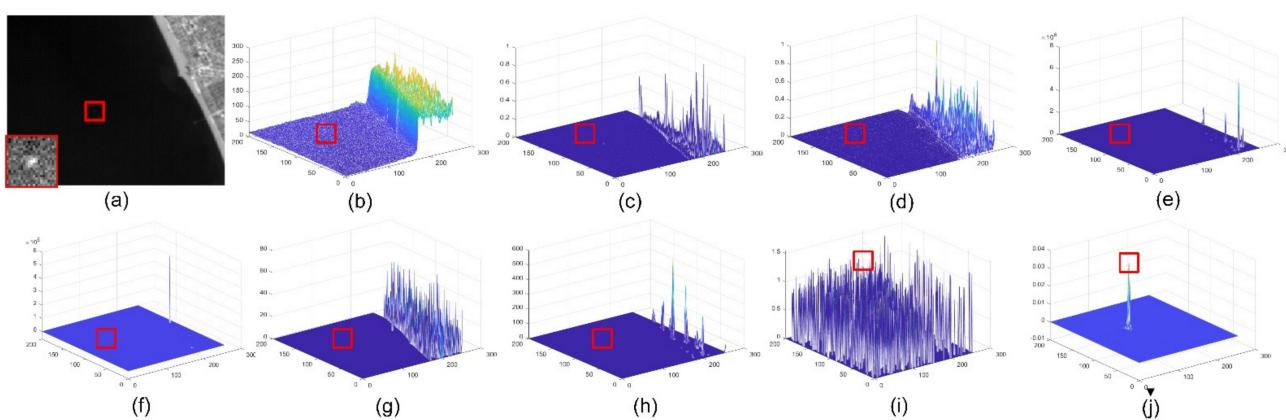


Figure 9. Detection results of different methods in Seq.4. (a) raw image, the details of target are magnified at the left bottom corner; (b) three-view of raw image; (c) result of STLCM; (d) result of STLCF; (e) result of FSM; (f) result of NSM; (g) result of STJP; (h) result of DNGM; (i) result of MLTC; (j) result of LSM.

Table 2. Average BSF values of different methods for four sequences.

Methods	Seq.1	Seq.2	Seq.3	Seq.4
STLCF	7.0835	3.7508	2.4573	4.2582
STLCM	14.9534	8.3428	13.0866	13.7626
FSM	11.4422	7.3382	11.1237	28.0876
NSM	9.0671	2.4982	13.9567	20.3182
STJP	5.7969	1.1265	1.9641	4.8315
DNGM	6.0336	4.1818	6.2833	14.9458
MLTC	1.2551	0.6357	1.5488	2.1215
Proposed	9.5974	6.7767	14.2828	24.0300

The results of detection effectiveness are shown by the ROC curves and AUC values in Figure 10 and Table 4. The P_{ds} s of LSM are more than 85% on the four sequences when P_f s are 10^{-4} and more than 97% when s are 10^{-3} . The AUC values of LSM on the four sequences are 0.9994, 0.9990, 0.9986, and 0.9995, respectively. However, the results of other methods are unstable on the four sequences. In Figure 10a,d, most AUC values of the most compared methods are less than 0.8 because Seqs.1 and 4 have backgrounds of sea and land, and the intensity of clutter is at least 10 times stronger than that of the target. On Seq.2, DNGM has an AUC value of more than 0.99 because it enhances the target better than LSM. However, the effectiveness of DNGM drops sharply when $P_f < 10^{-4}$ because DNGM enhances clutter better than target, and MLTC does the same on four sequences. On Seqs.3, the compared methods achieved better detection performance; the five methods obtained AUC values of more than 0.99 because the intensities of the target are close to those of clutter. The average AUC values are also calculated and given in Table 4; MLTC receives the highest average value of 0.9534 because of great target enhancement ability, while the minimum value is 0.7345 belonging to the single-frame method NSM. Compared with the seven methods, the average AUC value of LSM is 0.9991, indicating that LSM has the best detection effectiveness.

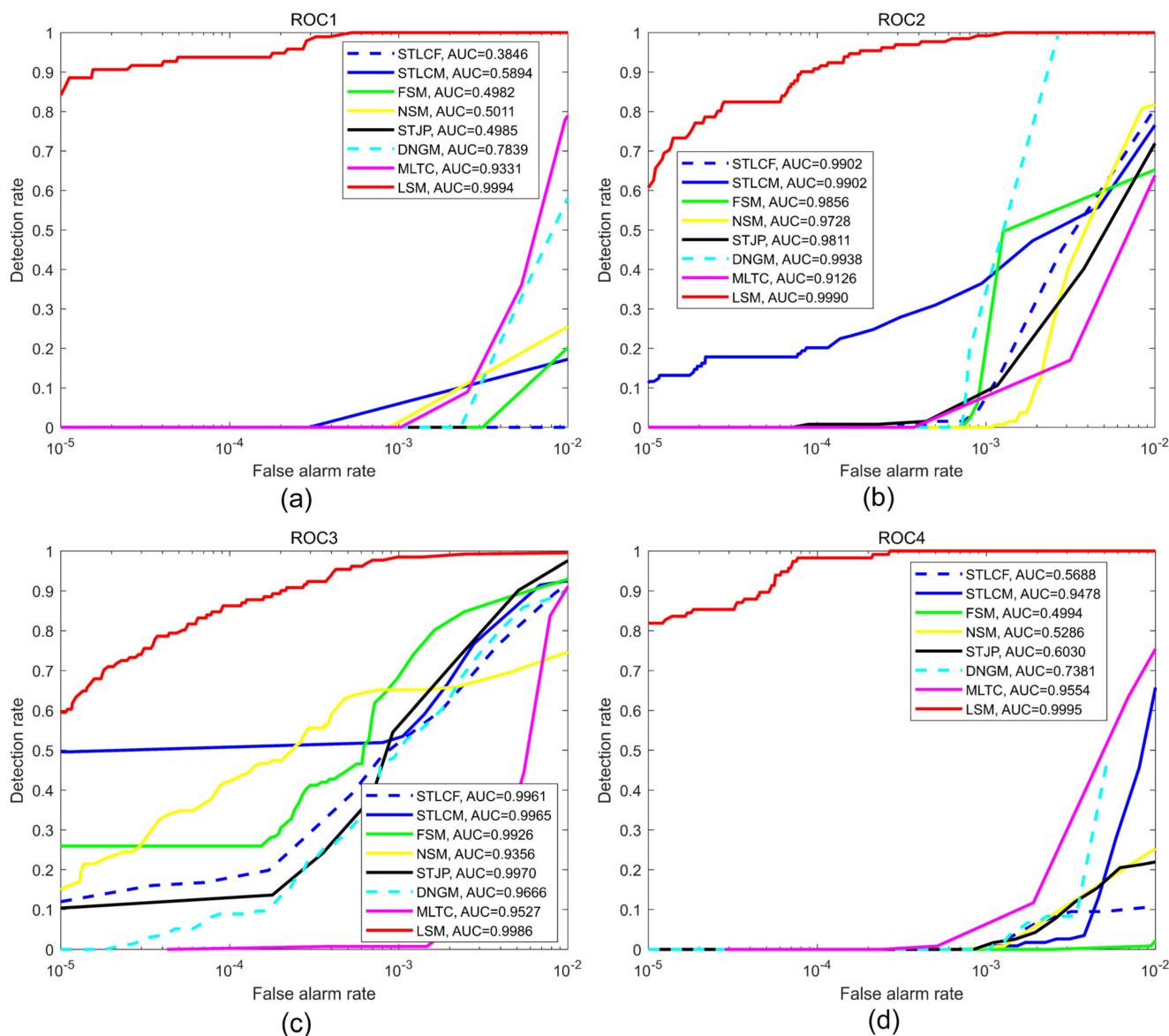


Figure 10. ROC curves and AUC values of different methods for the four sequences; (a–d) represent Seqs.1–4, respectively.

Table 3. Average GSCR values of different methods for four sequences.

Methods	Seq.1	Seq.2	Seq.3	Seq.4
$\overline{SCR}_{\text{before}}$	2.9400	1.7764	2.4502	2.2381
STLCF	1.7490	0.7818	0.6588	0.0897
STLCM	1.7231	2.8566	2.0138	0.5797
FSM	1.7213	1.3366	5.9280	0.0028
NSM	0.0118	0.7149	2.1257	0.0269
STJP	0.6507	0.1349	0.4681	0.0795

Table 3. Cont.

Methods	Seq.1	Seq.2	Seq.3	Seq.4
DNGM	0.7587	12.8578	7.0548	0.4568
MLTC	13.1471	1.2697	8.7553	28.2292
Proposed	17.5890	7.4260	11.7774	28.4823

Table 4. AUC values of different methods. The last column shows the average values.

Methods	Seq.1	Seq.2	Seq.3	Seq.4	Average
STLCF	0.3846	0.9902	0.9961	0.5688	0.7350
STLCM	0.5894	0.9902	0.9965	0.9478	0.8809
FSM	0.4982	0.9856	0.9926	0.4994	0.7440
NSM	0.5011	0.9728	0.9356	0.5286	0.7345
STJP	0.4985	0.9811	0.9970	0.6030	0.7699
DNGM	0.7839	0.9938	0.9666	0.7381	0.8706
MLTC	0.9331	0.9126	0.9527	0.9554	0.9534
Proposed	0.9994	0.9990	0.9986	0.9995	0.9991

According to the experimental results, LSM performs better than the seven compared methods and can detect the aerial targets more effectively on the space-based IR sequences with different backgrounds. The results prove that LSM can conquer the challenge of enhancing the weak target and suppressing the strong clutter simultaneously.

4. Analysis and Discussion

Seqs.1–3 contain 7×7 targets and Seq.4 contains 5×5 targets. According to the results, our method enhanced both targets significantly. The GSCR values in Table 3 revealed that our method enhanced the targets with different sizes considerably. As for the detection effectiveness, our method obtained $P_d > 98\%$ when the P_f s reached 10^{-3} on all sequences. Meanwhile, the values of AUC were more than 0.9986. The above results indicate our method can maintain its effectiveness when detecting targets of different sizes.

In order to analyze the influence of l , we selected Seq.1 as the example with which to explore the influence of parameter l , in which the target speed is 1.55 pixel/frame; the experiments were conducted with different values of l . Results of \overline{BSF} and \overline{GSCR} are given in Table 5, and the detection effectiveness is shown in Figure 11. The results indicate that there are few evidently different results between l and $-l$ but that the tendencies of BSF and $GSCR$ are different with $|l|$ increase. With $|l|$ increase, the matching coefficient r_1 decreases, but the dipole is clearer; thus, the target can be further enhanced, but the background cannot be well suppressed. The detection effectiveness shown in Figure 11 reveals that it generates similar detection results when $l = \pm 2$ or ± 3 , which is better than the result when $l = \pm 1$. Therefore, the value of l is recommended to be ± 2 , and set to two in this paper.

Table 5. The values of GSCR and BSF with different frame intervals.

l	−3	−2	−1	1	2	3
\overline{GSCR}	24.7905	17.2655	10.7153	10.7153	17.5890	24.7880
\overline{BSF}	9.3730	9.4456	9.5933	9.5933	9.5974	9.3730

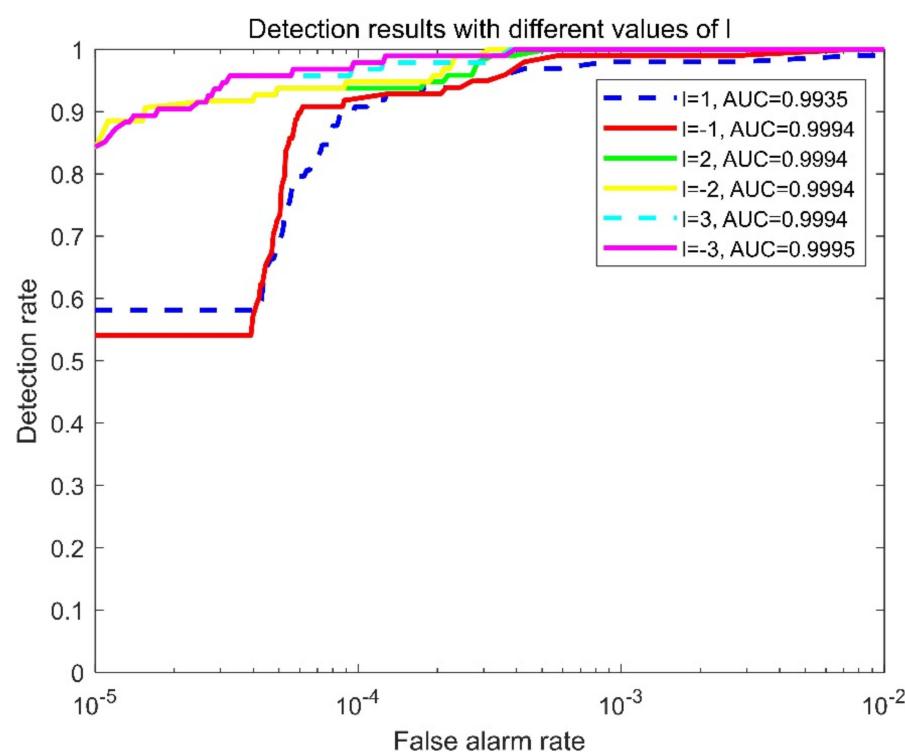


Figure 11. ROCs and AUC values with different frame intervals.

The segmentation parameter k directly influences the detection effectiveness; the relationships between k and detection results are shown in Figure 12. With the increase in the k , both detection rate and false alarm rate decrease. The detection rates of different experimental sequences showed similar trends, and the detection rates could be maintained above 90% when $10 \leq k \leq 30$. The variation trends of the false alarm rates of the five test sequences are nearly the same as well. When $k \geq 10$, the false alarm rates of all the sequences are less than 10^{-3} . In order to maintain the detection rate $\geq 90\%$ and false alarm rate $\leq 0.5 \times 10^{-3}$, the value range of k in this method is recommended to be $[10, 30]$, which was given in Section 2.5.

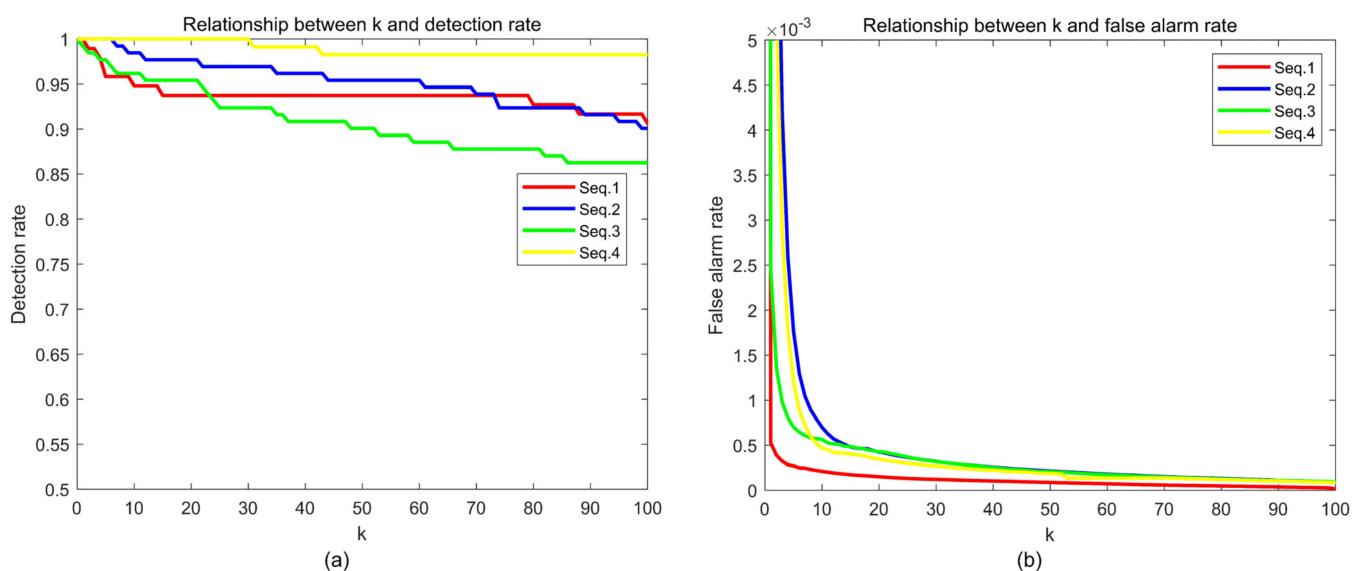


Figure 12. Influence of k . (a) Relationship between k and P_d ; (b) relationship between k and P_f .

5. Conclusions

This paper proposes a concise method, which is based on local spatial–temporal matching, for detecting an aerial target on a space-based IR platform. The experimental results determine that, compared with existing methods, LSM exhibits better detection performance when the clutter is much stronger than the aerial target. However, LSM is currently only suitable for the staring imaging mode and still needs to be optimized to adapt to other modes.

Author Contributions: Conceptualization, L.C. and X.C.; methodology, L.C.; software, L.C.; validation, X.C.; formal analysis, X.C.; resources, P.R. and M.H.; data curation, M.H.; writing—original draft preparation, L.C.; writing—review and editing, X.C. and P.R.; supervision, P.R.; funding acquisition, P.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China, grant number 62175251.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Liu, G.; Sun, X.; Fu, K.; Wang, H. Aircraft Recognition in High-Resolution Satellite Images Using Coarse-to-Fine Shape Prior. *IEEE Geosci. Remote Sens. Lett.* **2013**, *3*, 573–577. [[CrossRef](#)]
2. Zhao, B.; Xiao, S.; Lu, H.; Wu, D. Spatial-temporal local contrast for moving point target detection in space-based infrared imaging system. *Infr. Phys. Technol.* **2018**, *95*, 53–60. [[CrossRef](#)]
3. Cawley, S.J. The Space Technology Research Vehicle 2 Medium Wave Infra Red Imager. *Acta Astronaut.* **2003**, *52*, 717–726. [[CrossRef](#)]
4. Zhu, H.; Li, Y.; Hu, T.; Rao, P. Key parameters design of an aerial target detection system on a space-based platform. *Opt. Eng.* **2018**, *57*, 1. [[CrossRef](#)]
5. Wang, X.; Peng, Z.; Kong, D.; He, Y. Infrared dim and small target detection based on stable multisubspace learning in heterogeneous scene. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 5481–5493. [[CrossRef](#)]
6. Cao, Y.; Liu, R.; Yang, J. Small target detection using two-dimensional least mean square (TDLMS) filter based on neighborhood analysis. *Int. J. Infrared Millim. Waves* **2008**, *29*, 188–200. [[CrossRef](#)]
7. Zeng, M.; Li, J.; Peng, Z. The design of top-hat morphological filter and application to infrared target detection. *Infr. Phys. Technol.* **2006**, *48*, 67–76. [[CrossRef](#)]
8. Lu, Y.; Dong, L.; Zhang, T.; Xu, W. A robust detection algorithm for infrared maritime small and dim targets. *Sensors* **2020**, *20*, 1237. [[CrossRef](#)]
9. Chen, C.; Li, H.; Wei, Y.; Xia, T.; Tang, Y.Y. A Local Contrast Method for Small Infrared Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 574–581.
10. Pang, D.; Shan, T.; Ma, P.; Li, W.; Liu, S.; Tao, R. A Novel Spatiotemporal Saliency Method for Low-Altitude Slow Small Infrared Target Detection. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*. [[CrossRef](#)]
11. Wu, L.; Ma, Y.; Fan, F.; Wu, M.; Huang, J. A Double-Neighborhood Gradient Method for Infrared Small Target Detection. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 1476–1480. [[CrossRef](#)]
12. Han, J.; Moradi, S.; Faramarzi, I.; Zhang, H.; Zhao, Q.; Zhang, X.; Li, N. Infrared small target detection based on the weighted strengthened local contrast measure. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 1670–1674. [[CrossRef](#)]
13. Wang, H.; Zhao, Z.; Kwan, C.; Zhou, G.; Chen, Y. New Results on Small and Dim Infrared Target Detection. *Sensors* **2021**, *21*, 7746. [[CrossRef](#)] [[PubMed](#)]
14. Lv, P.; Sun, S.; Lin, C.; Liu, G. A method for weak target detection based on human visual contrast mechanism. *IEEE Geosci. Remote Sens. Lett.* **2018**, *16*, 261–265. [[CrossRef](#)]
15. Deng, L.; Zhu, H.; Tao, C.; Wei, Y. Infrared moving point target detection based on spatial–temporal local contrast filter. *Infr. Phys. Technol.* **2016**, *76*, 168–173. [[CrossRef](#)]
16. Gao, C.; Meng, D.; Yang, Y.; Wang, Y.; Zhou, X.; Hauptmann, A.G. Infrared Patch-Image Model for Small Target Detection in a Single Image. *IEEE Trans. Image Process.* **2013**, *22*, 4996–5009. [[CrossRef](#)]
17. Zhou, F.; Wu, Y.; Dai, Y.; Ni, K. Robust infrared small target detection via jointly sparse constraint of l_1/l_2 -metric and dual-graph regularization. *Remote Sens.* **2020**, *12*, 1963. [[CrossRef](#)]

18. Liu, H.-K.; Zhang, L.; Huang, H. Small target detection in infrared videos based on spatio-temporal tensor model. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 8689–8700. [[CrossRef](#)]
19. Kong, X.; Yang, C.; Cao, S.; Li, C.; Peng, Z. Infrared small target detection via nonconvex tensor fibered rank approximation. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–21. [[CrossRef](#)]
20. Yang, P.; Dong, L.; Xu, W. Infrared small maritime target detection based on integrated target saliency measure. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2021**, *14*, 2369–2386. [[CrossRef](#)]
21. Zhao, M.; Li, W.; Li, L.; Ma, P.; Cai, Z.; Tao, R. Three-order tensor creation and tucker decomposition for infrared small-target detection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–16. [[CrossRef](#)]
22. Zhang, T.; Peng, Z.; Wu, H.; He, Y.; Li, C.; Yang, C. Infrared small target detection via self-regularized weighted sparse model. *Neurocomputing* **2021**, *420*, 124–148. [[CrossRef](#)]
23. Dai, Y.; Wu, Y.; Zhou, F.; Barnard, K. Attentional local contrast networks for infrared small target detection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 9813–9824. [[CrossRef](#)]
24. Ying, X.; Wang, Y.; Wang, L.; Sheng, W.; Liu, L.; Lin, Z.; Zho, S. MoCoPnet: Exploring Local Motion and Contrast Priors for Infrared Small Target Super-Resolution. *arXiv* **2022**, arXiv:2201.01014.
25. Ju, M.; Luo, J.; Liu, G.; Luo, H. ISTDet: An efficient end-to-end neural network for infrared small target detection. *Infr. Phys. Technol.* **2021**, *114*, 103659. [[CrossRef](#)]
26. Hou, Q.; Wang, Z.; Tan, F.; Zhao, Y.; Zheng, H.; Zhang, W. RISTDnet: Robust infrared small target detection network. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [[CrossRef](#)]
27. Lv, P.-Y.; Sun, S.-L.; Lin, C.-Q.; Liu, G.-R. Space moving target detection and tracking method in complex background. *Infr. Phys. Technol.* **2018**, *91*, 107–118. [[CrossRef](#)]
28. Lv, P.-y.; Lin, C.-q.; Sun, S.-l. Dim small moving target detection and tracking method based on spatial-temporal joint processing model. *Infr. Phys. Technol.* **2019**, *102*, 102973. [[CrossRef](#)]
29. Lin, B.; Yang, X.; Wang, J.; Wang, Y.; Wang, K.; Zhang, X. A Robust Space Target Detection Algorithm Based on Target Characteristics. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [[CrossRef](#)]