

# Journal of Applied Remote Sensing

RemoteSensing.SPIEDigitalLibrary.org

## Ship detection and tracking method for satellite video based on multiscale saliency and surrounding contrast analysis

Haichao Li  
Liang Chen  
Feng Li  
Meiyu Huang

**SPIE.**

Haichao Li, Liang Chen, Feng Li, Meiyu Huang, "Ship detection and tracking method for satellite video based on multiscale saliency and surrounding contrast analysis," *J. Appl. Remote Sens.* **13**(2), 026511 (2019), doi: 10.1117/1.JRS.13.026511.

# Ship detection and tracking method for satellite video based on multiscale saliency and surrounding contrast analysis

Haichao Li,\* Liang Chen, Feng Li, and Meiyu Huang

China Academy of Space Technology, Qian Xuesen Laboratory of Space Technology,  
Beijing, China

**Abstract.** In port surveillance, monitoring based on satellite video is a valuable supplement to a ground monitoring system because of its wide monitoring range. Therefore, automatic ship detection and tracking based on satellite video is an important research field. However, because of the small size of ships without texture and the interference of sea noise, it is also a challenging subject. An approach of automatic detection and tracking moving ships of different sizes using satellite video is presented. First, motion compensation between two frames is realized. Then, saliency maps of multiscale differential image are combined to create dynamic multiscale saliency map (DMSM), which is more suitable for the detection of ships of different sizes. Third, candidate motion regions are segmented from DMSM, and moving ships can be detected after the false alarms are removed based on the surrounding contrast. Fourth, important elements such as centroid distance, area ratio, and histogram distance from moving ships are used to perform ship matching. Finally, ship association and tracking are realized by using the intermediate frame in every three adjacent frames. Experimental results on satellite sequences indicate that our method can effectively detect and track ships and obtain the target track, which is superior in terms of the defined recall and precision compared with other classical target tracking methods. © The Authors. Published by SPIE under a Creative Commons Attribution 4.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JRS.13.026511](https://doi.org/10.1117/1.JRS.13.026511)]

**Keywords:** satellite video; multiscale saliency map; ship detection; ship tracking.

Paper 190018 received Jan. 11, 2019; accepted for publication May 10, 2019; published online Jun. 6, 2019.

## 1 Introduction

The automatic monitoring of ships using video surveillance plays an important role in ocean security and maritime transportation, fishery management, ship traffic surveillance, and so on. In general, the video sources of ship monitoring systems are mainly taken from ground-based<sup>1,2</sup> or aerial-based<sup>3,4</sup> cameras. However, these systems have some shortcomings such as poor concealment, limited space coverage, and the required sensor installation and maintenance.

Fortunately, satellite-based monitoring can overcome the above shortcomings. However, video satellite was not available until recently. Therefore, in the past several years, many ship detection algorithms have been proposed based on synthetic aperture radar (SAR) images because of SAR's imaging capacity during day and night under most meteorological conditions.<sup>5-8</sup> Compared with SAR images, optical satellite images have higher spatial resolution that are more suitable for ship detection, and they are important supplements to SAR images.

Therefore, although optical satellite images are usually disturbed by weather conditions, such as clouds or ocean waves,<sup>9</sup> a number of ship detection methods have been developed from a single high-resolution optical satellite image.<sup>10-13</sup> Most of these methods adopt the coarse-to-fine strategies, which can be divided into ship candidate extraction stage and false alarm elimination stage. In the first stage, these methods extract candidate ships according to the differences of gray values between potential targets and background.<sup>14,15</sup> In the second stage, most of the algorithms utilize ship features with candidate classifiers to discriminate ships from false

---

\*Address all correspondence to Haichao Li, E-mail: [lihaichao2000@163.com](mailto:lihaichao2000@163.com)

alarms,<sup>16</sup> and an important issue is to find efficient descriptors to describe the ship targets. Further, some existing methods use *a priori* coastline data to detect sea area;<sup>17</sup> however, they are still disturbed by coastal areas due to low accuracy of coastline data.

However, the low-temporal resolution of existing satellite images limits the timeliness of ship detection and tracking. Fortunately, optical video satellites, such as Skysat and Jilin-1, have been developed due to the advancement of camera technology for highly spatial resolution and temporal resolution systems. In recent years, several researches on target detection and tracking based on optical satellite videos have been carried out.<sup>18–20</sup>

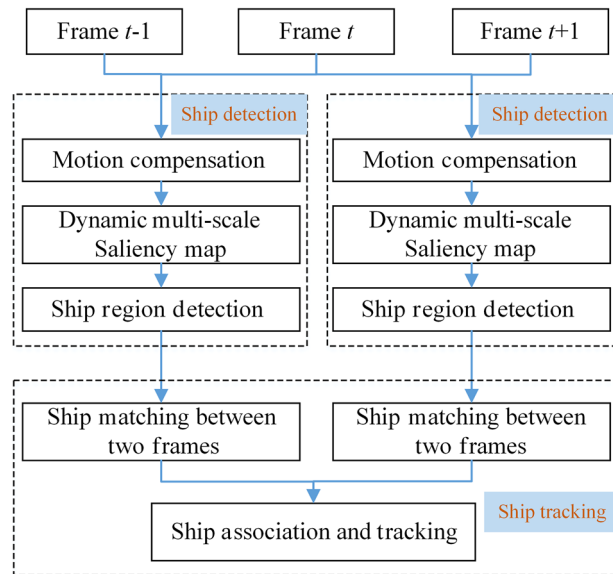
It should be noted that satellite video plays an important role in ship monitoring because of its strong concealment, wide monitoring range, and real-time continuous monitoring. Although much progress has been achieved in ship detection based on static satellite images, few studies have focused on moving ship detection and tracking based on satellite video, which will become a new research topic in the field of remote sensing. Rao et al.<sup>21</sup> proposed to estimate ship speed and direction by locating them and their tracks from multisatellite imagery. Yao et al.<sup>22</sup> and Zhang et al.<sup>23</sup> proposed ship-tracking methods in GF-4 satellite sequential imagery based on the automatic identification system data of ships, which are used for cooperative ships' surveillance. As shown in Fig. 1, ship features can be easily extracted and recognized from ground-based video [see Fig. 1(a)] and aerial-based video [see Fig. 1(b)] because of the sufficient spatial resolution, which are convenient for feature description in ships detection and tracking. Compared with general ground-based video and aerial-based video, there are several problems in ship detection and tracking based on satellite video (see Fig. 1). The moving ships in satellite-based video range from just a few pixels to dozens of pixels with similar brightness due to the limited resolution and even exhibit low contrast to the background due to the influence of cloud and sea wave. From the above characteristics of moving ships in satellite-based video, it is difficult to extract available appearance feature or texture information inside the ship. Maybe some small ships are submerged in a complex background. Another problem for the ship detection and tracking is the complex background due to the large field of view, in which there are disturbances, such as clouds, land moving targets, and so on. Accounting for the characteristics of satellite video, the traditional target detection and tracking methods are difficult to perform well for ships in satellite video.

The goal of this paper is to detect and track moving ships based on satellite video. Inspired by the multiscale selective cognition property of the human visual system, we propose an integrated algorithm of ship detection and tracking. The main contributions can be summarized as the following three aspects: (1) in order to detect ships of different sizes, a dynamic multiscale saliency map (DMSM) is proposed to compute the differential image between two frames, which can detect motion ships of small displacement and avoid the possible holes inside motion ships generated by frame difference method. (2) The surrounding contrast and ship characteristics are utilized to discriminate moving ships from false alarms such as clouds and land moving target. (3) An integrated ship detection and tracking scheme using the intermediate frame in every three adjacent frames is proposed, which avoids the problem that all frames are registered to the same reference frame.

The rest of this paper is organized as follows. Section 2 introduces the proposed moving ship detection and tracking method. Section 3 describes the experimental results and analysis, and Sec. 4 gives the conclusion.



**Fig. 1** Comparison of three types of videos. (a) Ground video frame. (b) Aerial video frame. (c) Satellite video frame.



**Fig. 2** Flowchart of the proposed ship detection and tracking method.

## 2 Methodology

The flowchart of the proposed ship detection and tracking method is shown in Fig. 2, which mainly includes two stages, ship detection stage and ship tracking stage.

At the first stage, for given two frames, first, camera motion compensation based on least square image matching is presented. Second, a saliency detection method is presented to construct DMSM based on the differential image. Third, the DMSM is segmented to obtain a binary image, and then moving ships are extracted with the help of ship characteristics and surrounding contrast.

At the second stage, first, important elements such as centroid distance, area ratio, and histogram distance are used to perform ship matching between two frames. And then, ship association and tracking is proposed by using the intermediate frame in every three frames.

### 2.1 Motion Compensation

The video captured by the satellite-mounted camera always includes camera motion. Therefore, given two frames  $t - 1$  and  $t$ , the natural solution is to estimate the transformation parameters and try to compensate the camera motion. Many approaches of affine transformation between two images have been reported in the literature. Most of these methods require a set of matched feature points to estimate the affine transformation parameters. However, the matched feature points are easy to fall on the moving objects which can result in the wrong affine parameters.

Therefore, the least square image matching method is used to compensate motion which does not require matched feature points. Denote frame  $t - 1$  and frame  $t$  as the source frame  $f_S(x, y, t - 1)$  and target frame  $f_T(x, y, t)$ , respectively. To account for intensity variations, the relationship between two frames can be modeled by the following transform:

$$m_7 f_S(x, y, t - 1) + m_8 = f_T(m_1 x + m_2 y + m_5, m_3 x + m_4 y + m_6, t), \quad (1)$$

where  $m_1 \sim m_8$  are the transform parameters. Among them,  $m_1 \sim m_4$  form the  $2 \times 2$  affine matrix,  $m_5$  and  $m_6$  are the translation vector, and  $m_7$  and  $m_8$  embody a change in contrast and brightness. In order to estimate these parameters, the following quadratic error is to be minimized:

$$E = \sum_{x, y \in \Omega} [m_7 f_S(x, y, t - 1) + m_8 - f_T(m_1 x + m_2 y + m_5, m_3 x + m_4 y + m_6, t)]^2. \quad (2)$$

Here, first, to simplify the minimization, the error function of Eq. (2) is derived through a Taylor-series expansion. A more accurate estimate of the actual error function can be determined using a Newton–Raphson style iterative scheme.<sup>24</sup> In particular, on each iteration, the estimated transformation is applied to the source image, and a new transformation is estimated between the newly registered source and target images. Second, in order to adapt to the large displacement between two frames and improve the registration speed, a coarse-to-fine registration scheme is adopted. The details can be found in Ref. 25. After the affine parameters are obtained, it is applied to the source frame to obtain the registered source frame. This method can achieve sub-pixel registration accuracy and adapt to the registration between two frames with illumination changes.

## 2.2 Dynamic Multiscale Saliency Map

If we have given two registered frames obtained at different time, the simplest way to detect motion regions is by frame differencing.<sup>26</sup> However, if the displacement of motion ship between two frames is small, the holes often appear inside motion ships in the differential image. If there are similar intensity values in the entire ship, it will also cause the holes inside motion ship. In all these cases, it is difficult to detect a complete moving ship, as shown in Fig. 3(c).

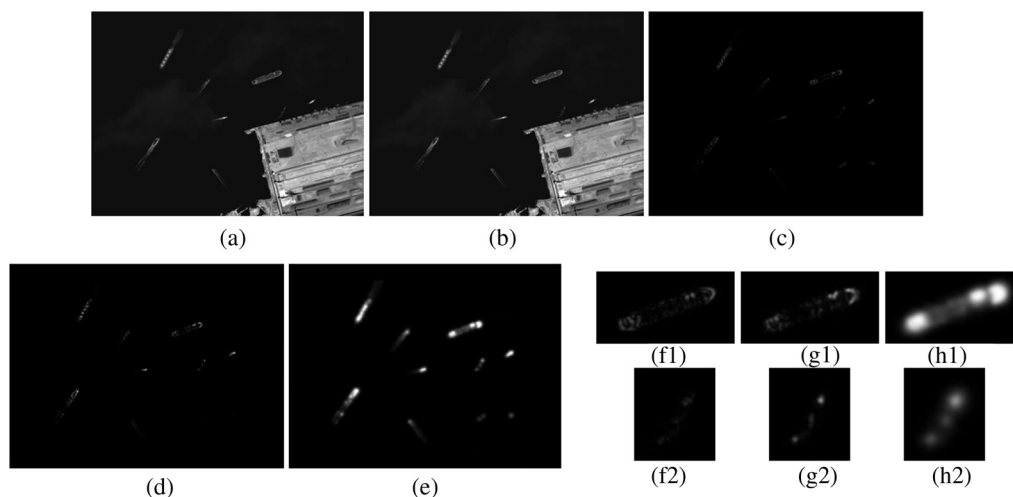
As is known, visual saliency is one of the preattentive processes which makes us to focus our eyes on attractive regions of the scene.<sup>27</sup> Due to the ability to capture the salient region, visual saliency has been widely applied in target detection, which is usually used to segment the salient target.<sup>28,29</sup> Therefore, we take advantage of this technique to highlight ship areas while suppressing the background in the differential image.

However, there may be many moving ships with different sizes and speeds in each frame, so it is difficult to extract a complete ship without holes from the single-scale saliency map of the differential image. As shown in Fig. 3(d), the single-scale saliency map with spectral residual (SR) model<sup>30</sup> cannot eliminate the holes inside motion ship effectively. Therefore, we propose to compute the DMSM of the differential image based on SR model, which is calculated with the following steps:

Step 1: A Gaussian pyramid of the differential image is built with  $n$  scales, expressed as  $\{G_i | i = 1, 2, \dots, n\}$ .

Step 2: Calculate the log amplitude and phase spectrum of image  $G_i$

$$F_i(f) = \mathcal{F}[G_i] \quad L_i(f) = \log[|F_i(f)|] \quad \Phi_i(f) = ph[F_i(f)], \quad (3)$$



**Fig. 3** Example of saliency map of differential image. (a) and (b) The registered source frame and the target frame. (c) The differential image of (a) and (b). (d) The saliency map based on SR model. (e) DMSM of proposed method. To display clearly, two regions possibly including ships are shown in detailed forms. (f1) and (f2) Two zoomed-in regions of (c). (g1) and (g2) Two zoomed-in regions of (d). (h1) and (h2) Two zoomed-in regions of (e).

where  $\mathcal{F}(\cdot)$  is the Fourier transform,  $ph(\cdot)$  is a function for computing phase spectrum,  $\Phi_i$  and  $\|F_i(f)\|$  are phase and amplitude spectra, respectively, and  $L_i(f)$  denotes the log amplitude spectrum.

Step 3: Calculate the SR of  $G_i$ :

$$R_i(f) = L_i(f) - h * L_i(f), \tag{4}$$

where  $h$  is the  $3 \times 3$  averaging filter.

Step 4: Do an inverse Fourier transform for SR by keeping the phase spectrum, and we obtain saliency map of  $G_i$ :

$$S_i = \|\mathcal{F}^{-1}\{\exp[R_i(f) + j\Phi_i(f)]\}\|^2. \tag{5}$$

Step 5: The saliency map  $S_i$  is resized to the size of differential image

$$\tilde{S}_i = \text{resize}(S_i), \tag{6}$$

where the resized saliency map  $\tilde{S}_i$  has the same width and height as that of differential image.

Step 6: DMSM is calculated based on the combination of all resized saliency maps of  $n$  scales with a two-dimensional Gaussian filter  $g$

$$\text{DMSM} = g * \frac{1}{n} \sum_{i=1}^n \hat{S}_i. \tag{7}$$

As shown in Fig. 3(e), regions of moving ships are highlighted in the output DMSM, and the holes inside the motion ship are successfully eliminated compared with the single SR model. As seen clearly from the zoomed-in regions [shown in Figs. 3(f1)–3(h1) and 3(f2)–3(h2)], our proposed method can obtain better performance in eliminating holes, because DMSM is generated under different resolutions just like the human visual system.

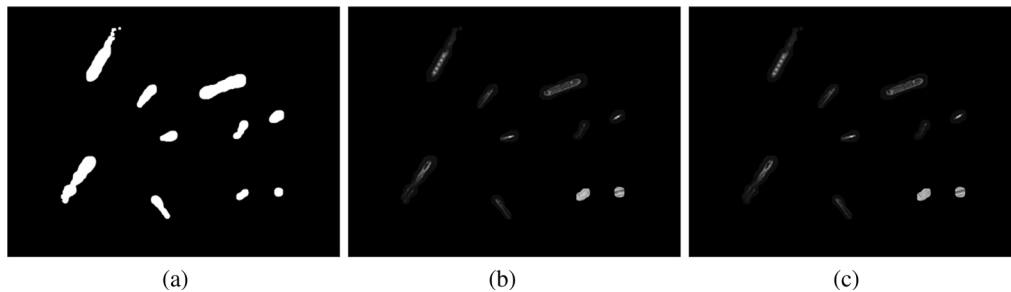
### 2.3 Ship Region Detection

Because the ship regions are relatively salient in DMSM, ship candidates can be obtained by the segmentation of DMSM. Here, a simple segmentation method based on the mean and standard deviation of DMSM is applied to compute an adaptive threshold with the following equation

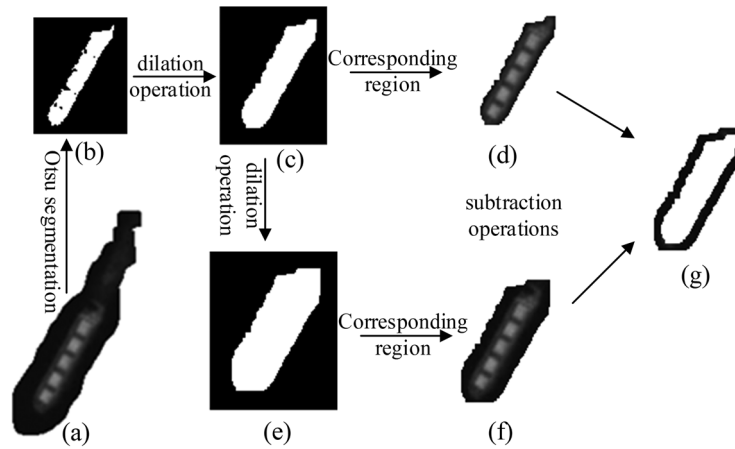
$$T_D = \mu_s + \lambda \cdot \sigma_s, \tag{8}$$

where  $\mu_s$  and  $\sigma_s$  are the mean and standard deviation of DMSM, respectively,  $\lambda$  is a coefficient which was empirically set to 1.0 to 2.0 in our experiments.

Further, mathematical morphology dilation operation is applied to eliminate the remaining holes within regions and a binary image is obtained [Fig. 4(a)]. Then, the candidate ship regions can be obtained by AND operation between a binary image and each frame of the two frames, respectively [Figs. 4(b) and 4(c)].



**Fig. 4** Ship candidate regions in two frames. (a) Regions after morphology operation. (b) The candidate motion regions in registered source frame. (c) The candidate motion regions in target frame.



**Fig. 5** Ship detection from candidate ship region based on surrounding contrast. (a) Candidate ship region. (b) Otsu segmentation result. (c) Dilation operation result of (b). (d) Foreground. (e) Dilation operation result of (c). (f) Foreground and background. (g) Background.

After candidate ship regions obtained, there may exist some false alarms, such as ship wakes, ocean waves, land, and clouds. As shown in Figs. 4(b) and 4(c), we can see that there are two regions without moving ships. Therefore, we need to further eliminate obvious false candidates with the following steps:

- Step 1: For each candidate region [Fig. 5(a)], it is segmented by Otsu segmentation method<sup>31</sup> [result shown in Fig. 5(b)]. Then, to eliminate the small holes possibly existed in the candidate region [Fig. 5(b)], the morphology dilation operation is applied to the segmentation result [result shown in Fig. 5(c)], and the corresponding moving ship is shown in Fig. 5(d).
- Step 2: Ships always have a limited area, length, and width range. According to these constraints, false candidate regions, such as very large or very small islands and clouds, can be eliminated with proper thresholds. Furthermore, ships are commonly long and thin. Therefore, the ratio of the length to the width is larger than a given threshold. According to this condition, obvious false alarms, including islands and clouds with very small ratios, are eliminated.<sup>13</sup>
- Step 3: To obtain more background region, the morphology dilation operation is further applied to Fig. 5(c) [result shown in Fig. 5(e)], and the corresponding region is shown in Fig. 5(f), which includes both the background and foreground. Figure 5(g) is background region after subtraction operations.
- Step 4: Because the intensity of the sea surface in the image is quite different from that of the ship, the surrounding contrast between sea and ship would be helpful to eliminate obvious false candidates. Based on the above characteristic, we regard ships which are not satisfied with the following conditions as obvious false alarms:

$$\mu_{FG} > \mu_{FG+BG} > \mu_{BG} \quad \sigma_{FG} > \sigma_{FG+BG} > \sigma_{BG}, \quad \sigma_{FG} > \gamma\sigma_{BG}, \quad (9)$$

where  $\mu_{FG}$  and  $\sigma_{FG}$  are the mean and standard deviation of the foreground region,  $\mu_{BG}$  and  $\sigma_{BG}$  are the mean and standard deviation of the background region, and  $\mu_{FG+BG}$  and  $\sigma_{FG+BG}$  are the mean and standard deviation of the foreground and background region, respectively,  $\gamma$  is empirically set to 1.5 to 2.0 in our experiments.

After false alarm elimination, the resulting regions are returned as moving ships.

## 2.4 Ship Matching between Two Frames

Moving ships are detected after false alarm elimination, respectively, in two frames, as previously discussed. Suppose the moving ships in the registered source frame are expressed as  $\{S_{1,t}, \dots, S_{m,t}\}$ , and in the target frame are expressed as  $\{S_{1,t-1}, \dots, S_{n,t-1}\}$ . The performance of ships matching is determined by Algorithm 1.

**Algorithm 1** Ships matching between two frames.

---

**Input:**  $\{S_{1,t}, \dots, S_{m,t}\}$  and  $\{S_{1,t-1}, \dots, S_{n,t-1}\}$ .

$T_d$ : Threshold of the distance between ships.

$T_\psi$ : Threshold of the matching ships.

**Output:** Matched ship pairs.

**for**  $i = 1$  to  $m$  do

**for**  $j = 1$  to  $n$  do

**Step 1.** Compute the centroid distance  $Dist_{(t,t-1)}^{i,j}$  between the  $i$ 'th ship  $S_{i,t}$  and the  $j$ 'th ship  $S_{j,t-1}$

$$Dist_{(t,t-1)}^{i,j} = \sqrt{(x_t^i - x_{t-1}^j)^2 + (y_t^i - y_{t-1}^j)^2}, \quad (10)$$

where  $(x_t^i, y_t^i)$  and  $(x_{t-1}^j, y_{t-1}^j)$  are the coordinates of centroids.

**Step 2.** If  $Dist_{(t,t-1)}^{i,j} < T_d$

a. Compute the area ratio  $AR(i, j)$  between the  $i$ 'th ship  $S_{i,t}$  and the  $j$ 'th ship  $S_{j,t-1}$ :

$$AR(i, j) = \frac{\min[\text{area}(S_{i,t}), \text{area}(S_{j,t-1})]}{\max[\text{area}(S_{i,t}), \text{area}(S_{j,t-1})]}, \quad (11)$$

where  $\text{area}(\bullet)$  is the area, that is, the pixels number. Area of  $S_{j,t-1}$  and  $S_{i,t}$  should be similar, even though the ship has been moved or rotated.

b. Compute the histograms of  $S_{i,t}$  and  $S_{j,t-1}$  with  $H$ -bin,  $p^i = \{p_k^i\}_{k=1,2,\dots,H}$  and  $q^j = \{q_k^j\}_{k=1,2,\dots,H}$ .

c. The Bhattacharya coefficient is

$$\rho(p^i, q^j) = \sum_{k=1}^H \sqrt{p_k^i q_k^j}. \quad (12)$$

Bhattacharya distance<sup>32</sup> between two distributions is defined as

$$BD(i, j) = \sqrt{1 - \rho(p^i, q^j)}. \quad (13)$$

d. The total metric is finally defined by the two metrics

$$\psi_{(t,t-1)}^{i,j} = \delta[1 - BD(i, j)] + (1 - \delta)AR(i, j) \quad (14)$$

where  $\delta$  is a weighting coefficient.

e. if  $\psi_{(t,t-1)}^{i,j} > T_\psi$ ,  $\{S_{i,t}, S_{j,t-1}\}$  is one pair of candidate matching ships.

**end for**  $j$

A high total metric value indicates a good matching with the target ship. If  $S_{i,t}$  have several candidate matching ships satisfied the above conditions, we take the ship pair with the highest total metric value as the matched ships.

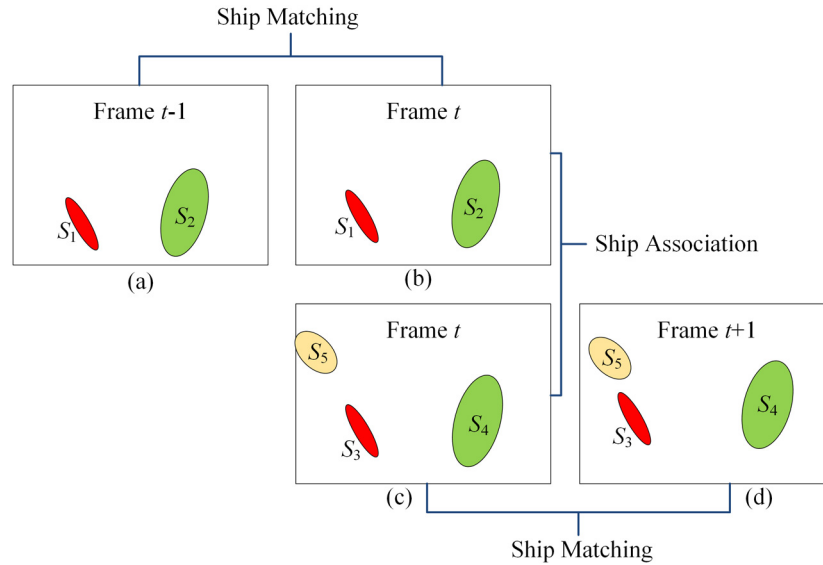
**end for**  $i$

---

## 2.5 Ship Tracking

From the above sections, the detected and matched moving ships are obtained from every two frames. Suppose two pairs of matched moving ships  $S_1$  and  $S_2$  have been obtained based on frame  $t - 1$  and frame  $t$  [Figs. 6(a) and 6(b)], and three pairs of matched moving ships  $S_3$ ,





**Fig. 6** An example of ship association and tracking in three frames. (a) Two ships detected from frame  $t - 1$  based on frames  $t - 1$  and  $t$ . (b) Two moving ships detected from frame  $t$  based on frames  $t - 1$  and  $t$ . (c) Three moving ships detected from frame  $t$  based on frames  $t - 1$  and  $t$ . (d) Three moving ships detected from frame  $t + 1$  based on frames  $t - 1$  and  $t$ .

$S_4$ , and  $S_5$  have been obtained based on frames  $t$  and  $t + 1$  [Figs. 6(c) and 6(d)]. Then, all the moving ships in frames  $t - 1$ ,  $t$ , and  $t + 1$  are associated using the intermediate frame  $t$ , which can realize ship tracking and give the ship trajectory in the satellite video.

Let  $S_{i,t}^{(t-1,t)}$  denote the ship- $i$  in frame  $t$  which is detected based on frames  $t - 1$  and  $t$ , and  $S_{j,t}^{(t,t+1)}$  denote the ship- $j$  in frame  $t$  which is detected based on frames  $t$  and  $t + 1$ . We define the overlap score  $R_{i,j}$  between the two ships as

$$R_{i,j} = \frac{\text{area}[S_{i,t}^{(t-1,t)} \cap S_{j,t}^{(t,t+1)}]}{\text{area}[S_{i,t}^{(t-1,t)} \cup S_{j,t}^{(t,t+1)}]}, \quad R_{i,j} \in [0,1], \quad (15)$$

where  $S_{i,t}^{(t-1,t)} \cap S_{j,t}^{(t,t+1)}$  and  $S_{i,t}^{(t-1,t)} \cup S_{j,t}^{(t,t+1)}$  are the intersection and the union of two ships, respectively, and  $\text{area}(\bullet)$  denotes the area.

If  $R_{i,j}$  is higher than a set threshold  $T_R$ ,  $S_{i,t}^{(t-1,t)}$  and  $S_{j,t}^{(t,t+1)}$  are regarded as the same ship, through which the ship association and tracking in three frames is realized. As shown in Figs. 6(b) and 6(c),  $S_1$  and  $S_3$ ,  $S_2$  and  $S_4$  are two pairs of associated ships, through which ship tracking can be realized.

The proposed algorithm resolves some problems in multiship tracking, such as the appearance of new moving ships and the disappearance of old moving ships. As shown in Fig. 6, it can be seen that a new moving ship  $S_5$  is detected and tracked based on frames  $t$  and  $t + 1$ . Furthermore, the proposed method avoids the registration problem of all frames to the same reference frame for moving ship detection and tracking.

### 3 Experiments Results

Considering the lack of benchmarking dataset of satellite video including moving ships, we use the sequences of Geostationary Orbit Space Surveillance System (GO3S) satellite video and our synthetic satellite video to evaluate the performance of the proposed method. All the experiments in the following are conducted on a desktop PC with a 1.40-GHz CPU and 4-GB memory, and our code are written in Microsoft Visual Studio 2013 with C++ and OpenCV library.



**Fig. 7** The frame 1871 in the GO3S satellite video, and the area of the red rectangle is image frame of Video 1.

### 3.1 Data Set Description

The first type of satellite image sequences is cut out from the image sequences of geostationary GO3S satellite and include moving ships from the 1871st frame (see Fig. 7). Video 1 is cut out from the frame 1871 (see Fig. 7) of synthetic GO3S satellite video.<sup>33</sup> The cutting frame size is  $650 \times 360$  pixels containing different sizes of ships, and the original frame rate is 10 frames per second (fps). However, because some ships may move slowly, if the video has a high frame rate, the change in the differential image is not obvious. Therefore, we reduce the frame rate to 5 fps to make the differential image change obvious, and the used video consists of 19 frames.

In Video 2, due to the lack of benchmarking datasets of satellite video including moving ships, we further evaluate our proposed algorithm with our synthetic satellite video including large and small ships. The used video consists of 19 frames, and the frame size is  $1024 \times 768$  pixels containing eight different sizes of ships (Fig. 9).

### 3.2 Parameter Selection

In this section, several parameters used in our method during detection and tracking are presented. First, to improve the processing speed, in motion compensation, we decompose the two each frame into five layers of pyramid, and each layer of pyramid image only needs three iterations to get better results. Second, in DMSM detection, we set pyramid scales  $n = 3$ , and the Gaussian kernel size of filter,  $g$ ,  $15 \times 15$  pixels. Third, in ship region detection, the coefficient,  $\lambda$ , is set to 1.0 for the two image sequences. The structuring element in all the morphology dilation operation is  $3 \times 3$ . Following, for ship matching, we set the three parameters during the ship matching process, i.e.,  $T_d = 70$ , threshold of the distance between ships,  $T_\psi = 0.7$ , threshold of the matching ships, and  $\delta = 0.5$ , the weighting coefficient. Finally, in ship associate and tracking, we set  $T_R$  to 70%. We employ Recall and Precision to evaluate the performance, which are defined as

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (16)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (17)$$

where TP (true positive) is a moving ship detected and tracked that turned out to be an actually moving ship, FN (false negative) is an actually moving ship but not detected and tracked, and FP (false positive) is a target detected and tracked that is not a moving ship.

In addition, we apply the widely used spatial overlap to measure whether a bounding box is true positive or false positive, and the threshold of the spatial overlap is set to 0.6. The spatial overlap is calculated by the following equation:<sup>34</sup>

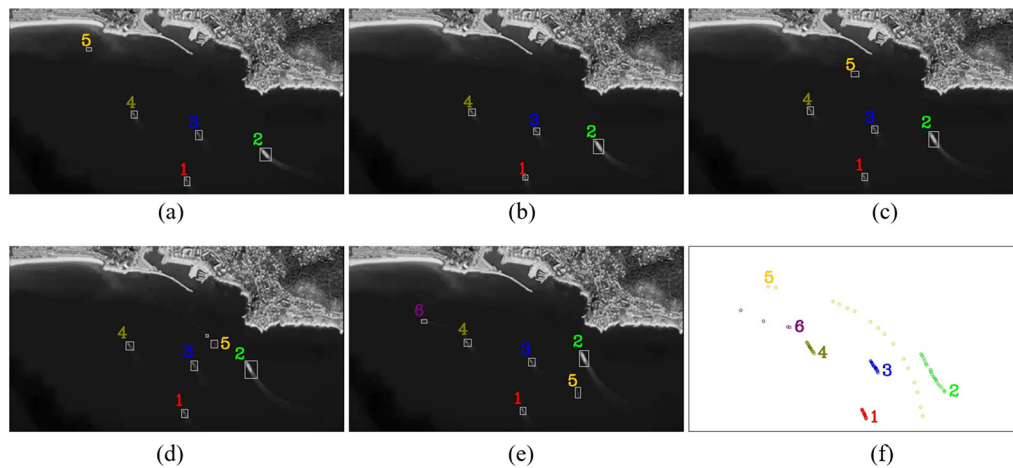
$$A(GT_{ik}, DT_{jk}) = \frac{\text{area}(GT_{ik} \cap DT_{jk})}{\text{area}(GT_{ik} \cup DT_{jk})}, \quad (18)$$

where  $\text{area}(GT_{ik} \cap DT_{jk})$  is denoted as the intersection of the ground truth bounding box  $GT_i$  and the detected bounding box  $DT_j$  in frame  $k$ , and  $\text{area}(GT_{ik} \cup DT_{jk})$  is denoted as the union of  $GT_{ik}$  and  $DT_{jk}$ .

### 3.3 Effectiveness of Our Method

**Example 1:** In Video 1 [Figs. 8(a)–8(e)], the ships are so small that they only take up dozens of pixels. Furthermore, the ships have similar shape and intensity values, and even are covered with thin clouds. All these factors introduce more difficulties for our detection and tracking task. Figure 8(a) shows the results of the proposed method for frame 2. We can observe that there are five moving ships, including a smaller ship or yacht (ship 5) that was leaving the coast and covered by thin clouds. However, this ship has been tracked failure in frame 3, frame 4 [see Fig. 8(b)], and frame 5, because the contrast of this ship has the similar contrast of thin cloud. Due to the ship 5 with very fast velocity, the ship wave is misjudged as the ship [see Fig. 8(d)]. As shown in Fig. 8(e), a yacht (the ship-6) is released from ship 4, fortunately, our algorithm can detect and track the new-emerging ship (ship 6).

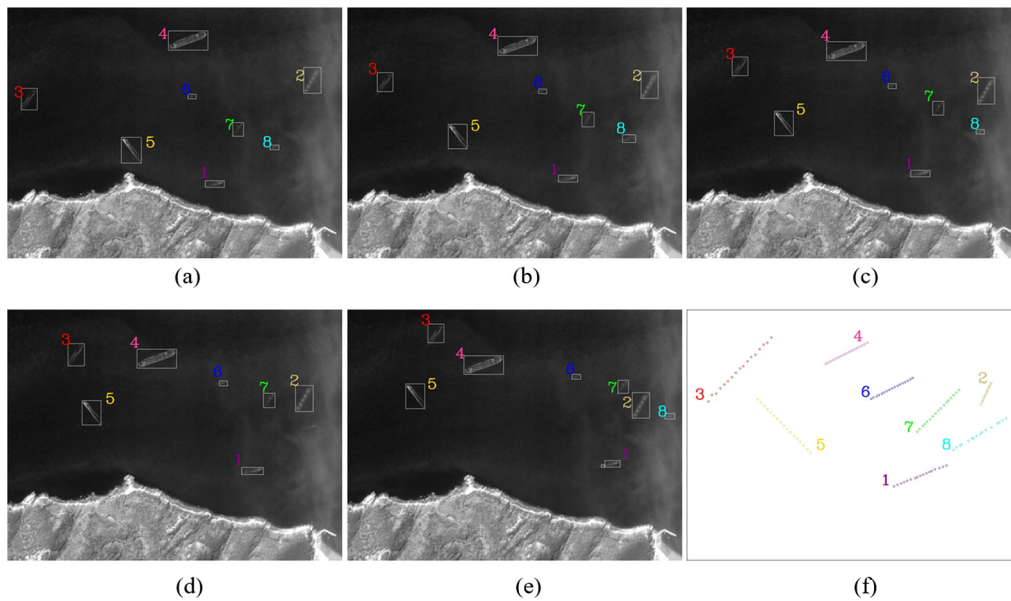
The estimated trajectory of the ships movement based on the proposed ship tracking algorithm is shown in Fig. 8(f). It can be seen that ship 1 to ship 6 have been tracked very well, except three false alarms are present in ship 5 and one false alarm in ship 6. The recall and precision are listed in Table 1 which are calculated with the total 19 frames.



**Fig. 8** Results of ship detection and tracking in the GO3S satellite video. (a)–(e) Frames 2, 6, 10, 14, 18, respectively. (f) The trajectory of moving ships.

**Table 1** Overall performance of the proposed algorithm for Video 1.

No. of ship	Recall	Precision
1	1.0	1.0
2	1.0	1.0
3	1.0	1.0
4	1.0	1.0
5	0.79	0.83
6	0.8	1.0



**Fig. 9** Results of ship detection and tracking in the synthetic satellite video. (a)–(e) Frames 2, 6, 10, 14, 19, respectively. (f) The trajectory of moving ships.

**Table 2** Overall performance of the proposed algorithm for Video 2.

No. of ship	Recall	Precision
1	1.0	0.94
2	1.0	1.0
3	1.0	1.0
4	1.0	1.0
5	1.0	1.0
6	1.0	1.0
7	1.0	1.0
8	0.89	1.0

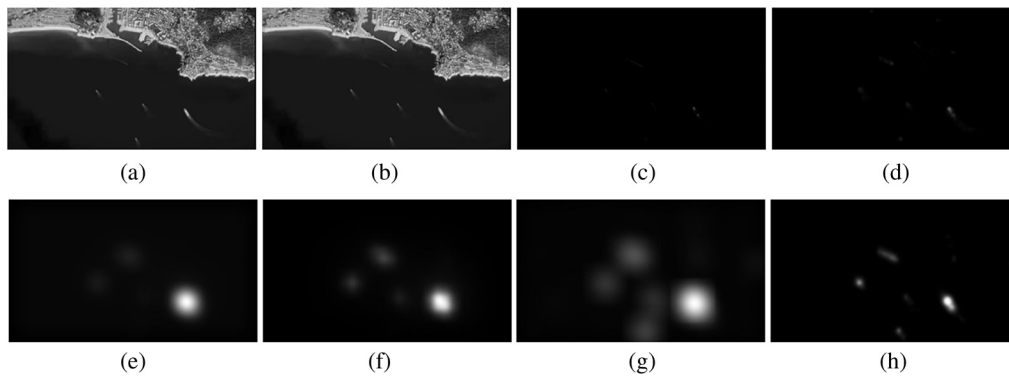
**Example 2:** Video 2 is covered by a thin cloud, especially in the right side of the image. These frames are shown in Figs. 9(a)–9(e) including eight moving ships. In frame 14 [Fig. 9(d)], one smaller ship (ship 8) has been tracked failure which is disappeared in the thin cloud, because the contrast of this ship has the similar contrast of the thin cloud. In frame- 19 [Fig. 9(e)], the wave of ship 1 is mistakenly detected as a moving ship (enclosed with a very small rectangle).

The estimated trajectory of the ships movement based on the proposed method is shown in Fig. 9(f). It can be seen that ship 1 to ship 8 have been tracked very well, except ship 8 is not tracked successfully in two frames, and there is a false alarm near ship 1 in frame 19. The recall and precision are listed in Table 2 which are calculated with the total 19 frames.

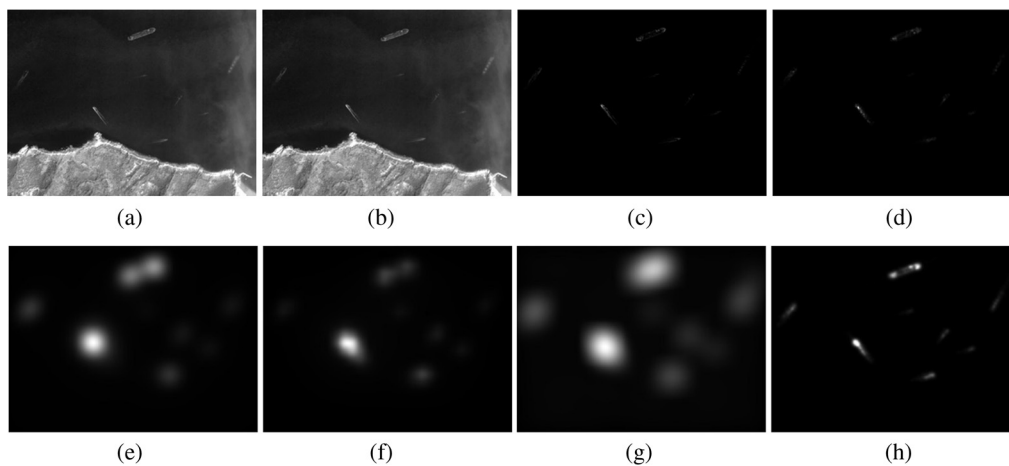
### 3.4 Comparison with Other Methods

#### 3.4.1 Saliency map

In this section, we conducted several experiments to compare our DMSM method with several saliency map methods (SR,<sup>30</sup> Itti,<sup>27</sup> GBVS,<sup>35</sup> and Signature algorithm<sup>36</sup>). Two examples in Figs. 10 and 11 validate that the saliency map obtained by our approach performs better than



**Fig. 10** The saliency map of differential image in Video 1, (AVI, 113 KB) [URL: <https://doi.org/10.1117/1.JRS.13.026511.1>]. (a) and (b) The compensated frame and the current frame. (c) Differential image of (a) and (b). (d) Saliency map for SR. (e) Saliency map for Itti. (f) Saliency map for GBVS. (g) Saliency map for signature algorithm. (h) DMSM.



**Fig. 11** The saliency map of difference image in Video 2, (AVI, 1,953 KB) [URL: <https://doi.org/10.1117/1.JRS.13.026511.2>]. (a) and (b) The compensated frame and the current frame. (c) Different image of (a) and (b). (d) Saliency map of SR. (e) Saliency map for Itti. (f) Saliency map for GBVS. (g) Saliency map for signature algorithm. (h) DMSM.

those obtained by other methods. As shown from the difference image [Figs. 10(c) and 11(c)], due to the ship displacement between two frames is small, and there are holes inside motion ships. In Figs. 10(d) and 11(d), the SR model cannot detect the moving ships without holes clearly. In Figs. 10(e)–10(f) and 11(e)–11(f), the Itti model and GBVS model cannot detect all the moving ships. In Figs. 10(g) and 11(g), the signature algorithm can highlight the moving ships; however, the true ship regions are enlarged too much, which causes ships close enough to mix up easily. Contrarily, our DMSM highlights the regions with moving ships, and these regions are not enlarged too much, as in Figs. 10(h) and 11(h).

### 3.4.2 Ship detection

To test the effectiveness of the proposed method, we compare it with the recent method of R2CNN\_head<sup>10</sup> and several background subtraction methods<sup>37</sup> using BGSLibrary: available at GitHub repository: <https://github.com/andrewsobral/bgslibrary>. Here, for R2CNN\_head method, we use the pretraining model “ResNet\_v1\_101.ckpt,” available at GitHub repository: [https://github.com/yangxue0827/R2CNN\\_HEAD\\_FPN\\_Tensorflow](https://github.com/yangxue0827/R2CNN_HEAD_FPN_Tensorflow), to initialize our network. For our dataset with different size of ships, we further train the model with a total of 10k iterations. The results for Video 1 and Video 2 are shown in Tables 3 and 4, respectively. As can be seen, the precision of the proposed method can reach 96% and 98.6%, and the recall of the

**Table 3** Comparison results of ship detection for Video 1.

Method	TP+FN	TP+FP	TP	Recall (%)	Precision (%)
Our method	100	96	93	93	96
R2CNN_head <sup>10</sup>	100	86	84	84	97.6
AdaptiveBackgroundLearning	100	155	91	91	58.7
FuzzyChoquetIntegral <sup>38</sup>	100	150	74	74	49.3
DPWrenGABS <sup>39</sup>	100	147	91	91	61.9
MixtureOfGaussianV1BGS <sup>40</sup>	100	39	35	35	89.74
MultiLayerBGS <sup>41</sup>	100	108	94	94	87.03
PixelBasedAdaptiveSegmenter <sup>42</sup>	100	32	28	28	87.5

**Table 4** Comparison results of ship detection for Video 2.

Method	TP+FN	TP+FP	TP	Recall (%)	Precision (%)
Our method	152	153	151	99.3	98.6
R2CNN_head	152	113	113	74.3	100
AdaptiveBackgroundLearning	152	202	150	98.68	74.25
FuzzyChoquetIntegral	152	145	99	65.13	68.27
DPWrenGABS	152	218	149	98.02	68.3
MixtureOfGaussianV1BGS	152	53	35	23.02	66.03
MultiLayerBGS	152	156	144	94.73	92.3
PixelBasedAdaptiveSegmenter	152	109	74	48.7	67.9

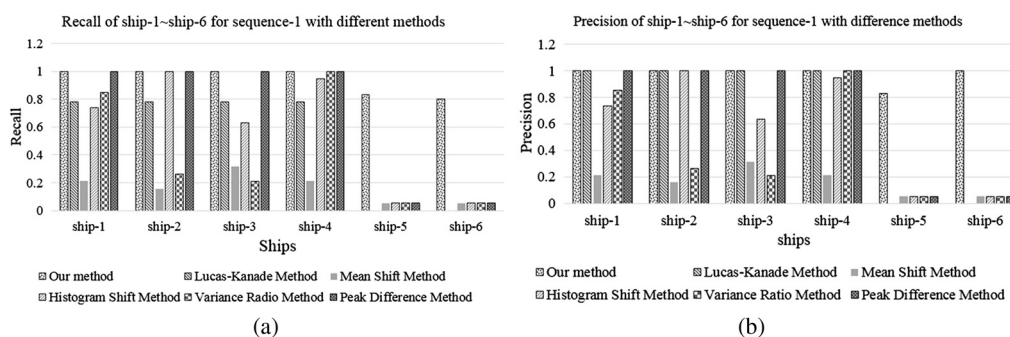
proposed method can reach 93% and 99.3% for Video 1 and Video 2 respectively, which indicates that the proposed method is of high accuracy. The precision of R2CNN\_head method can reach 97.6% and 100%, while the recall of this method is 75% and 64.5% for Video 1 and Video 2 with small ships, respectively.

### 3.4.3 Ship tracking

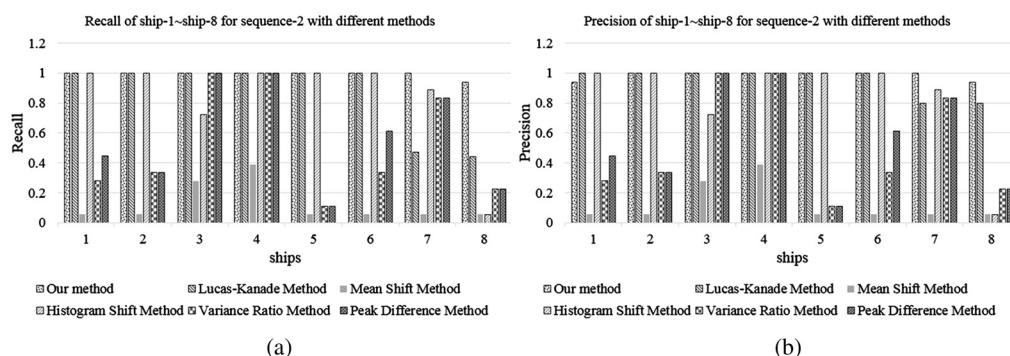
We compared our method with classical Lucas–Kanade tracker,<sup>43</sup> the other several tracking methods in VIVID Tracking Evaluation Testbed V3.0<sup>44</sup> (basic mean shift,<sup>34</sup> histogram shift, variance ratio,<sup>45</sup> and peak difference feature shift).

For Lucas–Kanade tracker,<sup>24</sup> corners were first detected in one frame and tracked in the other frame. For the other tracking methods in VIVID Tracking Evaluation Testbed V3.0, each moving ship was manually detected in the first frame, and tracked in the other frames using these tracking methods, and we compute Recall and Precision of this ship.

Figure 12 shows the tracking results of the six ships in Video 1. As can be observed, our method has high recall and precision at ship tracking than other five methods. Moreover, it has strong robustness to ships of different sizes and can provide a feasible way for ship detection and tracking in satellite video. As Fig. 12 shows, our proposed method performs the best among the tested methods even for ship 5 and ship 6 with very little size. Figure 13 shows the tracking results of eight ships in Video 2. Our method has high recall and precision at ship tracking than other methods. As shown in Fig. 13, our proposed method performs the best among the tested methods for ships of different sizes, which can provide a feasible way for ship detection and tracking.



**Fig. 12** Recall and precision of six ships with different methods in Video 1. (a) Recall. (b) Precision.



**Fig. 13** Recall and precision of eight ships with different methods in Video 2. (a) Recall. (b) Precision.

As shown in Figs. 12 and 13, the Lucas–Kanade method fails to track small ships, such as ship 5 and ship 6 in Video 1 and ship 7 and ship 8 in Video 2. The mean shift method gets the poor tracking results for ships of different sizes. The histogram shift method and the peak difference method are good for tracking large ships, but it is very poor for tracking small ships. However, the variance ratio method is tracking instability whether tracking large ships or small ships.

## 4 Conclusion

This paper provides a method of ship detection and tracking from satellite video. In the ship detection stage, the motion compensation between two adjacent frames is required to make the background stable. After that, the foreground can be extracted from the background based on DMSM of differential images. Then, a moving ship is detected based on the analysis of the surrounding contrast and ship characteristics. In ship tracking stage, the moving ships are matched based on the combination of centroid distance, area ratio, and histogram distance of ships between every two frames. Finally, the ship tracking is realized based on a ship association scheme. Our method has been tested using a set of satellite videos with different size of ships. The ships have been successfully detected and tracked, and the performance is analyzed by the calculation of recall and precision.

## Acknowledgments

The authors would like to thank the editors and anonymous reviews for their valuable comments and helpful suggestions, which greatly improved the paper’s quality. This work was supported by the National Natural Science Foundation of China under Grant Nos. 61773383 and 61702520.

## References

1. S. P. Zhang, Z. H. Qi, and D. L. Zhang, "Ship tracking using background subtraction and inter-frame correlation," in *Int. Cong. Image Sig. Process.*, pp. 1–4 (2009).
2. S. Fefilyat'ev, D. Goldgof, and C. Lembke, "Tracking ships from fast moving camera through image registration," in *Proc. IEEE Conf. Pattern Recognit.*, pp. 3500–3503 (2010).
3. J. W. Wu et al., "Ship target detection and tracking in cluttered infrared imagery," *Opt. Eng.* **50**(5), 057207 (2011).
4. S. X. Qi et al., "Low-resolution ship detection from high-altitude aerial images," *Proc. SPIE* **10608**, 1060805 (2018).
5. W. Liu et al., "Inshore ship detection with high-resolution SAR data using saliency map and kernel density," *Proc. SPIE* **10033**, 100333V (2016).
6. X. F. Wei, X. Q. Wang, and J. S. Chong, "Local region power spectrum-based unfocused ship detection method in synthetic aperture radar images," *J. Appl. Remote Sens.* **12**(1), 016026 (2018).
7. Q. P. Wang et al., "Inshore ship detection using high-resolution synthetic aperture radar images based on maximally stable extremal region," *J. Appl. Remote Sens.* **9**(1), 095094 (2015).
8. S. Tian, C. Wang, and H. Zhang, "Ship detection method for single-polarization synthetic aperture radar imagery based on target enhancement and nonparametric clutter estimation," *J. Appl. Remote Sens.* **9**(1), 096073 (2015).
9. F. Yang, Q. Xu, and B. Li, "Ship detection from optical satellite images based on saliency segmentation and structure-LBP feature," *IEEE Geosci. Remote Sens. Lett.* **14**(5), 602–606 (2017).
10. X. Yang et al., "Position detection and direction prediction for arbitrary-oriented ships via multitask rotation region convolutional neural network," *IEEE Access.* **6**, 50839–50849 (2018).
11. G. Yang et al., "Ship detection from optical satellite images based on sea surface analysis," *IEEE Trans. Geosci. Remote Sens. Lett.* **11**(3), 641–645 (2014).
12. C. H. Deng et al., "Ship detection from optical satellite image using optical flow and saliency," *Proc. SPIE* **8921**, 89210F (2013).
13. Y. Yao et al., "Ship detection in optical remote sensing images based on deep convolutional neural networks," *J. Appl. Remote Sens.* **11**(4), 042611 (2017).
14. J. Tang et al., "Compressed-domain ship detection on spaceborne optical image using deep neural network and extreme learning machine," *IEEE Trans. Geosci. Remote Sens.* **53**(3), 1174–1185 (2015).
15. Z. Shi et al., "Ship detection in high-resolution optical imagery based on anomaly detector and local shape feature," *IEEE Trans. Geosci. Remote Sens.* **52**(8), 4511–4523 (2014).
16. Z. Zou and Z. Shi, "Ship detection in spaceborne optical image with SVD networks," *IEEE Trans. Geosci. Remote Sens.* **54**(10), 5832–5845 (2016).
17. N. Proia and V. Pagé, "Characterization of a Bayesian ship detection method in optical satellite images," *IEEE Geosci. Remote Sens. Lett.* **7**(2), 226–230 (2010).
18. G. Kopsiaftis and K. Karantza, "Vehicle detection and traffic density monitoring from very high resolution satellite video data," in *IEEE Int. Geosci. and Remote Sens. Symp. (IGARSS)*, pp. 1881–1884 (2015).
19. T. Yang et al., "Small moving vehicle detection in a satellite video of an urban area," *Sensors* **16**(9), 1528 (2016).
20. S. Larsen, H. Koren, and R. Solberg, "Traffic monitoring using very high resolution satellite imagery," *Photogramm. Eng. Remote Sens.* **75**(7), 859–869 (2009).
21. N. S. Rao et al., "Estimation of ship velocities from MODIS and OCM," *IEEE Trans. Geosci. Remote Sens. Lett.* **2**(4), 437–439 (2005).
22. L. B. Yao et al., "A novel ship-tracking method for GF-4 satellite sequential images," *Sensors* **18**, 2007 (2018).
23. Z. X. Zhang et al., "Application potential of GF-4 images for dynamic ship monitoring," *IEEE Trans. Geosci. Remote Sens. Lett.* **14**(6), 911–915 (2017).
24. J. Shi and C. Tomasi, "Good features to track," in *IEEE Conf. Comput. Vision and Pattern Recognit. (CVPR)*, pp. 593–600 (1994).



25. S. Periaswamy and H. Farid, "Elastic registration in the presence of intensity variations," *IEEE Trans. Med. Imaging* **22**(7), 865–874 (2003).
26. R. Jain, "Dynamic scene analysis using pixel-based processes," *IEEE Comput.* **14**(8), 12–18 (1981).
27. L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(11), 1254–1259 (1988).
28. Z. Liang, L. Yu, and S. Yi, "Inshore ship detection via saliency and context information in high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens. Lett.* **13**(12), 1870–1874 (2016).
29. S. Wang et al., "New hierarchical saliency filtering for fast ship detection in high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.* **55**(1), 351–362 (2017).
30. X. D. Hou and L. Q. Zhang, "Saliency detection: a spectral residual approach," in *IEEE Conf. Comput. Vision and Pattern Recognit. (CVPR)*, pp. 1–8 (2007).
31. N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst. Man Cybern.* **9**(1), 62–66 (1979).
32. D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(5), 564–577 (2003).
33. Permanent Surveillance from Space, <https://www.military.com/video/space-technology/artificial-satellites/permanent-surveillance-from-space/5299910843001> (10 August 2018).
34. F. Yin, D. Markris, and S. Velastin, "Performance evaluation of object tracking algorithms," in *Proc. IEEE Int. Workshop Perform. Eval. Tracking Surveill.*, pp. 1–8 (2007).
35. J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proc. Adv. Neural Inf. Process. Syst.*, pp. 681–688 (2007).
36. X. Hou, J. Harel, and C. Koch, "Image signature: highlighting sparse salient regions," *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(1), 194–201 (2012).
37. A. Sobral and A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," *Comput. Vision Image Understanding* **122**, 4–21 (2014).
38. F. El Baf, T. Bouwmans, and B. Vachon, "Fuzzy integral for moving object detection," in *IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, pp. 1729–1736 (2008).
39. C. Wren et al., "Pfinder: real-time tracking of the human body," *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 780–785 (1997).
40. P. Kaewtrakulpong and R. Bowden, "An improved adaptive background mixture model for realtime tracking with shadow detection," in *Eur. Workshop Adv. Video Based Surveill. Syst. (AVSS)* (2001).
41. J. Yao and J.-M. Odobez, "Multi-layer background subtraction based on color and texture," in *IEEE Conf. Comput. Vision and Pattern Recognit. (CVPR)* (2007).
42. M. Hofmann, P. Tiefenbacher, and G. Rigoll, "Background segmentation with feedback: the pixel-based adaptive segmenter," in *IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recognit. Workshops (CVPRW)*, pp. 38–43 (2012).
43. J. Y. Bouguet, "Pyramidal implementation of the Lucas Kanade feature tracker," Technical Report, Intel Corporation, Microsoft Research Labs (2000).
44. R. T. Collins, X. H. Zhou, and S. K. Teh, "An open source tracking testbed and evaluation web site," in *IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance (PETS)* (2005).
45. R. Collins, Y. Liu, and M. Leordeanu, "On-line selection of discriminative tracking features," *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(10), 1631–1643 (2005).

**Haichao Li** received his BS and MS degrees in mechanical and electrical engineering from Beijing University of Chemical Technology, Beijing, China, in 2001 and 2004, respectively. And he received his PhD in the School of Instrument Science and Opto-Electronics Engineering, Beijing University of Aeronautics and Astronautics, Beijing, China, in 2008. He is currently an associate professor at Qian Xuesen Laboratory of Space Technology, China Aerospace Science and Technology Corporation. His research interests include motion detection and tracking, deep learning, and stereo vision.

**Liang Chen** received his BS and MS degrees in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2003 and 2006, respectively. And he received his PhD in the State Key Laboratory of Remote Sensing Science, Chinese Academy of Science, Beijing, China, in 2009. He is currently a senior engineer at Qian Xuesen Laboratory of Space Technology, China Aero-space Science and Technology Corporation. His research interests include remote sensing target tracking, data mining, and parameter retrieval.

**Feng Li** (M'07–SM'10) received his BSEE degree from the Lanzhou Railway University in 1999, his MEng degree from China Academy of Space Technology in 2002, and his PhD in electrical engineering from The University of New South Wales in 2009. Following several years working on astronomical image processing in CSIRO, Australia, and on remote sensing image processing in Chinese Academy of Science, respectively, he is currently a PI in the Qian Xuesen Laboratory of Space Technology. His research interests include image registration, super resolution and compressive sensing, moving detection.

**Meiyu Huang** received her BS degree in computer science and technology from Huazhong University of Science and Technology, Wuhan, China, in 2010, and her PhD degree in computer application technology from the University of Chinese Academy of Sciences, Beijing, China, in 2016. She is currently an assistant researcher in the Qian Xuesen Laboratory of Space Technology, China Academy of Space Technology, Beijing, China. Her research interests include machine learning, ubiquitous computing, human-computer interaction, computer vision and image processing.