

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.Doi Number

An Improved Correlation Filtering Method for Tracking Maritime Small Targets of GF-4 Staring Satellite Sequence Images

Deyang Zhang¹, Yao Xu^{2,3}, and Haimin Hu^{1,*}

¹Zhengzhou Campus, PLA Army Academy of Artillery and Air Defense, Zhengzhou 450052, China

²PLA Army Academy of Artillery and Air Defense, Hefei 230031, China

³Anhui Key Laboratory of Polarization Imaging and Detection, Hefei 230031, China

Corresponding author: Haimin Hu (e-mail: hmhu@163.com).

ABSTRACT Continuous, accurate and real-time earth observation plays an increasingly important role in battlefield situation awareness, environmental monitoring and many other fields. In this paper, GF-4 staring satellite's sequence images are taken as the research objects, and its features such as "short imaging interval, long image sequence and high resolution" are used to solve the tracking problem of large and medium ships at sea. Using the target information, the dark-channel background modeling method and multi-scale Retinex illumination algorithm are improved to achieve target enhancement. The multi-scale features of the target based on time and space information are extracted, and the target tracking is realized by improved kernel correlation filtering algorithm. By comprehensively utilizing unscented Kalman filter model and target significance features, tracking results are optimized to improve tracking accuracy. The experimental results show that the proposed algorithm can track dim and small targets in the staring satellite sequence images effectively, and it has reference significance for the application of similar high-precision satellites.

INDEX TERMS staring satellite; sequential image; target tracking; correlation filtering

I. INTRODUCTION

The Explorer-6 satellite launched by the United States took photos of the Earth from space for the first time in 1958, opened a new chapter in the field of satellite remote sensing. Since then, remote sensing technique has been regarded as an important direction for future development by advanced countries in the world. China's GF-4 satellite is one of the most advanced geosynchronous orbit imaging satellites in the world. The GF-4 satellite was officially put into service in June 2016. Due to the advantages of fast response speed, high resolution and wide imaging range, it provided a large number of timely and accurate observation data for earthquake prevention and disaster reduction, forestry monitoring, meteorological early warning and other fields. In addition, GF-4 satellite has shown great potential in the field of maritime ship monitoring, with the characteristics of high time resolution, high spatial resolution, wide imaging range and fast response, which meets the requirements of "long-term, wide area and high frequency" of maritime monitoring. It is of great significance for verifying the

concept of "one-star for multi-purpose, multi-star networking and multi-network collaboration", and can also provide important reference and guidance for the application research of GF-13 staring satellite, which was successfully launched on October 12, 2020, with higher spatial resolution. However, the size and energy of maritime targets traveling at sea in GF-4 satellite sequence images are small and weak due to the factors such as image resolution, target size and weather, and there are few effective feature types can be extracted for target tracking. As a result, the existing target tracking algorithms are not ideal when they are used for tracking targets in GF-4 satellite sequence images.

In this paper, we propose a method for tracking maritime targets in GF-4 sequence images. The main work of this paper includes:

1. Research on background suppression methods. Based on the difference of radiance information of the maritime target and its adjacent pixels, redundant pixels such as land and thick cloud are removed from the images, and

water area is extracted. According to the characteristics of GF-4 satellite continuous imaging, the traditional dark channel prior method is improved to model the background of the sequence images, remove the additive noise, and achieve the effect of suppressing the background. An improved multi-scale Retinex illuminance removal algorithm for GF-4 satellite sequence images is designed to remove the multiplicative noise and achieve the purpose of enhancing the targets.

2. Propose a spatio-temporal based target tracking methods. This paper analyzes the target tracking model based on correlation of Marine targets in GF-4 satellite sequence images, uses imaging information to determine the size of targets in each band, extracts the multi-scale gradient distribution characteristics of targets, increases the number of target features, and enhances the correlation of targets between adjacent frames. The improved kernel method is used to map the high-dimensional features to the kernel space, which reduces the complexity of the algorithm and improves the tracking efficiency.

3. Optimize target tracking results. Combining with the historical information of target trajectory, the unscented kalman filter model is used to predict and correct the wrong target position. Using weighted image local contrast as the index, the image saliency detection of targets' potential area is carried out to further improve the positioning accuracy.

II. RELATED WORKS

Since the GF-4 satellite was officially put into use, it has provided a large number of timely and accurate detailed data for research work in various fields. With the iterative upgrading of science and technology methods, the application of GF-4 satellite imaging data has gradually changed from extensive environmental monitoring to highly refined small target dynamic monitoring. Zhang[1] analyzed the performance characteristics of GF-4 satellite imaging load, combined time-sharing imaging with AIS data, and demonstrated the practicability of GF-4 satellite imaging data in Marine ship dynamic monitoring. Zhang [2] proposed a ship dynamic monitoring method based on single-frame image target detection. Firstly, the gray gradient threshold method was used to detect targets in the image, and the ship course was estimated according to the gray gradient distribution of the target. This method has a good effect on target monitoring in GF-4 multi-band image. Liu[3] calculated NDWI index by using the difference in reflectance of ground objects in green band and near infrared band, then performed mask processing on cloud and land, extracted water area in the image, used two-parameter constant false alarm algorithm to detect targets in the image, and then used RANSAC algorithm to reduce the false alarm detection rate of targets. Then MHT algorithm is used for multi-frame correlation of target to achieve the purpose of target dynamic monitoring. Yao[4]

used the local peak signal-to-noise ratio as an indicator to detect targets in a single frame image, and then corrected the detection results through AIS data. On this basis, MHT algorithm was used to track targets. Liu[5] demonstrated the effect of multi-source data fusion in Marine target tracking. After two-parameter constant false alarm rate target detection in GF-4 satellite sequence images and target tracking based on multi-hypothesis tracking algorithm, the ship trajectory generated was fused with the de-sampled AIS data for track level data, which significantly improved the ship positioning accuracy. Liu [6] studied the application of high and low orbit optical satellite data fusion in maritime target monitoring, conducted target detection and tracking on the images generated by GF-4 satellite in geostationary orbit, and carried out target detection and feature extraction on the images generated by GF-1 satellite in low Earth orbit, and then conducted multi-level correlation between the two results. The ship correlation accuracy is improved effectively. Lin[7] used deep learning methods to conduct dynamic monitoring of moving targets in GF-4 satellite images which contained cloud fragments, used two-parameter CFAR algorithm to extract local candidate regions, used trained convolutional neural networks to detect targets in candidate regions, and combined the middle latitude method with the global nearest neighbor algorithm to achieve target association and tracking.

By summarizing the above methods, it can be seen that the current research on dynamic monitoring of dim small targets in GF-4 satellite sequence images is mainly carried out with Marine ships as the observation objects. The main methods adopted are as follows: firstly, targets in the images are detected on the basis of single or multi-frame images, and then the detected results are correlated between frames to obtain the target motion trajectory. This detection before tracking method is commonly used in tracking dim targets in video, and it also has certain reference and enlightening significance for solving the target tracking problem of GEO optical satellite sequence images with higher precision. However, due to the limitations of satellite imaging load technical indicators and imaging mode, this method is not that suitable for GF-4 satellite, as the length of ships operating at sea usually ranges from 100m to 400m, while the optimal spatial resolution of GF-4 satellite sequence images is only 50m. The size of targets in the images is only a few to a dozen pixels, while seawater and sea fog pixels will affect the target radiance of ships. This is to say, for both the traditional method and the deep learning method, it is difficult to extract enough features from a small number of pixels to identify details such as ship type without prior knowledge. Different from the short interval between traditional video frames, the frame interval of visible near-infrared multispectral images generated by GF-4 satellite is usually tens of seconds to several minutes, and the

ship's running speed at sea is about 30 knots. The position of the same target between adjacent frames of the sequence image is almost no coincidence, and the correlation between frames is weak. The reliability of the correlation between the target frames based on the sequence image detection results is not high, especially when the distance between adjacent ships is relatively close, the error rate of the generated ship trajectory results will be significantly increased. In order to solve the above problems, some researchers have integrated the target detection results of GF-4 satellite sequence images with AIS and low-orbit optical satellite data to improve the tracking accuracy of targets. However, from the perspective of practical application, the efficiency of this method is not high, which belongs to the right direction with improper method, because in the case of the ship position known, the target detection of GF-4 satellite sequence images is time-consuming and unnecessary.

III. METHODS

The detailed flow of GF-4 sequence images' maritime targets tracking method proposed in this paper is shown in Figure 1.

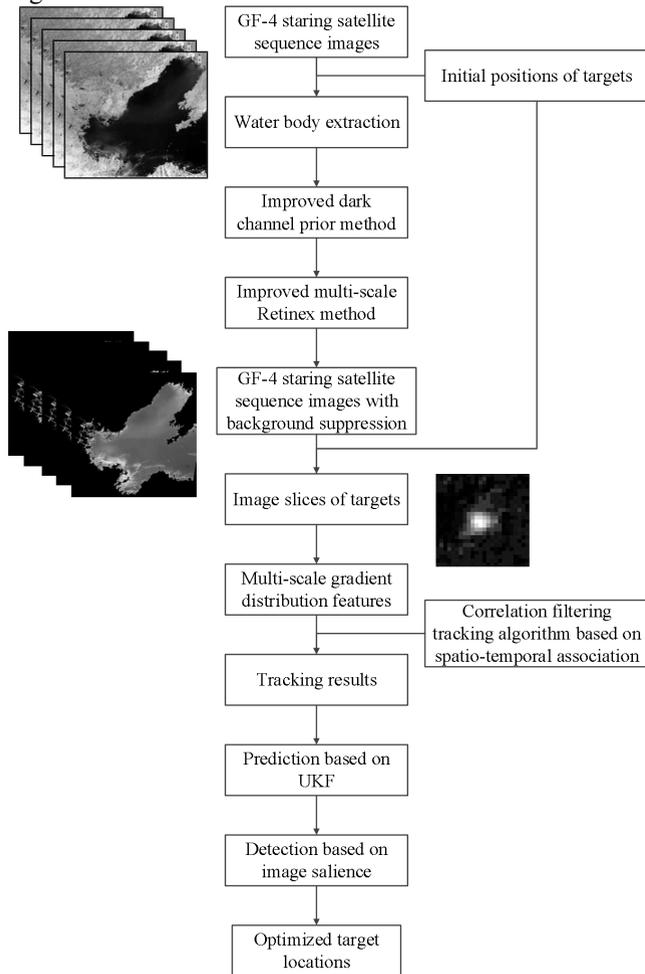


FIGURE 1. The flow of the proposed target tracking algorithm in this paper.

The target tracking method can be divided into three parts, namely, background suppression of sequential images, target tracking based on spatial-temporal association, and tracking result optimization based on target characteristics. Taking GF-4 sequence images as an example, we are able to extract the water body of the images where the targets locate by using the initial location information of the targets firstly. With the characteristics of multi-band sequential imaging, we can improve the dark channel method and the retinex method to suppress the images' background and enhance the ship targets. Secondly, an improved correlation filtering algorithm is used to extract multi-band and multi-scale features of the sequence image to improve the feature dimension of the targets. Finally, the tracking results are predicted and corrected through ukf method and significance detection principle, and then the high efficiency tracking of maritime targets in GF-4 satellite sequence images is realized.

A. SEQUENTIAL IMAGES BACKGROUND SUPPRESSION

1) WATER BODY EXTRACTION

According to the USGS ground object spectrum database, in the visible near-infrared band, the reflected radiance of land and thick clouds is higher than that of sea water and ships, and the reflected radiance of ships is slightly higher than that of sea water. The study of multiple sets of GF-4 satellite sequence images shows that the differences in reflected radiance of objects in different places in the images conform to the above rules, that is, affected by factors such as season, time, illumination, climate and region, the reflected radiance of the same category of objects in GF-4 satellite images is also different, while the differences in reflected radiance of different categories of ground objects remain consistent in the overall trend. If the reflection radiance range of ship and water pixels is known, the category of objects can be distinguished by threshold method, and the water region of the sequence image can be extracted by the calculation method as follows:

$$I_{mask} = \begin{cases} 1, & I_{min} \leq I(x, y) \leq I_{max} \\ 0, & others \end{cases} \quad (1)$$

where, I_{max} is the maximum radiance of all targets and their adjacent pixels; I_{min} is the minimum radiance value of all targets and their adjacent pixels; I_{mask} represents the mask region obtained by the threshold segmentation operation.

By applying formula (1) to the GF-4 sequence images, the complete water body of the sequence images can be obtained without redundant information such as land and thick cloud.

2) BACKGROUND SUPPRESSION

Earth imaging by staring satellites is a complicated process. The pixel information of satellite image mainly includes: the reflection information of ground objects in the area to be observed; the scattering information of adjacent ground objects; scattering and reflection information of attenuated

atmospheric light and sea fog. The mathematical expression of GF-4 satellite imaging model is as follows:

where $I(x, y)$ represents the image collected by GF-4 satellite; $J(x, y)$ represents the true reflection information of the area to be observed; $H(x, y)$ stands for atmospheric attenuation function, that is, multiplicative noise of imaging system; $n(x, y)$ represents the additive noise of the imaging system.

The reflection information of objects in the attenuated area to be observed is the objective function, and its calculation expression is as follows:

$$J(x, y) = \frac{I(x, y) - n(x, y)}{H(x, y)} \quad (3)$$

Due to the wide imaging width of GF-4 satellite, the noise of different positions in the image is also different, and it constantly changes with the change of climate, season, time, light and many other factors. In order to get high quality surface object reflection images, the additive noise and multiplicative noise were estimated and removed through the dark channel prior and multi-scale Retinex illuminance respectively.

The atmospheric scattering model proposed by Narasimhan et al. [8,9] has been widely used in the research field of optical remote sensing image degradation, and its mathematical expression is as follows:

$$I(x) = J(x)t(x) + A[1 - t(x)] \quad (4)$$

where $I(x)$ is the image obtained by the acquisition device; $J(x)$ is the image to be recovered; $t(x)$ is atmospheric transmittance; A is the atmospheric light value.

There are many methods to solve $J(x)$. He et al [10] proposed the dark channel prior theory after analyzing a large number of visible fog free images, and combined it with the atmospheric scattering model to achieve better fog removal effect for visible images. Pan et al [11] found in the research process that compared with the visible light image, the gray statistical histogram of remote sensing multi-spectral image has an overall right shift phenomenon, and proposed an improved dark channel defogging enhancement algorithm applicable to single frame remote sensing multi-spectral image: After extracting the dark channel, the gray statistical histogram of the dark channel is moved to the left to meet the requirements of the dark channel defogging method. The dark channel extraction method is as follows:

$$(J(x) - C)^{dark} = \min_{y \in \Omega(x)} (\min_{c \in \{R, G, B\}} (\max(J(y) - C, 0)^c)) \quad (5)$$

where C is obtained by statistics of non-fog remote sensing images; $(J(x) - C)^{dark}$ represents the dark channel of remote sensing image containing fog; $\Omega(x)$ is some neighborhood centered on the pixel x .

Based on the successful application of the above dark channel prior method in the fog removal of single-frame multi-spectral remote sensing images, the background

modeling of GF-4 satellite sequence images can be conducted by using the imaging characteristics of GF-4

$$I(x, y) = J(x, y) \times H(x, y) + n(x, y) \quad (2)$$

satellite and improving the dark channel prior method. The specific methods are as follows:

$$J_{i,j}^{background}(x) = \min_{y \in \Omega(x)} (\min_{c \in \{S_{i-1,j}, S_{i,j}, S_{i+1,j}\}} (J^c(y))) \quad (6)$$

where $J_{i,j}^{background}$ represents the background modeling result of the j th band of the i th frame image; $\Omega(x)$ is a neighborhood of the center of the pixel x ; $S_{i,j}$ represents the j th band of the i th frame image; J^c indicates a band.

To improve the reliability of the image background model, multi-frame images are selected for background modeling based on the improved dark channel prior method, and the results are weighted. The calculation method is as follows:

$$J_{(i,j)}^{BG}(x) = \sum_k \omega_k \times J_{(k,i,j)}^{background}(x) \quad (7)$$

where $J_{(i,j)}^{BG}$ represents the weighted background modeling result of the j th band of the i th frame image; ω_k is the weighting coefficient, $\omega_k \in (0, 1)$; $J_{(k,i,j)}^{background}$ represents the background modeling result between the j th band of the k th frame image and the j th band of the i th frame image, $J_{(k,i,j)}^{background}(x) = \min_{y \in \Omega(x)} (\min_{c \in \{S_{k,i-1,j}, S_{k,i,j}, S_{k,i+1,j}\}} (J^c(y)))$.

By differentiating the above results with the original sequence image, the additive noise from the surrounding pixels in the image can be removed, and the calculation expression is as follows:

$$D_{(i,j)}(x) = I_{(i,j)}(x) - J_{(i,j)}^{BG}(x) \quad (8)$$

where $D_{(i,j)}$ represents the removing additive noise result of the j th band of the i th frame image; $I_{(i,j)}$ represents the j th band of the i th frame image; $J_{(i,j)}^{BG}$ represents the weighted background modeling result of the j th band of the i th frame image.

In order to restore the real reflection information of ground objects, multi-scale Retinex illumination processing is needed to remove the multiplicative noise and enhance the targets. According to the Retinex theory proposed by Land, the object information received by human eyes or imaging equipment is determined by the reflection information of the object itself and the incident light information [12], and the imaging model expression is as follows:

$$I(x, y) = R(x, y) \times L(x, y) \quad (9)$$

where $I(x, y)$ is the image information received by the imaging device; $R(x, y)$ is the reflection information of the object; $L(x, y)$ is the incident illuminance information.

To solve the problem that the single-scale illuminance method cannot achieve optimal results in both image detail and color range, some researchers proposed use Gaussian kernels of different scales to filter the original image

respectively in the approximate estimation of incident light component, and then the filtering results are weighted according to different weights, that is, multi-scale Retinex illuminance:

$$R_{MSR}(x, y) = \sum_{k=1}^N \omega_k \times [\ln(I(x, y)) - \ln(I(x, y) * G_k(x, y))] \quad (10)$$

where $R_{MSR}(x, y)$ is the image's multi-scale Retinex illumination result; N is the scale of Gaussian kernel function; ω_k is the weight coefficient; $G_k(x, y)$ represents a Gaussian function on the k th scale.

Since the imaging process of GF-4 satellite is multi-band sequential imaging, and the image width is large, the improved multi-scale Retinex illuminance removal method is considered to effectively remove the multiplicative noise in the image band by band, and its mathematical expression is as follows:

$$D_{MSR(i,j)}(x, y) = \ln(D_{(i,j)}(x, y)) - \sum_{k=1}^N \omega_{k,j} \times [\ln(D_{(i,j)}(x, y) * G_{(k,j)}(x, y))] \quad (11)$$

where $D_{MSR(i,j)}(x, y)$ represents the result of removing multiplicative noise in the j th band of the i th frame image; $D_{(i,j)}(x, y)$ is the result of removing additive noise in the j th band of the i th frame image by improved dark channel prior method; $\omega_{k,j}$ is the weighting coefficient of the k th scale Gaussian filtering results in the j th band; $G_{(k,j)}(x, y)$ represents a Gaussian function in the j th band on the k th scale.

In the process of image Gaussian filtering, to reduce the influence of ship pixels on peripheral pixels and improve the accuracy of incident light component estimation, the background model $J_{D(i,j)}^{BG}$ of image $D_{(i,j)}$ is obtained, and the image and background model are weighted, and we can get:

$$D_{(i,j)}^{BG}(x, y) = \omega_j \times D_{(i,j)} + (1 - \omega_j) \times J_{D(i,j)}^{BG} \quad (12)$$

where ω_j is the weighting coefficient of the j th band; $D_{(i,j)}^{BG}(x, y)$ is the optimized result of image $D_{(i,j)}$.

By substituting equation (12) into equation (11), we get:

$$D_{MSR(i,j)}(x, y) = \ln(D_{(i,j)}(x, y)) - \sum_{k=1}^N \omega_{k,j} \times [\ln(D_{(i,j)}^{BG}(x, y) * G_{(k,j)}(x, y))] \quad (13)$$

According to formula (13), the multiplicative noise in the GF-4 satellite sequence image can be effectively removed, and the reflected radiance information of the objects can be restored. In order to facilitate visual observation and further enhance the contrast difference between different objects in the image, the image is linearly stretched, and the processing method is as follows:

$$I_{D(i,j)}(x, y) = \frac{D_{MSR(i,j)}(x, y) - D_{\min}}{D_{\max} - D_{\min}} \times 255 \quad (14)$$

where $I_{D(i,j)}(x, y)$ represents the stretching result of the j th band of the i th frame image; D_{\min} and D_{\max} are the

minimum and maximum values of the gray scale of image $D_{MSR(i,j)}(x, y)$.

The additive noise and multiplicative noise are suppressed, and the ship targets are enhanced, by applying formulas (7), (8), (12), (13) and (14) to the sequence images with the water body obtained in the previous part.

B. TARGETS TRACKING

1) REVIEW OF KERNEL CORRELATION FILTERING TRACKING ALGORITHMS

Bolme et al [13] introduced the correlation theory into target tracking for the first time: the targets' information in images were used to construct a filter template, and the subsequent images were correlated filtered. The maximum response position was the target position in the subsequent images. It significantly improved the tracking speed and accuracy compared with traditional methods[14-30], and thus opened up the research direction of target tracking based on correlation filtering. Henriques et al [31] proposed to construct a cyclic matrix by shifting the basic samples, which not only expanded the training samples and saved the storage space, but also reduced the algorithm complexity. Henriques et al [32] mapped the image HOG feature [33] into the kernel space for the first time, transformed the nonlinear regression problem into a linear regression problem through the kernel function, realized the fast tracking of the target, and the algorithm complexity was significantly optimized. The general flow of kernel correlation filtering tracking is:

Let a set of basis of kernel space be $\Phi(X) = [\Phi(x_1), \Phi(x_2), \dots, \Phi(x_n)]$, then the regression function of nonlinear regression problem can be expressed as:

$$f(x_i) = \omega^T \Phi(x_i) \quad (15)$$

where ω is the column vector of the weight coefficient, $\omega = \Phi(X)^T \alpha$.

In this case, the loss function is:

$$\min_{\alpha} \left(\left\| \Phi(X) \Phi(X)^T \alpha - y \right\|^2 + \lambda \left\| \Phi(X)^T \alpha \right\|^2 \right) \quad (16)$$

Taking the derivative of equation (19) and setting the derivative to 0, we get:

$$\alpha = \left(\Phi(X) \Phi(X)^T + \lambda I \right)^{-1} y \quad (17)$$

Let $K = \Phi(X) \Phi(X)^T$, then:

$$\alpha = \left(K + \lambda I \right)^{-1} y \quad (18)$$

When the basis function of the kernel space is a Gaussian kernel function, the matrix K is a cyclic matrix. Using the cyclic matrix property, equation (18) can be simplified to:

$$\hat{\alpha} = \frac{\hat{y}}{\left(\hat{K}^{xx} + \lambda \right)^*} \quad (19)$$

where \hat{K}^{xx} is the first row of the cyclic matrix K , which is also a Gaussian function, and its mathematical expression is as follows:

$$K^{xx'} = \exp\left(-\frac{1}{\sigma^2}\left(x^2 + x'^2 - 2\mathcal{F}^{-1}\left(x^* \odot \hat{x}'\right)\right)\right) \quad (20)$$

The objective function is:

$$\hat{f}(\mathbf{z}) = \hat{K}^{xx} \odot \hat{\alpha} \quad (21)$$

2) MULTI-CHANNEL FEATURE EXTRACTION BASED ON SPATIO-TEMPORAL INFORMATION

In GF-4 satellite image, the targets' sizes are often slightly larger than the theoretical value, due to the influence of the imaging load integration time, the satellite system conversion time and the targets' own scattering. The calculation method of targets' size is as follows:

$$l_{image} = (\alpha_{ship} \times l_{ship} + v_{ship} \times t_{band}) / \beta_{image} \quad (22)$$

where l_{image} represents the size of the target in the image; α_{ship} represents the expansion coefficient of the target in the image due to light scattering. l_{ship} is the target length of the ship, generally between 150m and 350m; v_{ship} is the target speed of the ship, usually between 20 knots and 35 knots; t_{band} represents the image imaging time of a certain band, including the integration time of the imaging load and the conversion time of the satellite system, which can generally be obtained by querying the image header file; β_{image} is the spatial resolution of the image.

If the location (x_0, y_0) of the target is known, the four-point coordinates of the ship's area are:

$$\begin{cases} D_{left,top}(x_0, y_0) = x_0 - l_{image} / 2, y_0 - l_{image} / 2 \\ D_{right,top}(x_0, y_0) = x_0 + l_{image} / 2, y_0 - l_{image} / 2 \\ D_{left,bottom}(x_0, y_0) = x_0 - l_{image} / 2, y_0 + l_{image} / 2 \\ D_{right,bottom}(x_0, y_0) = x_0 + l_{image} / 2, y_0 + l_{image} / 2 \end{cases} \quad (23)$$

$D_{left,top}(x_0, y_0)$, $D_{right,top}(x_0, y_0)$, $D_{left,bottom}(x_0, y_0)$ and $D_{right,bottom}(x_0, y_0)$ respectively represent the vertex coordinates of the image centered on (x_0, y_0) : upper left, upper right, lower left and lower right

In this paper, the initial values of the targets are the positions in the near-infrared band, while the GF-4 satellite image is sequential imaging in each band, and the positions of the moving targets in each band are different. Therefore, when calculating the size of the targets region in other bands, the images' size should be appropriately expanded. If the size $l_{ship-nir}$ of the targets in the near infrared band are known, their sizes in other bands are calculated as follows:

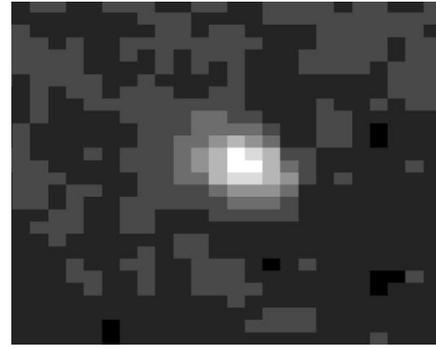
$$l_{band_i} = v_{ship} \times \Delta T_i / 50 + l_{ship-nir} \quad (24)$$

where l_{band_i} represents the size of the region where the target is located in the i th band; ΔT_i represents the imaging interval between the i th band and the near-infrared band.

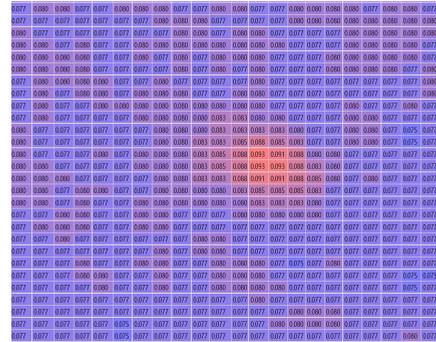
According to the image timing information, taking the sample size in the near-infrared band as the benchmark, the calculation method of the training sample size l'_{band-i} and the sample size l'_{band-i} of the target in the i th band is as follows:

$$\begin{cases} l'_{band-i} = k_{band-i} \times (\alpha_{ship} \times l_{ship} + 2 \times v_{ship} \times t_{band-i}) / \beta_{tiff} \\ l'_{band-i} = 2 \times v_{ship} \times \Delta t_{tiff} / \beta_{tiff} + l_{band-i} \end{cases} \quad (25)$$

Taking the near-infrared band of GF-4 satellite as an example, the ship neighborhood image and its reflected radiance data are shown in Figure 2.



(a)



(b)

FIGURE 2. Target neighborhood image and reflection radiance data. (a) is target neighborhood image in near infrared band. (b) is radiance of target neighborhood in near infrared band.

In order to extract HOG features of targets, the horizontal and vertical gradients of each pixel in the image are required first, and the calculation method is as follows:

$$\begin{cases} G_x(x_0, y_0) = I(x_0 + 1, y_0) - I(x_0 - 1, y_0) \\ G_y(x_0, y_0) = I(x_0, y_0 + 1) - I(x_0, y_0 - 1) \end{cases} \quad (26)$$

where $G_x(x_0, y_0)$ and $G_y(x_0, y_0)$ represent the horizontal and vertical gradients of pixel (x_0, y_0) respectively; $I(x_0, y_0)$ represents the reflected radiance of the pixel (x_0, y_0) .

The gradient amplitude and direction of each pixel in the image are calculated as follows:

$$\begin{cases} G(x_0, y_0) = \sqrt{G_x(x_0, y_0)^2 + G_y(x_0, y_0)^2} \\ \phi(x_0, y_0) = \arctan(G_y(x_0, y_0)/G_x(x_0, y_0)) \end{cases} \quad (27)$$

where $G(x_0, y_0)$ and $\phi(x_0, y_0)$ represent the gradient amplitude and gradient direction of pixel (x_0, y_0) .

The direction circle is divided into n_θ angle intervals with angle θ as the interval, and the interval of each pixel gradient direction is counted, and the pixel gradient amplitude belonging to the same angle interval is superimposed to obtain the statistical histogram of the gradient distribution of the image. The gradient distribution is calculated as follows:

$$M_i = \sum_j G(x_j, y_j), \quad \phi(x_j, y_j) \in [n_i \times \theta, (n_i + 1) \times \theta) \quad (28)$$

where n_i is the serial number of the angle interval, $n_i \in [1, n_\theta]$; M_i is the total amplitude of the gradient in the n_i Angle interval.

When the image contains a large number of pixels, the statistical histograms of gray gradients of different features may show similar structures or distributions. In order to avoid the above situation, the image is divided into several sub-blocks, and the gray gradient histogram of each sub-block is counted respectively, and the proportion of each sub-block in its neighborhood is calculated. With step length k as the interval, the image is divided into several sub-blocks, each sub-block is called a cell, and each cell contains $k \times k$ pixels. Then, the image is divided into several regions with step length $k \times l$ as the interval. Each region is called a block, and each block contains $l \times l$ cells. When $k = 2, l = 3$, each block in the image is of size 6×6 and contains 3×3 cell subblocks of size 2×2 . Figure 3 shows a block region in the ship neighborhood image in Figure 2. The yellow dashed line divides the region into 3×3 cell subblocks. The weight of each cell subblock in the block area is calculated as follows:

$$\omega_i = cell_i / \sum_j^{n_{cell}} cell_j \quad (29)$$

where n_{cell} represents the number of cell subblocks contained in the block region; $cell_i$ represents the gray gradient histogram of the i th cell; ω_i indicates the weight of the i th cell subblock in the block area to which it belongs.

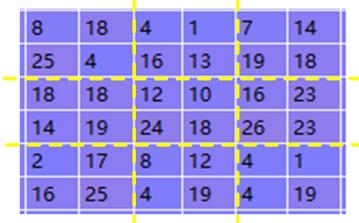


FIGURE 3. Subblock partitioning method of image.

In order to further enhance the uniqueness and exclusivity of the gray gradient histogram of each cell subblock, a block region is selected by taking each cell subblock in the image

as the center and extending a certain distance around it. The proportion of gray gradient histogram of center cell subblock in each subregion is calculated respectively, and the weight distribution of the center cell subblock in the image is represented by all the gravity. Taking the central cell subblock in Figure 3 as an example, the size of the block region obtained by expanding the size of one cell around it is 6×6 . Assuming that the size of each selected subregion is 4×4 pixels, there are four subregions in the block region including the central cell subblock, and the division of each subregion is shown by the colored dashed line in Figure 4. The weight of the center cell subblock in each subarea is calculated as follows:

$$\begin{cases} \omega_{1,2,4,5} = cell_5 / (cell_1 + cell_2 + cell_4 + cell_5) \\ \omega_{2,3,5,6} = cell_5 / (cell_2 + cell_3 + cell_5 + cell_6) \\ \omega_{4,5,7,8} = cell_5 / (cell_4 + cell_5 + cell_7 + cell_8) \\ \omega_{5,6,8,9} = cell_5 / (cell_5 + cell_6 + cell_8 + cell_9) \end{cases} \quad (30)$$



FIGURE 4. Block segmentation method in the image.

When the angle interval is separated by $\theta = 25^\circ$, the angle interval can be divided into 4 unoriented intervals and 8 directional intervals. In this case, the feature dimension of each cell subblock increases from the $(4+8)$ dimension of equation (29) to the $4 \times (4+8)$ dimension of equation (30).

Using the above method to extract target features from single-band images, the dimensions of the features obtained are as follows:

$$N_{image} = 3 \times N_\theta \times N_{cell} \times N_{block} \quad (31)$$

where, N_{image} is the characteristic dimension of the target in the single-band image; N_θ represents the number of angular intervals obtained with θ as the interval without distinguishing the direction; N_{cell} represents the number of cell subblocks contained in the image; N_{block} indicates the number of subregions used to calculate the weight of cell subblocks.

As mentioned above, the sequence image of GF-4 satellite has multiple bands, and the size of the target in the images of different bands is different, so the feature dimension of its multi-scale gray gradient distribution in the images of different bands is also different. The imaging mode of GF-4 satellite is sequential imaging by band, so the features of targets in images of different bands are independent of each other. By paralleling the above features, multi-scale features of targets based on image timing information can be obtained, and the feature dimension is as follows:

$$N_{diff} = \sum_{k=1}^{n_{band}} N_{image_k} \quad (32)$$

where N_{diff} represents the characteristic dimension of the target in the multi-spectral image; k is the image band number; N_{image_k} is the characteristic dimension of the target in the k -band image.

Features of different dimensions are extracted according to different Angle interval division, image subblock size and other parameters, and the results are connected in parallel according to a certain weight as the multi-scale gradient distribution features of targets. The mathematical expression is as follows:

$$\mathbf{x}'_i = \omega_i \times \mathbf{x}_i \quad (33)$$

where \mathbf{x}'_i represents the weighted feature; ω_i is the weighting coefficient, $\sum_i \omega_i = 1$; \mathbf{x}_i is the target feature when the i th parameter is set.

At this time, the maximum response $\hat{f}_{max}(\mathbf{Z})$ is calculated as follows:

$$\hat{f}_{max}(\mathbf{Z}) = \max_{1 \leq i \leq n_z} (\hat{K}^{x_i} \odot \hat{\alpha}) \quad (34)$$

where \mathbf{Z} is the image to be detected; n_z represents the number of sub-blocks; z_i represents the multi-scale gradient distribution feature of the sub-block image to be detected with the same size as the training sample.

When the above samples are used to calculate the correlation of adjacent frames, the response results of each band are different, and the response of each band is superimposed with different confidence degrees. The maximum value of the superimposed result is the target position. At this time, the maximum response $\hat{f}_{\Sigma-max}(\mathbf{Z})$ in image \mathbf{Z} to be detected is calculated as follows:

$$\hat{f}_{\Sigma-max}(\mathbf{Z}) = \max_{1 \leq i \leq n_z} \left(\sum_{j=1}^{n_{band}} \omega_j \times (\hat{K}^{x_j z_i} \odot \hat{\alpha}_j) \right) \quad (35)$$

where, n_z is the number of subblocks; n_{band} is the number of image bands; ω_j represents the confidence of the j th band; x_j is the training sample feature of the j th band; $z_{i,i}$ is the characteristic of the i th sub-block in the samples to be measured in the j th band; $\hat{\alpha}_j$ is the filter template of the j th band.

The size information of ship targets in different bands of the GF-4 sequence images can be obtained according to formula (22)~(25), and the multi-channel features of the ship targets in the sequence images can be extracted through formula (26)~(30). By substituting the above results into formula (35), we can get the ship targets tracking results of the GF-4 sequence images.

C. TRACKING OPTIMIZATION

1) TARGETS PREDICTION

Let \mathbf{x}_k be the state variable of the target at time k , and its mathematical expression is as follows:

$$\mathbf{x}_k = [p_{x,k} \quad p_{y,k} \quad v_k \quad \varphi_k \quad \dot{\varphi}_k]^T \quad (36)$$

In the formula, $p_{x,k}$ and $p_{y,k}$ represent the horizontal and vertical coordinates of the target at time k , respectively. v_k , φ_k and $\dot{\varphi}_k$ represent the velocity, course, and course angle change rate of the target at time k , respectively.

Let \mathbf{q}_k be the process noise at time k , and its mathematical expression is as follows:

$$\mathbf{q}_k = \begin{bmatrix} q_{a,k} \\ q_{\dot{\varphi},k} \end{bmatrix} \quad (37)$$

where $q_{a,k}$ and $q_{\dot{\varphi},k}$ are the noise of target acceleration and angular acceleration respectively, $\mathbf{q}_k \sim (0, \mathbf{Q}_k)$.

In this case, the equation of state can be expressed as:

$$\begin{aligned} \mathbf{x}_{k+1} &= f(\mathbf{x}_k, \mathbf{q}_k) \\ &= \mathbf{x}_k + \int (\dot{\mathbf{x}}_k + \mathbf{q}_k) dt \end{aligned} \quad (38)$$

The derivative of state variable \mathbf{x}_k at time k is:

$$\dot{\mathbf{x}}_k = \begin{bmatrix} \dot{p}_{x,k} \\ \dot{p}_{y,k} \\ \dot{v}_k \\ \dot{\varphi}_k \\ \dot{\dot{\varphi}}_k \end{bmatrix} = \begin{bmatrix} v_k \cos(\varphi_k + \dot{\varphi}_k t) \\ v_k \sin(\varphi_k + \dot{\varphi}_k t) \\ 0 \\ \dot{\varphi}_k \\ 0 \end{bmatrix} \quad (39)$$

By substituting equations (37) and (39) into equation (38), we get:

$$\mathbf{x}_{k+1} = \begin{bmatrix} p_{x,k} + (\sin(\varphi_k + \dot{\varphi}_k \cdot t) - \sin(\varphi_k)) \times v_k / \dot{\varphi}_k + q_{a,k} \times \cos(\varphi_k) \times t^2 / 2 \\ p_{y,k} + (\cos(\varphi_k) - \cos(\varphi_k + \dot{\varphi}_k \cdot t)) \times v_k / \dot{\varphi}_k + q_{a,k} \times \sin(\varphi_k) \times t^2 / 2 \\ v_k + q_{a,k} \times t^2 \\ \varphi_k + \dot{\varphi}_k \cdot t + q_{\dot{\varphi},k} \times t^2 / 2 \\ \dot{\varphi}_k + q_{\dot{\varphi},k} \times t \end{bmatrix} \quad (40)$$

Let \mathbf{z}_{k+1} be the measurement prediction at time $k+1$, and its mathematical expression is as follows:

$$\mathbf{z}_k = \begin{bmatrix} p_{x,k+1} \\ p_{y,k+1} \end{bmatrix} \quad (41)$$

Then the measurement prediction equation of the unscented kalman filter model in this section is as follows:

$$\begin{aligned} \mathbf{z}_{k+1} &= h(\mathbf{x}_{k+1}, \mathbf{r}_{k+1}) \\ &= \mathbf{H}\mathbf{x}_{k+1} + \mathbf{r}_{k+1} \end{aligned} \quad (42)$$

Where \mathbf{H} is the state transition matrix, $\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$; \mathbf{r}_{k+1} is the measurement noise at time $k+1$, $\mathbf{r}_k \sim (0, R_k)$

According to formula (42)~(48), the abnormal positions in the target tracking results can be screened out and the correct ship positions can be predicted.

2) TARGETS DETECTION

In recent years, with the deepening of the research on human visual system, the detection method based on image saliency has been widely used in the field of object detection, which has remarkable effects on background suppression and object enhancement. By observing the sequence images of GF-4 satellite, it can be seen that the salience of different ground objects in the images is different, that is, the contrast between the pixel of different types of ground objects and its neighboring pixel is different. The schematic diagram of the structural operator used in this section to calculate the local contrast of the image is shown in Figure 5.

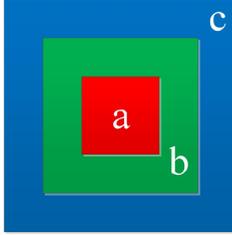


FIGURE 5. Structure of local contrast operator.

As is shown in Figure 5, the local contrast structure operator is divided into three layers: the innermost region a represents the target region where the contrast strength is to be calculated; The outermost c region represents the background region used to calculate contrast; Middle layer b is the isolation layer that separates the target area from the background area. Region local contrast(RLCM) is calculated as follows:

$$RLCM = (\bar{I}_a \times \bar{I}_a - \bar{I}_c \times \bar{I}_c) / \bar{I}_c \quad (43)$$

Where \bar{I}_a represents the mean value of the central region of the local contrast structure operator; \bar{I}_c represents the mean of the background region of the local contrast structure operator.

After background suppression processing, the thick cloud and land elements in the GF-4 satellite sequence image have been removed. The local contrast structure operator is used to traverse the target location neighborhood image predicted by the unscented Kalman filter model, and the location whose RLCM value is greater than the threshold value α is likely to be the target.

In order to determine the position of the target in the image, taking the last position of the correct historical segment data of the target as the center, the potential area of the ship in the current image is calculated as follows:

$$r_{image} = (\alpha_{ship} \times l_{ship} / 2 + v_{ship} \times \Delta t_{diff}) / \beta_{image} \quad (44)$$

where r_{image} is the radius of the potential area of the ship; α_{ship} is the coefficient of expansion; l_{ship} is length; v_{ship} is speed; Δt_{diff} is the imaging interval of adjacent images; β_{image} is the spatial resolution of the image.

The target in the image can be detected by calculating the local contrast of the potential area. The target detection results are compared with the ship position predicted by the unscented Kalman filter model. The closer the target is to the predicted position, the higher the probability that it is the correct target. Using the predicted position of the unscented Kalman filter model as the image center of gravity, the weighted Gaussian mask of the potential region is calculated:

$$I_w(x, y) = \exp\left(-\frac{(x-x_0)^2 + (y-y_0)^2}{\sigma^2}\right) \quad (45)$$

where $I_w(x, y)$ is the mask weight of (x, y) in the image; (x_0, y_0) is the predicted position coordinate of the unscented Kalman filter model.

The RLCM value of the potential area is multiplied by the weighted Gauss mask pixel by pixel, and the target's position is that where the maximum value of the result locates. The weighted local contrast of the location $I_{ship}(x, y)$ is calculated as follows:

$$I_{ship}(x, y) = \max(I_w(x, y) \times RLCM(x, y)) \quad (46)$$

where $I_w(x, y)$ is the mask weight of (x, y) ; $RLCM(x, y)$ is the local contrast of (x, y) .

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. DATASETS

In this paper, 4 sets of multi-spectral sequence images taken by GF-4 satellite in different regions, different seasons and different intervals are selected in the process of studying the tracking problem of staring satellite sequence images. In order to evaluate the following target tracking effect, it is necessary to determine the number and location of targets in each sequence image. Since the imaging mode of GF-4 satellite multi-spectral image is sequential imaging by band, the positions of moving targets in each band are different, the difference between different bands can be used to conduct expert discrimination of targets, and the position of targets in near-infrared bands is mainly marked as the truth value of follow-up tracking effect evaluation. The basic information of each sequence image is shown in Table I.

TABLE I
STATISTICS OF GF-4 SATELLITE SEQUENCE IMAGE INFORMATION

Serial number	Imaging date	Central coordinate	Image quantity /fram	Imaging interval	Target quantity
1	20161128	121.7° E, 38.9° N	14	186	162

2	20170303	119.5° E, 38.7° N	10	67	139
3	20191208	112.8° E, 22.9° N	3	69	150
4	20170405	118.6° E, 13.1° N	4	246	35

We use ENVI software, a professional image processing tool, to process radiometric correction, atmospheric correction and orthographic correction of the sequence images, and then make the first frame of each sequence image as the basis to achieve internal registration of the image sequence through SIFT algorithm.

B. EVALUATION INDEX

1) EVALUATION INDICATORS OF BACKGROUND SUPPRESSION

The contrast difference between the target and the neighborhood background is used as the index to evaluate the target enhancement effect. The mathematical expression of the difference measurement BSR(background-shipratio) between the Background and the ship is as follows:

$$BSR = \left(\sum_i^{n_i} I_i^{sea}(x, y) \div n_i \right) / \left(\sum_j^{n_j} I_j^{ship}(x, y) \div n_j \right) \quad (47)$$

where $I^{sea}(x, y)$ represents the gray level of the ship neighborhood background; $I^{ship}(x, y)$ represents the gray level of the ship; n_i and n_j represent the number of background and ship pixels used to calculate the BSR, respectively.

To reduce the influence of background pixels in the target area, the average value of some maximum values in the target area is taken as the reflection radiance. While the average value of some intermediate values in the background region is taken as the background reflection radiance, in order to reduce the influence of other ship pixels in the background region. In the GF-4 satellite sequence image, the reflectance of the target is higher than that of the water body, so the BSR value ranges from 0 to 1.

2) EVALUATION INDICATORS OF TARGET TRACKING

The distance between the real position of the target and the tracking result is used as the index to evaluate the target tracking effect. d represents the distance between the real location of the target and the tracking result. When d is larger than the threshold value l , the tracking result is judged to be wrong. The ratio of the number of errors to the number of all targets in one frame is defined as the single-frame loss rate, which is calculated as follows:

$$p_i = \frac{m_i}{M} \quad (48)$$

where p_i represents the probability of target tracking error in the i th frame; m_i represents the number of the targets with tracking errors in the i th frame; M represents the total number of the targets.

The average of the single frame loss rate of the sequence images is defined as the average loss rate, which is calculated as follows:

$$p_{avg} = \frac{1}{N} \sum_{i=1}^N p_i \quad (49)$$

It can be calculated by substituting the four sequence images' imaging intervals of this study into equation (22) that the length of the targets in the near infrared band ranges from 14 to 25 pixels. The tracking result is judged to be wrong when the distance between the tracking result and the real position of the ship is larger than 5 pixels.

C. EXPERIMENTAL RESULTS

The calculated and statistical BSR values of the targets and their neighborhood background in the image are shown in Table II.

TABLE II
BSR OF TARGETS AND THEIR NEIGHBORHOOD IN EACH SEQUENCE IMAGES

Serial number	BSR_max	BSR_min	BSR_max after suppression	BSR_min after suppression
1	0.9999	0.4186	0.0000	0.0000
2	0.9444	0.3478	0.2000	0.0000
3	0.8666	0.2174	0.0400	0.0000
4	0.8750	0.4074	0.0000	0.0000

As can be seen from Table 3, the difference between target and background in the four sequence images has been improved, achieving the purpose of target enhancement, indicating that the proposed method is effective in background suppression of GF-4 satellite sequence images.

The kernel correlation filtering method(KCF) in reference [32] is selected as a comparison, as it performs better among the traditional correlation filtering methods. ICF represents the improved correlation filtering algorithm with multi-channel feature extraction. The above two methods are used to background suppressed sequence images, which are represented by BS-KCF and BS-ICF. The proposed method in this paper is represented by TSCF(temporal-spatio based correlation filtering method). The single-frame and average loss rate of targets tracking in each sequence images are calculated, and the statistics of the single frame loss rate and average loss rate of targets in each sequence images are shown from Table III to Table VI.

TABLE III
TARGET TRACKING LOSS RATE OF THE 1ST SEQUENCE IMAGES

Serial number	KCF	ICF	BS-KCF	BS-ICF	TSCF
1	0.8951	0.7099	0.1914	0.0494	0.0494
2	0.9568	0.8580	0.2407	0.0617	0.0556
3	0.9506	0.9074	0.2346	0.1235	0.0926

4	0.9383	0.9259	0.2901	0.1235	0.0926
5	0.9444	0.9383	0.2963	0.1420	0.1111
6	0.9444	0.9444	0.3148	0.1358	0.0988
7	0.9568	0.9444	0.3272	0.1481	0.1049
8	0.9198	0.9506	0.3333	0.1481	0.0864
9	0.9753	0.9506	0.3580	0.1543	0.1235
10	0.9630	0.9630	0.3704	0.2037	0.1605
11	0.9630	0.9753	0.4074	0.2037	0.1481
12	0.9506	0.9815	0.3889	0.2222	0.1605
13	0.9568	0.9753	0.4630	0.3210	0.2407
Mean value	0.9473	0.9250	0.3243	0.1567	0.1173

3	0.9429	1.0000	0.2571	0.2286	0.1429
Mean value	0.8952	0.9524	0.2286	0.1905	0.1238

By analyzing the statistical results in the upon tables, the following conclusions can be drawn:

(1)When tracking ship targets in sequence images, the single frame loss rate of all the target tracking algorithms based on correlation filtering mentioned in this paper increases with the increase of the number of frames;

(2)The ship targets in the original sequence image are submerged in complex background noise, while the background suppression method removes the sequence images' background noise and enhances the targets. Moreover, the single frame loss rate and average loss rate of the ship targets in the original sequence images are higher than that in the suppressed sequence images, indicating that the quality of the sequence image will affect the tracking results;

(3)When using the method proposed to track ship targets in sequence images, the average target loss rate of the 2nd sequence images is the lowest, and that of the 4th sequence images is the highest. According to the sequence image header file, the imaging intervals of the sequence images are 186s, 67s, 69s and 246s, respectively. It shows that the tracking effect of the method proposed decreases with the increase of the interval between frames;

(4)For the same sequence images, the ship target loss rate of the target tracking method proposed in this study is lower than that of the methods used as comparison, indicating that the method proposed in this aticle is better than the comparison algorithm in terms of the tracking effect of small dim targets on the sea in GF-4 satellite sequence images;

(5)By suppressing image background, improving kernel correlation filtering algorithm and optimizing target tracking results, the method proposed in this paper achieves the best tracking effect and the lowest loss rate, indicating that the method has adaptability and good robustness to sequence images of different quality.

TABLE IV
TARGET TRACKING LOSS RATE OF THE 2ND SEQUENCE IMAGES

Serial number	KCF	ICF	BS-KCF	BS-ICF	TSCF
1	0.4748	0.0791	0.0432	0.0072	0.0000
2	0.4748	0.1511	0.0647	0.0144	0.0072
3	0.4748	0.1655	0.0504	0.0072	0.0000
4	0.5252	0.2158	0.0719	0.0072	0.0000
5	0.5827	0.2086	0.1007	0.0144	0.0072
6	0.5683	0.2446	0.1439	0.0216	0.0000
7	0.6187	0.2878	0.1295	0.0432	0.0072
8	0.5827	0.2734	0.1511	0.0504	0.0072
9	0.5971	0.2806	0.1727	0.0576	0.0000
Mean value	0.5443	0.2118	0.1031	0.0248	0.0032

TABLE V
TARGET TRACKING LOSS RATE OF THE 3RD SEQUENCE IMAGES

Serial number	KCF	ICF	BS-KCF	BS-ICF	TSCF
1	0.4067	0.1933	0.2000	0.1000	0.0800
2	0.4533	0.2000	0.2000	0.1200	0.1000
Mean value	0.4300	0.1967	0.2000	0.1100	0.0900

TABLE VI
TARGET TRACKING LOSS RATE OF THE 4TH SEQUENCE IMAGES

Serial number	KCF	ICF	BS-KCF	BS-ICF	TSCF
1	0.8571	0.8857	0.2000	0.1714	0.1143
2	0.8857	0.9714	0.2286	0.1714	0.1143

V. Conclusion

In this paper, the problem of tracking small targets at sea in GF-4 satellite sequence images is studied. The multi-spectral sequence images of GF-4 satellite were used as the research object to extract the water region by using the difference of the radiance of different ground objects. By analyzing the satellite imaging model, an improved dark channel prior method is used to complete the image background modeling, and the original image and background are differentiated to remove the influence of additive noise on the targets in the image and achieve background suppression. Then, the improved multi-scale Retinex illuminance removal method is used to remove the multiplicative noise in the image, recover the real reflection information of the targets, and achieve the effect of the targets enhancement. Based on the sequence characteristics of the targets in sequential images, the target

tracking model based on image correlation is analyzed. Based on the gray gradient distribution, the multi-scale features of the targets in each band are extracted. By introducing the kernel method, the high-dimensional features of the targets are mapped to the kernel space, making it suitable for the tracking model based on image correlation, reducing the algorithm complexity and improving the tracking effect. The observation parameters of targets' motion are determined, and the tracking results of targets anomaly are predicted by unscented Kalman filter model, and the trajectory correction is realized initially. The local contrast of the image of the potential area is calculated, and then the weighted mask of the image is constructed with the predicted position as the center to realize the detection of the targets in the image, improve the positioning accuracy, and complete the tracking result optimization based on the target characteristics. The research of this paper fully demonstrates the application potential of staring satellites in the field of target tracking, and has a guidance for the subsequent application and development of staring satellites with higher precision and low-orbit optical satellite constellation.

ACKNOWLEDGMENT

We sincerely thank the editor and reviewers for their constructive comments.

REFERENCES

- [1] Z. Zhang, Y. Shao, W. Tian, Q. Wei, Y. Zhang and Q. Zhang. Application Potential of GF-4 Images for Dynamic Ship Monitoring. *IEEE Geoscience and Remote Sensing Letters*, 2017, 14, 911-915.
- [2] Z. Zhang, Q. Xu, and C. Zhang. Ship motion feature detection based on single-view Gaofen-4 satellite multi-spectral images(in Chinese). *Remote sensing technology and application*, 2019, 34, 892-900.
- [3] Y. Liu, L. Yao, W. Xiong, T. Jing & Z. Zhou. Ship target tracking based on a low resolution optical satellite in geostationary orbit[J]. *International Journal of Remote Sensing*, 2018, 39, 2991-3009.
- [4] L. Yao, Y. Liu, Y. He, et al. A Novel Ship-tracking Method for GF-4 Satellite Sequential Images. *Sensors*, 2018, 18, 1-14.
- [5] Y. Liu, L. Yao, W. Xiong, Z. Zhou. GF-4 Satellite and Automatic Identification System Data Fusion for Ship Tracking. *IEEE Geoscience and Remote Sensing Letters*, 2019, 16, 281-285.
- [6] Y. Liu, P. Guo, L. Cao, M. Ji and L. Yao. Information Fusion of GF-1 and GF-4 Satellite Imagery for Ship Surveillance. 2021 IEEE International Geoscience and Remote Sensing Symposium, 2021, 5044-5047.
- [7] X. Lin, L. Yao. Target Tracking of Moving Ships by Gf-4 Satellite under Cloud Fragmentation Environment(in Chinese). *Space return and remote sensing*, 2021, 42, 127-139.
- [8] S. G. Narasimhan, S. K. Nayar. Chromatic framework for vision in bad weather. *Computer Vision and Pattern Recognition*, 2000, 1, 598-605.
- [9] S. G. Narasimhan, S. K. Nayar. Vision and the Atmosphere. *International Journal of Computer Vision*, 2002, 48, 233-254.
- [10] He K, Jian S, Fellow, et al. Single Image Haze Removal Using Dark Channel Prior, *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2011, 33, 2341-2353.
- [11] Xiao Xi Pan, Feng Ying Xie, Zhi Guo Jiang, et al. Haze Removal for a Single Remote Sensing Image Based on Deformed Haze Imaging Model. *IEEE Signal Processing Letters*, 2015, 22, 1806-1810.
- [12] Land E H. The Retinex theory of color vision. *Scientific American*, 1978, 237, 108-128.
- [13] Bolme D S, Beveridge J R, Draper B A, et al. Visual object tracking using adaptive correlation filters. *Computer Vision and Pattern Recognition*, 2010, 2544-2550.
- [14] L. Yuepeng, Z. Shuyan, Z. Lirui and W. Xiaochen. Robust visual tracking via an online multiple instance learning algorithm based on SIFT features. 2016 IEEE International Conference on Signal and Image Processing, 2016, 85-89.
- [15] Z. Wang and X. Yang. Moving Target Detection and Tracking Based on Pyram-id Lucas-Kanade Optical Flow. 2018 IEEE 3rd International Conference on Image, Vision and Computing, 2018, 66-69.
- [16] N. Xie and Y. Shang. An Object Tracking Method by Concatenating Structural SVM and Correlation Filter. 2018 IEEE 3rd International Conference on Im-age, Vision and Computing, 2018, 701-705.
- [17] W. Wang, C. Wang, S. Liu, T. Zhang and X. Cao. Robust Target Tracking by Online Random Forests and Superpixels. *IEEE Transactions on Circuits and Systems for Video Technology*, 2018, 28, 1609-1622.
- [18] E. Dong, M. Deng and J. Tong. An Improved Struck Tracking Algorithm Based on Scale Adaptation and Selective Updating. 2021 IEEE International Conference on Mechatronics and Automation, 2019, 2309-2313.
- [19] Gomaa A, Abdelwahab M M, Abo-Zahhad M. Efficient vehicle detection and tracking strategy in aerial videos by employing morphological operations and feature points motion analysis. *Multimedia Tools and Applications: An International Journal*, 2020, 26023-26043.
- [20] L. Wang, H. Li and C. Han. Improved TLD gesture tracking algorithm based on SURF and Camshift. 2020 International Conference on Virtual Reality and Visualization, 2020, 1-4.
- [21] H. Yao, A Survey for target tracking on Meanshift algorithms. 2021 IEEE International Conference on Consumer Electronics and Computer Engineering, 2021, 476-479.
- [22] H. Wang, Q. Zhang, L. Yu and Z. Wang. Research on CAMshift Algorithm Based on Feature Matching and Prediction Mechanism. 2021 IEEE International Conference on Mechatronics and Automation, 2021, 773-778.
- [23] Long X, Ying W, Shengxing S. Discriminative Deep Non-Linear Dictionary Learning for Visual Object Tracking. *Neural Processing Letters*, 2022, 4007-4030.
- [24] Ahmed G, Tsubasa M, M. M A, et al. Faster CNN-based vehicle detection and counting strategy for fixed camera scenes. *Multimedia Tools and Applications*, 2022, 25443-25471.
- [25] Yuansheng L, Ping W, Mingyi Y, et al. Joint Detection, Tracking, and Classification of Multiple Extended Objects Based on the JDTC-PMBM-GGIW Filter. *Remote Sensing*, 2023, 887-887.
- [26] Dongsheng C, Zexiang G, Olusola B, et al. Multi-objective tracking for smart substation onsite surveillance based on YOLO Approach and AKCF. *Energy Reports*, 2023, 1429-1438.
- [27] Anggi D M, Carmadi M, Lenni Y, et al. Deep features fusion for KCF-based moving object tracking. *Journal of Big Data*, 2023,10(1).
- [28] Bao F, Zhang Z, Zhang G. An ensemble score filter for tracking high-dimensional nonlinear dynamical systems. *Computer Methods in Applied Mechanics and Engineering*, 2024, 117447-117447.
- [29] Wan G, Su Z, Wu Y, et al. High-Precision Multi-Object Tracking in Satellite Videos via Pixel-Wise Adaptive Feature Enhancement. *Sensors*, 2024, 6489-6489.
- [30] Gomaa A, Abdalrazik A. Novel Deep Learning Domain Adaptation Approach for Object Detection Using Semi-Self Building Dataset and Modified YOLOv4. *World Electric Vehicle Journal*, 2024, 255-255.
- [31] Henriques J F, Caseiro R, Martins P, et al. Exploiting the circulant structure of tracking-by-detection with kernels. *Proceedings of the 2012 International Conference on European Conference on Computer Vision*, 2012, 702-715.
- [32] Henriques J F, Caseiro R, Martins P, et al. High-speed tracking with kernelized correlation filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37, 583-596.
- [33] Navneet Dalal, Bill Triggs. Histograms of Oriented Gradients for Human Detection. *Computer Vision and Pattern Recognition*, 2005, 886-893.



Deyang Zhang received the master degree from School of Electronics and Information Engineering, Harbin Institute of Technology, in 2022. He is currently a lecturer of PLA Army Academy of Artillery and Air Defense(Zhengzhou Campus). His research interests include image processing, computer vision, and machine learning.



Yao Xu received the master degree from National University of Defense Technology. She is currently a lecturer of PLA Army Academy of Artillery and Air Defense. Her research interests include Electro-Optical Countermeasures and aero optics. She has co-authored about 10 papers indexed by SCI/EI and holds 3 China Inventive Patents.



Haimin Hu received the Ph.D. degree from the Nuclear Science & Engineering, University of Chinese Academy of Sciences, in 2008. He is currently an Associate Professor of PLA Army Academy of Artillery and Air Defense(Zhengzhou Campus). His research interests include image processing, computer vision, and VR. He has co-authored about 30 papers indexed by SCI/EI and holds 5 China Inventive Patents.