Robust Region Feature Extraction With Salient MSER and Segment Distance-Weighted GLOH for Remote Sensing Image Registration

Zilu Zhao, Feng Wang, and Hongjian You

***Abstract*—Remote sensing image registration is one of the crucial steps in remote sensing image processing, where ground control information is essential. Maintenance of control point databases is complex and expensive. Consequently, lightweight feature databases are emerging. Lightweight feature databases need to store stable and reproducible features. In this context, region features exhibit a distinct advantage. In feature registration methods, the reproducibility of regional features is typically stronger than with individual points. A popular feature region matching method is currently the combination of maximally stable extremal regions (MSER) and scale-invariant feature transform (SIFT). However, the direct combining of MSER and SIFT has difficulties primarily due to redundancy and overlap in regions extracted by MSER, as well as the conflict in applying texture descriptors on homogeneous regions. In this research, we first suggest a salient MSER detection approach that combines frequency-tuned salient region detection and effective nonmaximum suppression filtering to get rid of redundant information and enhance the stability and dependability of the feature region; afterward, we describe the feature region using the unique, enhanced segment distance-weighted gradient location-orientation histogram, which aims to comprehensively describe the feature regions by incorporating more information about the gradient at the edges of the regions. In the experimental phase, we validate the proposed method using multiple remote sensing images. The experimental results confirm the superiority of the proposed method and demonstrate the significant potential and advantages of feature region matching in the context of lightweight feature databases and remote sensing image registration.**

***Index Terms*—Feature region matching, gradient location-orientation histogram (GLOH), maximally stable extremal regions (MSERs), remote image registration, scale-invariant feature transform (SIFT).**

Manuscript received 27 October 2023; revised 11 December 2023; accepted 15 December 2023. Date of publication 19 December 2023; date of current version 8 January 2024. This work was supported by the Key Research Program of Frontier Sciences, Chinese Academy of Sciences, under Grant ZDBS-LY-JSC036. *(Corresponding author: Feng Wang.)*

Zilu Zhao and Hongjian You are with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China, and with the Key Laboratory of Technology in Geo-spatial Information Processing and Application System, Chinese Academy of Sciences, Beijing 100190, China, and also with the School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 101408, China (e-mail: zhaozilu16@mails.ucas.ac.cn; hjyou@mail.ie.ac.cn).

Feng Wang is with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China, and also with the Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, Chinese Academy of Sciences, Beijing 100190, China (e-mail: wangfeng003020@aircas.ac.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3344474

I. Introduction

REGISTRATION of remote sensing images is the process of aligning remote sensing images of the same scene taken from various perspectives, periods, or sensors [1], [2]. It plays an important role in applications, such as change detection [3], image fusion [4], and target detection [5]. Moreover, it is necessary to analyze remote sensing data and obtain remote sensing information.

Geographic registration in remote sensing requires ground control points (GCPs) to ensure consistent geometric positioning across images. The traditional method involves manually collecting and analyzing GCPs, which can be time-consuming, labor-intensive, and requires duplication of effort. To make better use of existing data, we can establish a control point database, which facilitates quick point matching [6]. Nevertheless, maintaining a control point database is complex and expensive. The development of local invariant features now allows us to automatically extract accurate control information from remote sensing images and geometrically align them with the target image [7]. Therefore, certain researchers propose substituting high-resolution reference images and control point databases with lightweight feature databases, such as Ji et al. [8] proposed a method to generate lightweight feature databases based on feature vectors from satellite imagery, achieving lightweight treatment and storage for descriptors. Lightweight feature databases require consideration of feature stability, especially when dealing with images with significant differences in time, perspective, and sensor. Therefore, it is crucial to store features in the database that demonstrate robustness and high reproducibility.

Feature extraction methods are fast and robust in detecting salient features, such as points or regions, from an image rather than complete information [9]. In addition, features are described subsequently matched based on the similarity of their descriptors. The most prevalent methods for feature point matching are scale-invariant feature transform (SIFT) [10] and its variations [11], such as speeded up robust feature (SURF) proposed by Bay et al. [12], uniform robust SIFT (UR-SIFT) proposed by Sedaghat et al. [13], and KAZE proposed by Alcantarilla et al. [14]. In addition, there are enhanced approaches specialized for SIFT descriptors, such as gradient location-orientation histogram (GLOH) [15], DAISY [16], and adaptive binning SIFT (AB-SIFT) [17].

In recent years, researchers have tackled the issue of intensity differences in multimodal remote sensing images by developing novel feature descriptors that utilize phase congruency (PC) [18]. Ye et al. [19] developed the histogram of oriented phase congruency, which leverages PC eigenvalues and orientations to better capture the geometric structure of the image. Fan et al. [20] developed a PC structural descriptor by extracting PC features at various scales. Li et al. [21] proposed the radiation-variation insensitive feature transform, which identifies feature points in PC image and employs descriptors on the maximum index map derived from the phase coherence magnitude. It is important to note that the majority of these techniques predominantly depend on corner point extraction algorithms, such as FAST [22] and Harris [23], as the primary method for feature detection.

Feature point matching methods often extract features from key points correlated with local gradients and corner points. However, they face challenges in reproducing accurate results in multitemporal multiview remote sensing image matching due to variations in image textures, changes in perspective, and noise. It is often necessary to increase the number of feature points to improve the chance of correct matching. The crucial aspect in constructing a feature control database is to ensure effective matching by storing a limited number of features, thereby achieving stable feature matching across a wide range. In this context, region features possess certain advantages over point features. Matching region features considers a group of related pixels as a whole, providing more contextual information. Homogeneous regions are less sensitive to variations in perspectives, illumination, or other conditions, displaying similar structures [24]. By selectively sampling feature regions, comparable results can be achieved with fewer regions than by matching a large number of feature points. This reduces storage and computation overhead, enhancing the overall matching efficiency. Thus, this article delves into issues concerning the registration of stable feature regions.

Feature region detection algorithms typically identify regions in an image that exhibit affine invariance and high contrast, such as lakes, reservoirs, buildings, and shadows [25]. The detected region features have high invariance and stability, which can be repeatedly detected in most images, and are complementary to other detectors [26]. Common region detection algorithms include methods, such as intensity extrema-based regions [27] and edge-based regions proposed by Tuytelaars and Van Gool[28], and maximally stable extremal regions (MSERs) proposed by Matas et al.[29]. MSER is widely used and known for its stability in detecting areas. The method applies the concept of watershed to threshold an image with different gray levels so that the image undergoes a process from all black to all white, and in the process multiple homogeneous extreme regions are extracted using the stability threshold as a criterion.

MSER is essentially an intensity-based region detector that processes connectivity components and extracts extreme regions that are stable to intensity perturbations, and the ideal image for this method is a well-structured homogeneous image with separated intensity variations [25]. Therefore, many improvements to MSER focus on information enhancement of the image. For example, Liu et al. [30] proposed an edge-enhanced MSER (EMSER), which performs MSER on the image after combining the original image and the edge image. Martins et al. [31] proposed the feature-driven MSER (fMSER) that extracts MSERs on a salient feature map where boundary-related features are highlighted, providing an improvement over standard MSER in terms of completeness. Xin et al. [32] performed MSER for spatial detection of urban roads on gradient images obtained by the Sobel operator. There are also some scholars who have enhanced MSER by incorporating scale invariance. For example, Forssen and Lowe [33] introduced a technique for extracting MSERs across multiple scales. Śluzek [34] extended the MSER detection mechanism to the 2-D space with scale-insensitive MSERs. However, most existing methods do not take into account the problem of nested overlapping regions between MSERs, which can lead to the extraction of much greater numbers of MSERs than the number of homogeneous regions visually seen after MSER detection. Excessive overlap between feature regions diminishes their differentiation, thereby limiting the effective utilization of features and compromising the accuracy of feature matching.

Several of the region feature extraction methods based on MSER employ texture descriptors, such as gradient direction histograms or phase direction histograms, to describe the acquired feature regions [30], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43]. Some methods calculate the information in the entire feature region to construct descriptors. As an example, Liu et al. [35] proposed the maximally stable PC, which computes the histogram of phase feature directions within a circular region. In contrast, some other approaches focus solely on the centers of feature regions as points of interest and employ texture descriptors for their characterization. For instance, the hyperspectral image MSER introduced by Ordóñez et al. [37] constructs SIFT descriptors within an area 16×16 pixels centered on the region. However, most of the papers simply combine MSER and SIFT without explicitly specifying whether the description is performed on the feature region or on the feature centroid neighborhood. Nevertheless, the use of texture descriptors may have a significant effect on the performance and matching results. MSERs extracted by the MSER are mostly uniform and stable homogeneous regions, which may lack texture information and gradient information by themselves; it is inherently contradictory to use texture descriptors to characterize the regions that are not rich in texture information. Particularly for descriptions of region centroid neighborhoods, which are typically homogeneous, the uniqueness of descriptors is greatly reduced. Therefore, it is important to clearly articulate and explain the range of feature region descriptors in the papers.

In this article, we propose the robust region feature extraction with salient MSER (SMSER) and segment distance-weighted GLOH (SDGLOH), addressing the demand for stable features in lightweight feature databases and improving abovementioned MSER redundancy overlap and the application of texture descriptors to homogeneous regions. For feature detection, we proposed SMSER to improve the stability of extracted regions as well as to remove overlapping feature redundancy. For feature description, SDGLOH is proposed to enhance the characterization of significant homogeneous regions by focusing on their edginess and wholeness. Finally, region matching uses

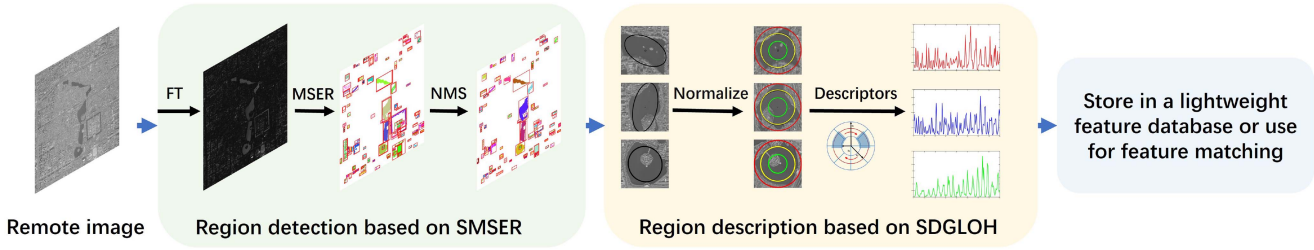


Fig. 1. Schematic diagram of the proposed feature region extraction method.

the nearest-neighbor distance ratio (NNDR) [10] combined with fast sample consensus (FSC) [44] to filter error matching [45], [46]. Our proposed enhanced region matching method reduces the number of extracted features while enhancing their reproducibility, resulting in improved performance in feature matching. Although reducing the number of features may lead to information loss, and careful selection and processing of these features enhance their stability, which is crucial for efficient matching in lightweight feature databases. A schematic of the flow of the method in this research is shown in Fig. 1.

This article has the following three main contributions.

1) Facing the need for stable features in lightweight feature databases, we propose the robust region feature extraction with SMSER and SDGLOH for remote sensing image registration.
2) For feature detection, SMSER is proposed. Initially, MSER is applied to the salient map acquired through frequency-tuned (FT) salient region detection [47], emphasizing crucial image components and enhancing the distinctiveness and reproducibility of MSERs. Later, nonmaximum suppression (NMS), relying on shape complexity, is applied to remove redundant and intricate areas, ensuring the validity and stability of chosen features.
3) For feature description, SDGLOH is proposed. We augment the statistical subregion blocks of GLOH, broadening the statistical scope to capture more extensive regional contextual information. Subsequently, the segmented distance-weighted method and the overlap statistical technique are proposed to enhance the edge gradient weights and mitigate edge effects, ensuring the precision and distinguishability of the descriptors.

The rest of this article is organized as follows. Sections II and III describe the detection and description of our proposed region extraction method, respectively. Sections IV and V conduct experiments and discuss the results in depth, respectively. Finally, Section VI concludes this article.

II. Salient MSER

To improve the existing problem of nested overlapping of MSERs and provide MSER with intensity variations separated and well-structured images, we propose SMSER. SMSER is able to effectively extract SMSERs and reduce regional overlapping and information redundancy by combining FT and NMS. First, FT is used to create a salient map from the original remote sensing image, then MSER is performed on the salient map, NMS based on shape complexity is used to filter the regions, and finally, MSERs with excellent visual contrast and fairly regular shape are created.

A. FT Salient Region Detection

In this article, the saliency map of a remote sensing image is produced using FT. FT analyzes the image from a frequency perspective. In general, the high-frequency portion of the image represents the detail information of the image, whereas the low-frequency portion of the image typically reflects the overall information of the image. Our goal is to obtain the low-stable homogeneous region in the image, but we also need to keep the high frequency in order to obtain a clear boundary. In the actual calculation, FT uses Gaussian smoothing of the 5×5 window to roundoff the highest frequencies and remove the fine texture and noise in the image. Subsequently, the luminance feature is used to estimate the contrast of the image region with respect to the surrounding environment, and a salient map with clear boundaries and full resolution can be generated. The specific formula for calculating the salient map S can be seen in the following equation:

$$S(x, y) = ||I_\mu - I_{\omega hc(x,y)}||_2 \tag{1}$$

where I_μ is the arithmetic mean of the whole image and $I_{\omega hc}$ is the Gaussian blur image of the original image after being blurred with a 5×5 Gaussian kernel, which is used to both remove noise and fine textures. $\|\|$ is the L_2 norm.

To make the contrast between the salient regions and the background stronger, a simple gamma correction is then performed on the salient map that has been created. The gamma correction can be seen in the following equation:

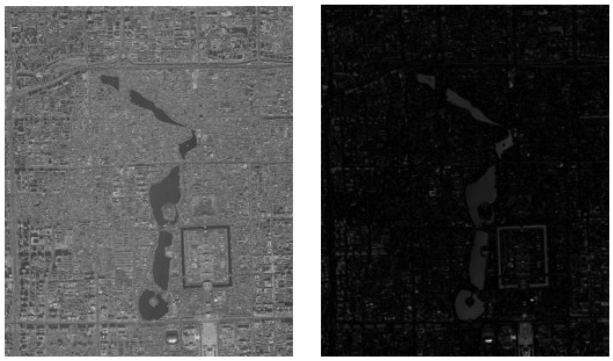
$$s = cr^\gamma. \tag{2}$$

Here, r is the original input value of the grayscale image and s is the grayscale output value after gamma correction. c is the scaling factor. γ is the gamma factor, which is used to control the degree of scaling of the entire correction.

Fig. 2 shows the comparison schematic of the salient map of an area in Beijing after FT and gamma correction. It is evident that saliency detection accentuates the visually prominent regions of the image.

B. Maximally Stable Extremal Regions

Next, MSER is performed on the salient map. The salient map obtained by FT is taken as the input image $S : D \subset N^2 \to T$



(a) (b)

Fig. 2. Salient region detection before and after comparisons. (a) Test image of Beijing area. (b) Salient map of test image.

(D stands for pixel space). T takes values between 0 and 255, and a series of thresholds $t \in T$ is selected from T to segment the image into binary images [32]

$$B_t = \begin{cases} 1 & , S \geq t \\ 0 & , \text{other.} \end{cases} \tag{3}$$

Q is the region of all extremes consisting of connected B_t. Let $Q_1, Q_2, \ldots, Q_{i-1}, Q_i, Q_{i+1} \ldots$ form the sequence of extremes $Q_i \subset Q_{i+1}, i = 1, 2, \ldots$, with the rate of change of the area $q(i)$ as follows

$$q(i) = \frac{|Q_{i+\Delta} - Q_{i-\Delta}|}{|Q_i|}. \tag{4}$$

Here, Q_i represents the area of the ith connected region and Δ represents a small threshold change. When the value of $q(i)$ is less than the given threshold, the region is recognized as the ER. It is also necessary to invert the original image and then perform the binarization process with the threshold value from 0 to 255. The two operations before and after are also known as MSER+ and MSER-, respectively.

C. Nonmaximum Suppression

There is still a lot of overlap between the feature regions obtained after MSER on the salient map. These overlapping regions can be removed by NMS to prevent too large overlapping information between nested MSERs [48].

In remote sensing image registration, the task of matching feature regions with more complex shapes is often more challenging. Especially in the elliptical regions obtained after fitting, the centering of complex regions may have a large deviation relative to regions with simpler shapes. Therefore, we employ the shape complexity of the feature region as an index for the filtering of NMS, and set the reciprocal of each region's shape complexity as the response value of that region, then, the smaller the shape complexity of the region, the higher the response value obtained. This reduces the number of nested regions and selects regions with simple shapes for easy matching after filtering the MSERs by NMS. The calculation of shape complexity is shown in the following:

$$C = \frac{L^2}{A}. \tag{5}$$

Here, L is the perimeter of the feature region and A is the area of the feature region.

Specifically, the response values of all MSERs are sorted in a descending order, the region with the highest score from the sorted feature regions is selected as the reference region, and the remaining regions are traversed one by one in a descending order. If the intersection over union (IoU) of the currently traversed region and the reference region exceeds the set threshold, the two regions are highly overlapped and the region is deleted. Following this, the above operation is repeated until all regions are traversed, and finally, we obtain a set of relatively independent, nonintersecting feature regions. These regions exhibit high saliency and representativeness, effectively reducing the redundancy of feature information. In addition, another score threshold C_t is set while performing NMS, and when the score of the traversed region is less than C_t, it indicates that the shape complexity of the feature region is high, and the feature region is immediately removed. The calculating process of NMS is shown in the following:

$$s_f = \begin{cases} s_i, & \text{IoU}(R, b_i) < N_t \\ 0, & \text{IoU}(R, b_i) \geq N_t. \end{cases} \tag{6}$$

Here, b_i indicates the ith detected ER, s_i denotes the raw score of b_i, and s_f corresponds to the final score of b_i. R represents the region detected with the highest score in each round of the detection cycle, while $\text{IoU}(R, b_i)$ signifies IoU of b_i and R. Moreover, N_t is a predetermined filtering threshold.

Fig. 3 depicts a schematic illustration of the comparison between MSER and SMSER conducted on the test image. The diagram distinctly illustrates that the feature regions derived from SMSER correspond to salient areas perceptible to the human eye. These regions exhibit relatively regular and simple shapes, with minimal overlap observed among different regions.

III. Segment Distance-Weighted GLOH

After completing the process of extracting feature regions, to effectively address the extracted feature region information and improve the contradiction of the existing texture descriptors applied in homogeneous regions, we introduce a novel SD-GLOH. In this article, GLOH is selected to describe MSERs in place of SIFT and is improved through two key aspects. First, we augment the count of subregions for the statistical analysis performed by GLOH and expand its statistical scope. This facilitates the acquisition of a richer set of regional contextual information, enabling a more comprehensive representation of feature region characteristics. Second, we implement a segmented distance-weighted and overlap statistical approach for subregion gradient information based on concentric circles. Consequently, this enhances the weighting of the edge gradient, mitigates edge effects, and captures additional information within the feature region, thus elevating the stability and reliability of the matching process.

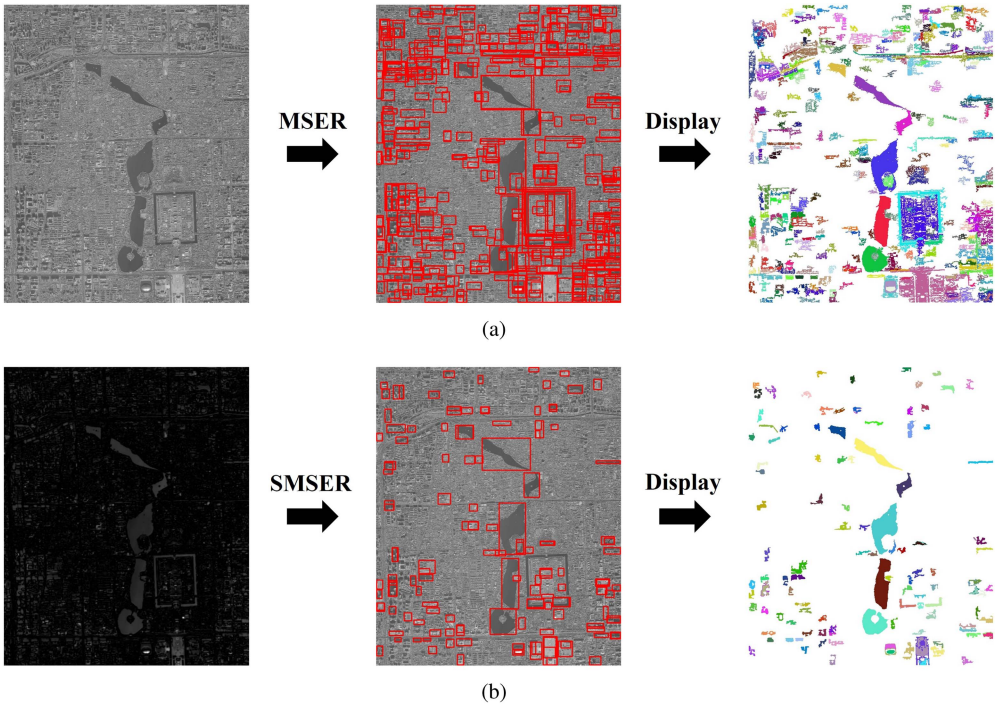


Fig. 3. Schematic comparison of MSER and SMSER. (a) MSER. (b) SMSER.

A. Gradient Location-Orientation Histogram

GLOH represents an extended variant of SIFT. Compared with the neighborhood square statistics of SIFT, GLOH transforms the neighborhood subblocks in standard SIFT into affine concentric circles in log-polar coordinates. In GLOH, the concentric circle radius ratios are set to 6, 11, and 15, respectively, and divided into eight equal parts in the angular direction so that together with the circular subblock in the center a total of 17 image subregions are produced. After that, the gradient direction is then divided into 8 or 16 direction intervals, and the pixel gradient direction in each subregion is counted, using a method similar to the Gaussian-weighted gradient method in SIFT. GLOH demonstrates enhanced capability in capturing and portraying the holistic attributes of the circular region postnormalization of the elliptical region fitted by MSERs. It also exhibits greater adaptability to the shape attributes of the acquired region.

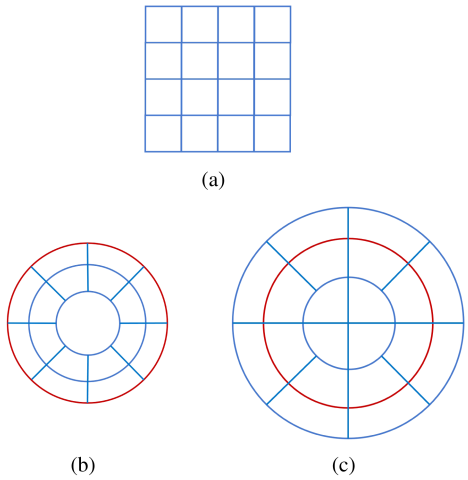


Fig. 4. Schematic of the different descriptors. (a) SIFT. (b) GLOH. (c) SDGLOH.

B. Descriptor Statistics Improvement

Considering the limited internal texture within the MSERs extracted via MSER and the irregularities in the contour of the region, we expand the statistical range of GLOH. While adhering to the 6:11:15 ratio for the feature region, we use the original ellipse normalized circle as the intermediate concentric circle to expand the statistical description of the region. The expansion of the statistical range of the gradient information of the MSERs is conducive to obtaining more texture information and context information of the region, increasing the matching accuracy.

In particular, when describing regions using the logarithmic pole structure, if the inner circular region is not divided into radial sectors, this process reduces the discriminatory power of the descriptor due to the loss of some spatial information. As a result, we proceed to subdivide the central circular region into four subregions, forming 20 subregions. As shown in Fig. 4, the different descriptors are schematically depicted. Then, for each subregion, the gradient histograms in eight directions are counted. Eventually, a 160-D feature region descriptor is produced by combining the information of these subregions.

The method proposed in this study involves normalizing the extracted elliptical region by utilizing the elongated axis of the ellipse as the normalization direction. Hence, the orientation of the ellipse is not considered when computing the final statistical descriptors. Fig. 5 illustrates the schematic representation of the statistical analysis performed on GLOH and SDGLOH during

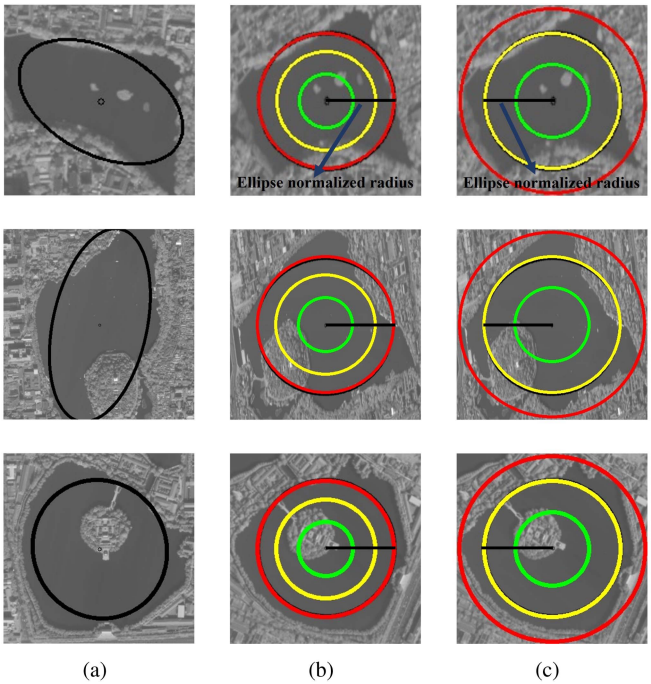


Fig. 5. Elliptic normalization and schematic representation of the statistical region of descriptors. (a) Region slices before normalization. (b) Region slices after normalization and GLOH statistical regions. (c) Region slices after normalization and SDGLOH statistical regions.

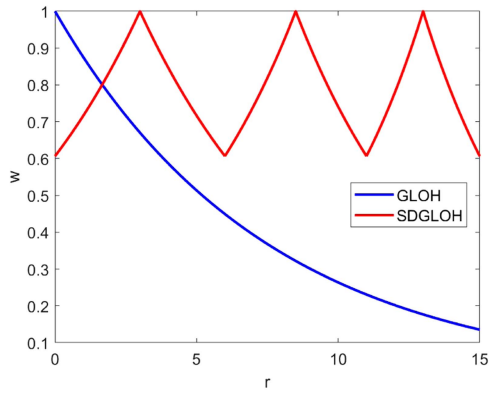


Fig. 6. Schematic of descriptor forward and inverse weighting when $m(\cdot)$ is 1.

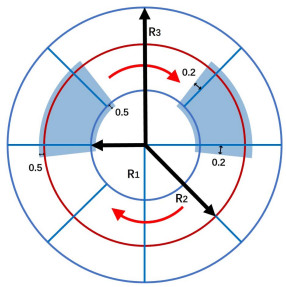


Fig. 7. SDGLOH descriptor statistical diagram.

experimentation on the test image. In this figure, the colors green, yellow, and red correspond, respectively, to concentric circles of the 6:11:15 ratio. It can be observed that each concentric circle layer has captured more gradient information, while the outermost concentric circle contains a greater amount of gradient information closer to the edge of the region.

C. Segmented Distance-Weighted and Overlap Statistical

The conventional Gaussian weighting method applied to gradient magnitudes in traditional SIFT and GLOH amplifies the homogeneous information inside the region even more when dealing with MSERs that possess thin internal texture information alongside rich edge gradient characteristics. Hence, SDGLOH not only improves the statistical approach of GLOH but also modifies the way of assigning Gaussian weights. We applied a segmented distance-weighted method, allocating weights within each concentric circle, to reduce the influence of missing texture information in the center of homogeneous regions and increase the contribution of edge information. Equations (7)–(9) correspond to Gaussian weighting approaches that we used for the descriptors SIFT, GLOH, and SDGLOH, respectively. w is the magnitude of the weighted gradient

$$w = m(a+x, b+y) * e^{-\frac{(x')^2+(y')^2}{2\times(0.5^{*}d)^2}}. \tag{7}$$

Here, $m(\cdot)$ denotes the gradient amplitude, (a, b) refers to the coordinates of the statistical point within the global coordinate system, (x, y) signifies the coordinates of the statistical point within the local coordinate system centered on the feature point, and (x', y') represents the coordinates of the rotated sampling point within the local coordinate system. The number of blocking regions for SIFT statistics is $d \times d$, d is typically 4.

$$w = m(a+x, b+y) * e^{-\frac{r}{R_3 * 0.5}} \tag{8}$$

$$w = m(a+x, b+y) * \exp\left(-\frac{\left|r - \left(\frac{R_i+R_{i+1}}{2}\right)\right|}{R_{i+1}-R_i}\right), \quad i = 0, 1, 2, 3 \tag{9}$$

where r denotes the polar length of the point, and R_0 is 0. R_1, R_2, and R_3 are the radius ratios set in GLOH, which are 6, 11, and 15, respectively. Fig. 6 presents a schematic of the Gaussian weighting function used by the descriptor, with the red line depicting our specific Gaussian weighting function. The disparity between the two weighting methods is evident, and our weighting method preserves the percentage of edge information.

Moreover, we refer to M-SURF [49] utilized in KAZE and introduce an overlapping gradient statistics approach to SDGLOH circular descriptor. This modification aims to alleviate boundary effects and enrich the information captured within the feature region, as shown in Fig. 7. The radius ratio of the three concentric circles is 6:11:15. It signifies the relative sizes of the circles. In the range with a radius of 6, which corresponds to the first concentric circle, the direction along the radius axis remains constant. While the direction along the radius axis for the final statistical range is set to 6 in the region between the first and second concentric circles. This subblock extends 0.5 units on both sides, ranging from 5.5 to 11.5 units, as shown in Fig. 7. The

TABLE I
OVERVIEW OF COMPARATIVE METHODS

Detector	MSER	Feature region detection using MSER on images
	EMSER	Feature region detection using EMSER on images
	fMSER	Feature region detection using the proposed fMSER on images
	SMSER	Feature region detection using the proposed SMSER on images
Descriptor	GLOH-R	Describe the entire feature region using GLOH
	SDGLOH-R	Describe the entire feature region using the proposed SDGLOH
	AB-SIFT-R	Describe the entire feature region using AB-SIFT
	DAISY-R	Describe the entire feature region using DAISY
	SIFT-R	Describe the entire feature region using GLOH
	SIFT-P	Use SIFT only to describe the central neighborhood of the feature region of approximately 16×16 pixels
	M-SURF-R	Describe the entire feature region using M-SURF
	M-SURF-P	Use M-SURF only to describe the central neighborhood of the feature region of approximately 16×16 pixels

same pattern is also applied for the third concentric circle, with an extension of 1 unit on both sides. Similarly, an overlapping statistics strategy is applied on the angular axis, which expands the angular range of each subregion by 0.2 units on either side of the angular axis. After statistical analysis, 160-D descriptors are obtained to provide a comprehensive description of MSERs.

IV. EXPERIMENT AND RESULT ANALYSIS

In this section, the effectiveness of our proposed method will be comprehensively evaluated through experiments conducted on multisource optical remote sensing images. Furthermore, a comparative analysis will be conducted between our method and other existing techniques to provide a comprehensive assessment. The primary comparison focuses on the stability of features extracted using our methods and their subsequent matching performance. The specific combinations of comparative methods used in the experiment are presented in Table I.

A. Image Data

We have selected multioptical remote sensing images captured from different perspectives and at different times by multiple satellite sensors. Fig. 8 shows the approximate coverage of the images. Subsequently, these remote sensing images were used for experiments on feature region repeatability and feature region matching. Detailed parameters used for the different experiments including the sources, dates, scales, viewing perspectives, and resolutions of the images are documented in Tables II and IV.

B. Experimental Parameters and Evaluation Criterion

This article presents two experiments designed to validate the effectiveness of the proposed method. In these experiments, the

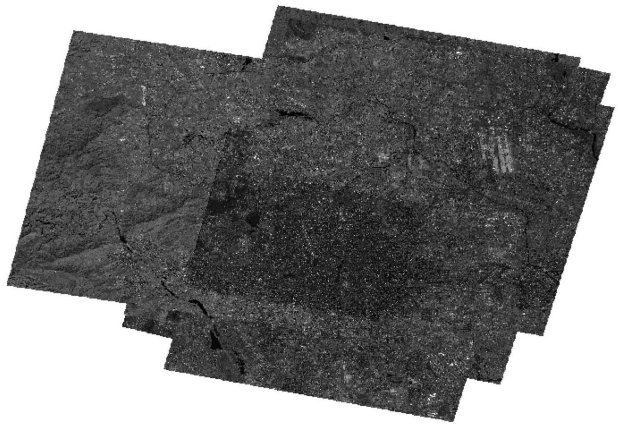


Fig. 8. Schematic illustration of the optical images coverage in the experiments.

TABLE II
SPECIFIC PARAMETERS OF THE REMOTE SENSING IMAGES USED IN THE EXPERIMENT ON FEATURE REPEATABILITY

Image No.	Source/Date	Size(pixels×pixels)	Roll angle (°)	Resolution(m)
1	JL-1/20191203	28651×28720	11.27	0.75
2	JL-1/20200323	28676×28674	-4.99	0.75
3	JL-1/20200621	28676×28630	8.97	0.75
4	GF-2/20161209	27619×29200	0	0.81
5	GF-2/20161223	27620×29200	-10.20	0.81
6	GF-2/20221017	29276×29200	-9.01	0.81
7	GF-1/20191027	18194×18194	0	2.00
8	GF-1/20220507	18236×18194	0	2.00
9	GF-1/20220906	18236×18190	-20.00	2.00

TABLE III
FEATURE REPEATABILITY CALCULATION RESULTS

Image No.		2	3	4	5	6	7	8	9
Repeatability With Image 1	MSER	0.169	0.110	0.242	0.236	0.309	0.300	0.274	0.210
	EMSER	0.219	0.152	0.224	0.233	0.337	0.280	0.277	0.212
	fMSER	0.187	0.103	0.262	0.279	0.379	0.318	0.332	0.233
	SMSER	**0.343**	**0.214**	**0.353**	**0.346**	**0.412**	**0.481**	**0.488**	**0.403**

Bold values represent the highest repeatability obtained in the experiments with our proposed SMSER.

TABLE IV
SPECIFIC PARAMETERS OF THE REMOTE SENSING IMAGES USED IN THE EXPERIMENT ON FEATURE MATCHING

Image Pairs No.	Source/Date	Size(pixels×pixels)	Roll angle (°)	Resolution(m)
1	GF-1/20191027	18194×18194	0	2.00
	JL-1/20200323	28676×28674	-4.99	0.75
2	GF-2/20161209	27619×29200	0	0.81
	JL-1/20200323	28676×28674	-4.99	0.75
3	GF-1/20231001	18236×18194	0	2.00
	GF-1/20220906	18236×18190	-20.00	2.00
4	GF-2/20161209	27619×29200	0	0.81
	GF-2/20221017	29276×29200	-9.01	0.81
5	GF-2/20230522	29276×29200	-4.00	0.81
	GF-1/20191203	18236×18190	7.71	2.00
6	GF-2/20190703	29276×29200	0	0.81
	GF-1/20220825	18236×18194	-9.50	2.00

remote sensing images were standardized to a 4-m resolution and processed in blocks with a block size of 2500 × 2500 pixels for the matching area. The number of feature regions detected by MSER is controlled by defining a threshold value. In our case, we specified the MSER area range between approximately 900 and 62 500 pixels, meaning that the dimensions of the regions fall within the range of about 30–250 pixels. In addition, the upper limit for the region change rate of MSER was set at 0.35. Referring to EMSER, fMSER used in this article enhances the information by overlaying the boundary feature map with the original map. Except for the proposed SDGLOH, the parameters of the descriptors use the default values of the corresponding methods, except for the statistical region range. The remaining relevant parameter values are $c = 1$, $\gamma = 0.75$, $N_t = 0.85$, and $C_t = 0.01$.

Evaluation metrics employed in this study are the following.

1) *Repeatability:* It is used to quantify the similarity of features extracted by different methods. A higher feature repeatability indicates greater stability and reproducibility of features extracted by the same method across the multitemporal and multiview conditions of remote sensing images. Repeatability is defined as the ratio between the number of repeated feature regions (N_r) and the total number of feature regions (N_{orig})

$$\text{Repeatability} = \frac{N_r}{N_{\text{orig}}}. \tag{10}$$

This article utilizes the intersection ratio IoU to quantify the similarity of regional features, where a higher value of IoU indicates a stronger regional feature similarity

$$\text{IoU}(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{11}$$

where *A* and *B* represent the individual feature regions extracted from each of the two images.

2) *Correct matching ratio (CMR):* It is employed to gauge the overall performance of the feature matching algorithm, primarily assessing accuracy. In this article, CMR is defined as the ratio of the number of total matches (TM) acquired after eliminating false matches to total number of feature regions N_{orig}. A higher CMR corresponds to an improved overall method performance

$$\text{CMR} = \frac{\text{TM}}{N_{\text{orig}}}. \tag{12}$$

3) *F-measure:* It is used to evaluate the reconciliation of precision and recall. A higher F-measure value indicates better matching performance

$$F - \text{measure} = 2 \times \frac{\text{Precision} \times \text{recall}}{\text{Precision} + \text{recall}}. \tag{13}$$

Precision is used to evaluate the effectiveness of the matching results. The precision rate is defined as the ratio of the number of correct matches (CM) obtained by the algorithm to the number of TM

$$\text{Precision} = \frac{\text{CM}}{\text{TM}}. \tag{14}$$

Recall is used to assess the algorithm's performance in terms of matching completeness. Recall is defined as the ratio of the number of CM obtained by the algorithm to the total correspondences (TC) originally associated between the two images

$$\text{Recall} = \frac{\text{CM}}{\text{TC}}. \tag{15}$$

4) *Root mean square error (RMSE):* It is used to reflect the geometric localization accuracy of the feature matching method. The lower the RMSE, the higher the matching accuracy of the adopted method

$$\text{RMSE} = \sqrt{\frac{1}{\text{NCM}} \sum_{i=1}^{\text{NCM}} \left(\left(x_i^r - x_i^{t'}\right)^2 + \left(y_i^r - y_i^{t'}\right)^2 \right)}. \tag{16}$$

Here, (x_i^r, y_i^r) denotes the coordinates of the feature points in the reference image, while $(x_i^{t'}, y_i^{t'})$ signifies the corresponding coordinates of the matching points after correction.

C. Experiment on Feature Repeatability

The objective of this experiment is to assess the stability and reproducibility of feature regions extracted using different feature detection methods on remote sensing images captured at various time phases and viewing perspectives. The experimental dataset comprises all the images listed in Table II. Detailed steps are outlined as follows.

1) *Data preprocessing:* Image 1 (which is in the center of the selected image set) was selected as the reference image in Table II (each image with certain overlapping regions). All other images are prealigned with reference to the geographic location of Image 1, after which the latitude and longitude of the experimental area are set according to the overlapping latitude and longitude of all images, which facilitates the calculation of the subsequent repeatability.
2) *Feature extraction:* Various feature detection methods are employed to extract feature regions within the predefined latitude and longitude regions of each image.
3) *Calculation of repeatability:* Repeatability is calculated by comparing Image 1 with each remaining image for its features. Locate the centers of feature regions in other images within a 5×5 pixels neighborhood of the center of each feature region in Image 1 and record their positions. Subsequently, IoU is calculated between the searched feature region and the selected region. If there is a region with an IoU larger than the specified threshold, the count of feature region repetitions in Image 1 is incremented by one. Image 1 is traversed to identify all feature regions using the aforementioned method. Subsequently, the feature repetition rate is calculated. (To prevent double counting of regions that overlap at the same location due to the redundancy of MSERs, NMS is applied to each feature extraction method.)

Fig. 9 shows a schematic diagram of the calculation of the overlap area for two different remote sensing images. The degree of overlap for different IoU values can be seen in the figure. In this study, regions with an IoU greater than 0.6 are considered larger overlapping regions and are used for subsequent repeatability calculations. The experimental results will be presented in Table III and visualized in Figs. 10 and 11. Furthermore, an effective quantitative and qualitative analysis will be provided on the basis of the experimental results.

Fig. 11 presents a schematic diagram illustrating the overlapping area between the selected reference image (Image 1) and an image block from Image 7 under different feature detection methods. In this depiction, the green ellipse delineates the extracted region from Image 1, while the red ellipse highlights the overlapping zone between Image 1 and Image 7 (IoU > 0.6). The schematic image block comes from Image 1. From the diagram, it is evident that our enhanced SMSER approach considerably reduces the extracted region compared with other

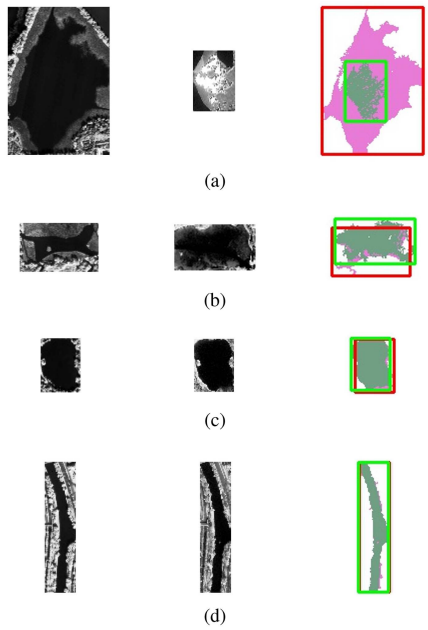


Fig. 9. Schematic of overlapping areas with different IoU values. (a) IoU $= 0.17$; (b) IoU $= 0.59$; (c) IoU $= 0.78$; (d) IoU $= 0.94$.

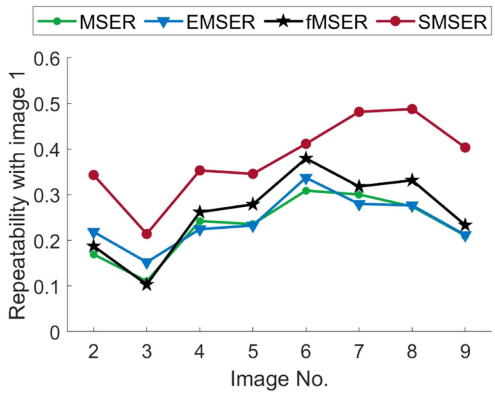


Fig. 10. Schematic representation of the repetition rate of feature regions in multiple images with Image 1.

methods. Nevertheless, the distribution of the overlapping red regions remains consistent with the previous three techniques, indicating the high reproducibility of the SMSER-detected regions.

Referring to Fig. 10, it is clear that when calculating the repeatability of feature detection methods for various remote sensing images under identical reference image conditions, SMSER demonstrates superior performance. The observed high feature region repetition rate reveals the high reproducibility of our extracted features across varying imaging conditions. This not only enhances matching opportunities but also augments the reliability of the matching process. Therefore, it can be regarded as an ideal choice for remote sensing image matching and registration. The application of these highly reproducible features to the feature database would significantly enhance its stability. Refer to Table III for specific repeatability data.

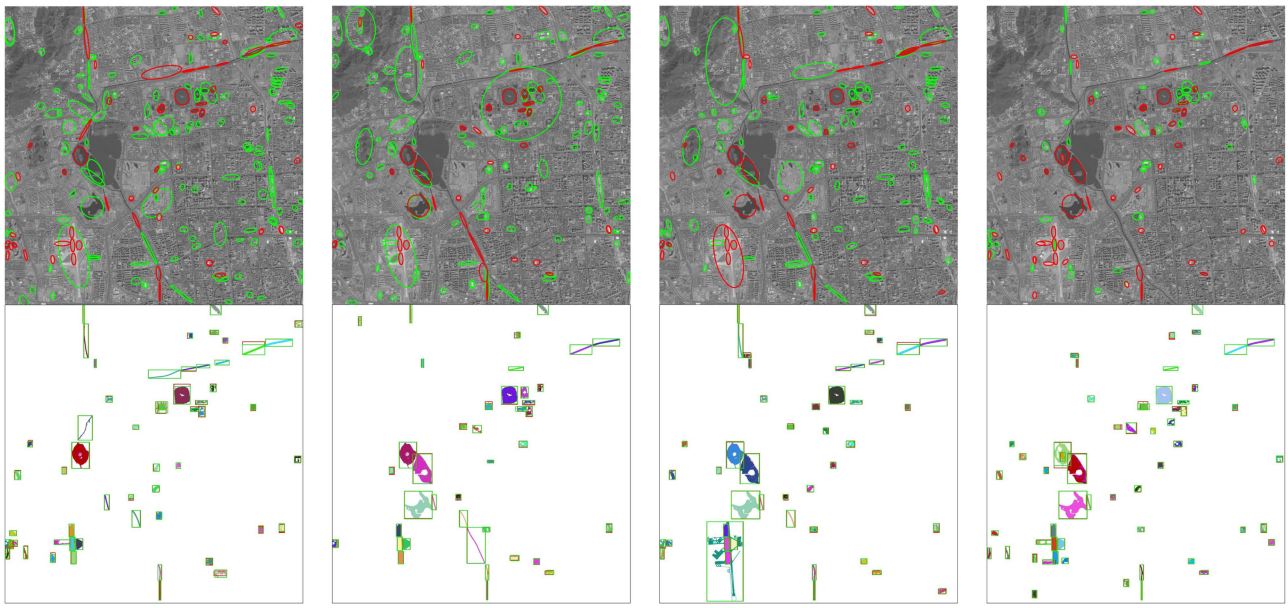


Fig. 11. This schematic representation illustrates the duplicate feature regions in Image 1 blocks derived from different feature detection methods. (a) MSER. (b) EMSER. (c) fMSER. (d) SMSER.

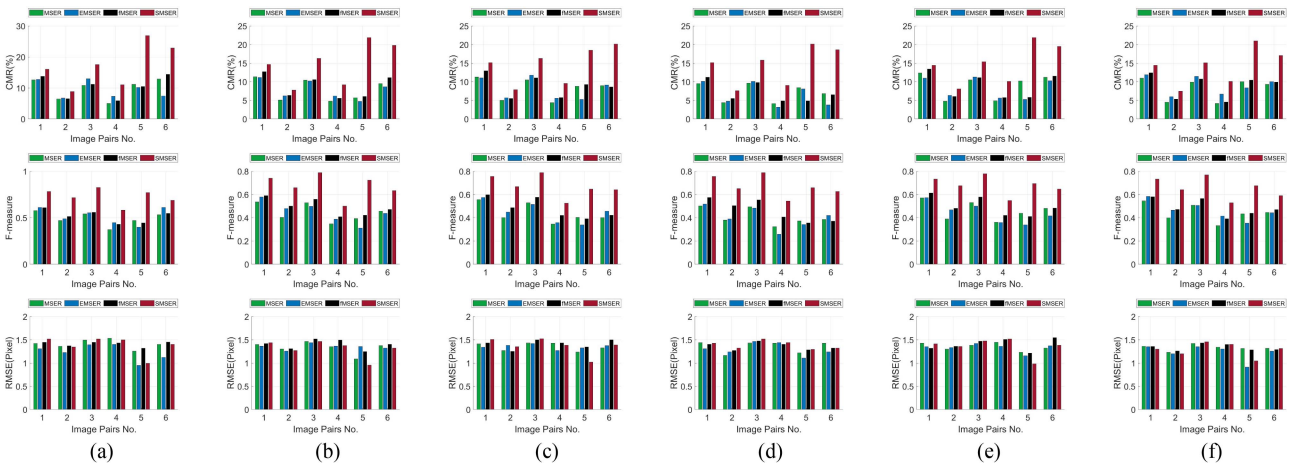


Fig. 12. Schematic representation of matching results using different feature detection methods under the same descriptor scenario. (a) SDGLOH-R. (b) GLOH-R. (c) AB-SIFT-R. (d) DAISY-R. (e) SIFT-R. (f) M-SURF-R.

D. Experiment on Feature Matching

The objective of this experiment is to evaluate the performance of the proposed region feature extraction method, i.e., the performance of the features when they are used in a matching task. Detailed steps are described below.

1) *Data preparation:* Six pairs of images for matching experiments are selected and summarized in Table IV. These image pairs demonstrate considerable variations in temporal phases and viewing perspectives.
2) *Feature region matching:* For each set of image pair, we perform feature region matching using various feature detection and description techniques. The similarity between feature region descriptors is assessed through distance measurements, and false matches are filtered out using a combination of NNDR and FSC methods [44].
3) *Results evaluation:* The matching performance of region features under various combinations of feature detection and description methods is assessed using CMR, F-measure, and RMSE metrics.

The experimental results will be presented in Figs. 12–14, and we will provide a comprehensive analysis based on these results.

First, six pairs of images are used to evaluate the performance of different feature detection methods under the same feature region descriptors. Specific results are depicted in Fig. 12. Fig. 12 shows the matching results of the six image pairs based on CMR, F-measure, and RMSE.

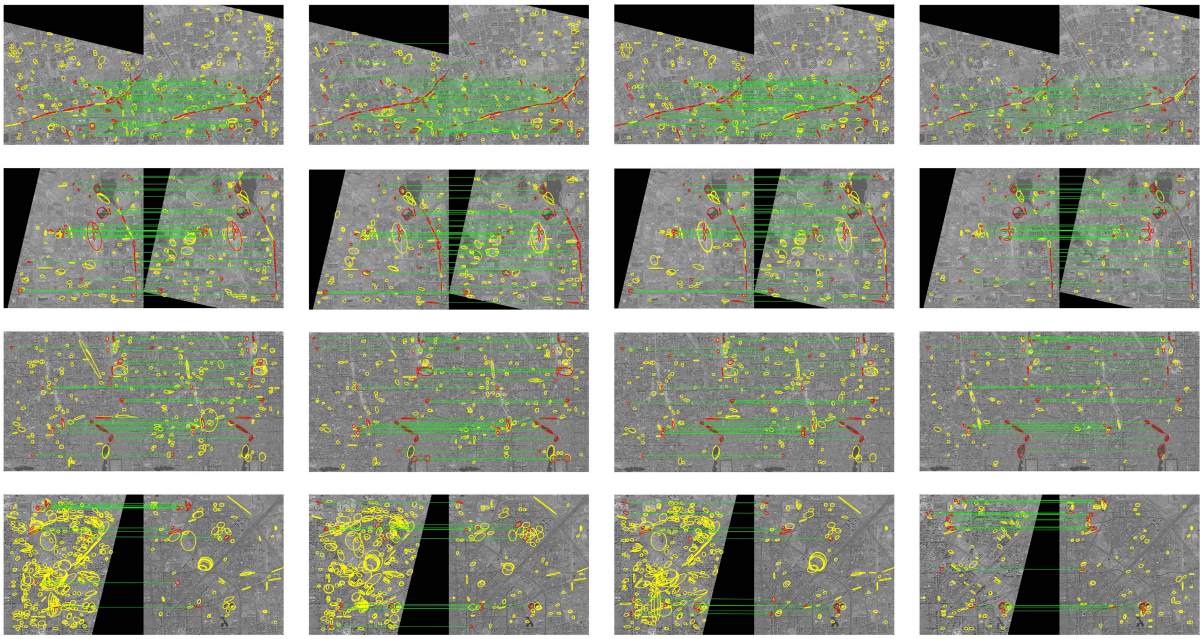


Fig. 13. Schematic representation of the matching results in the same block using different extraction methods (using the SDGLOH descriptor). (a) MSER. (b) EMSER. (c) fMSER. (d) SMSER.

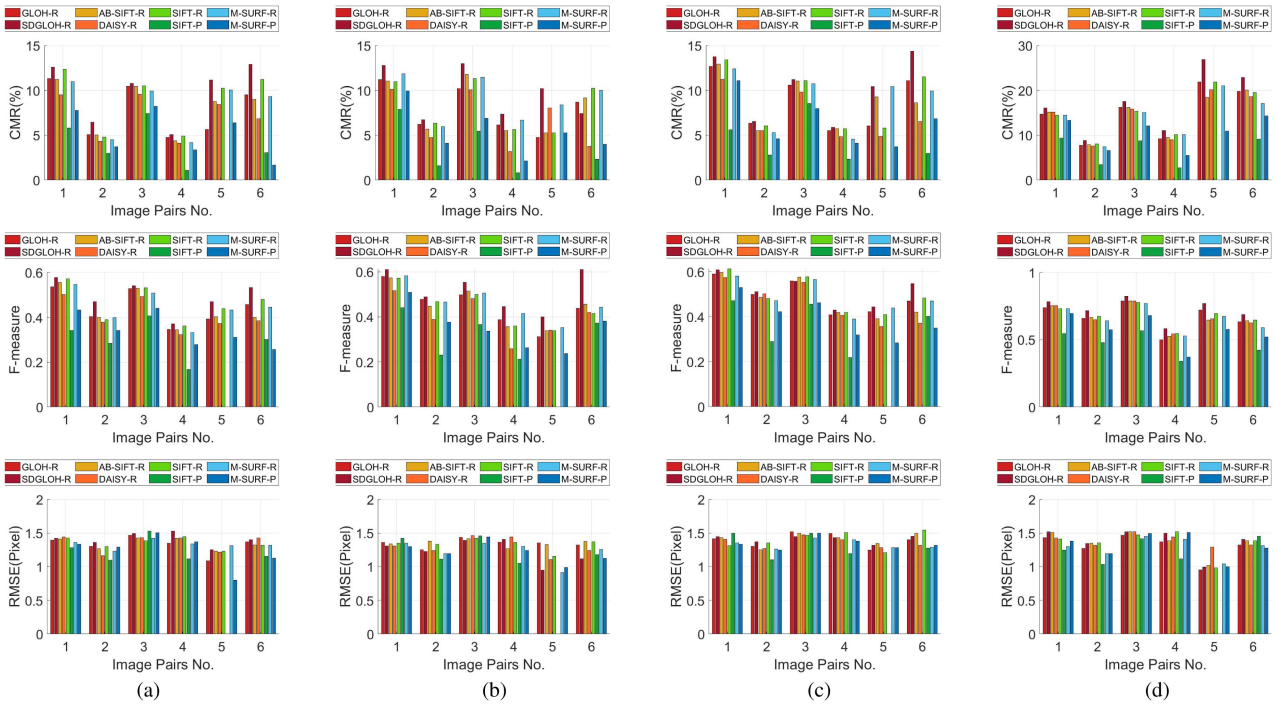


Fig. 14. Schematic representation of matching results using different descriptors under the same feature detection method scenario. (a) MSER. (b) EMSER. (c) fMSER. (d) SMSER.

It is evident that SMSER outperforms the original detector, EMSER, and fMSER in terms of CMR in every image pair, yielding improved matching results. The accuracy rate of feature region matching is significantly enhanced. Before the improvement, MSER had more redundant and overlapping information, bringing some difficulties to the matching process. Through our improvement, the quality of the obtained feature regions is significantly improved and the distinguishability and stability are enhanced, indicating superior performance in the matching task.

We can also clearly observe from the figure that SMSER demonstrates a higher F-measure in the performance metrics.

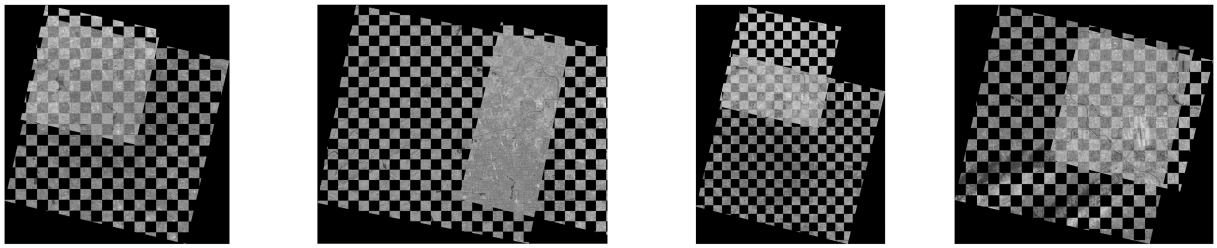


Fig. 15. Partial image registration results using our method SMSER and SDGLOH. (a) Image Pair 1; (b) Image Pair 3 (c) Image Pair 5; (d) Image Pair 6.

F-measure is a comprehensive performance evaluation metric that considers two critical aspects: recall and precision. A high F-measure value indicates that SMSER outperforms other algorithms in terms of finding all features while maintaining high matching precision. This underscores the algorithm's excellent performance in the matching task.

The registration accuracy depicted in Fig. 12 shows that the RMSE obtained using SMSER is similar to the pre-improvement results, without any discernible difference. This suggests that the enhanced method does not significantly improve matching accuracy. The accuracy of registration is affected by the inherent errors introduced by the region detection method. For instance, at the boundaries or texture-rich regions, the extraction results may suffer from certain levels of blurriness and misalignment, and fitting ellipsoidal regions may also generate some errors. These errors will directly affect the region registration accuracy. To address the issue of accuracy, we can enhance performance by reducing the area of the extracted feature regions. This aspect will be evaluated in subsequent experiments to determine appropriate region sizes. We can optimize the performance of region feature matching based on practical requirements and experimental results.

Fig. 13 illustrates the matching schematic for different detection methods across four distinct latitude and longitude range image blocks in the matching experiments. The yellow elliptical regions represent the initial MSERs extracted from the image blocks by the feature detection methods, while the red elliptical regions represent the final matching regions (utilizing SDGLOH for matching). It is evident from the figures that SMSER exhibits a lower count of regions compared with MSER, EMSER, and fMSER. However, they have a more regular shape, making them easier to match. Despite the smaller number of features that we extracted, the feature regions that are ultimately successfully matched by the four methods demonstrate high consistency in terms of similarity. This indicates that our methods effectively eliminate redundant information from feature regions while preserving valuable matching data.

Overall, the proposed SMSER detector consistently outperforms and yields favorable matching results across the majority of image pair groups sharing the same descriptor.

Under the same feature detection method, Fig. 14 presents the results of matching six image pairs using different descriptors, evaluating CMR, F-measure, and RMSE metrics. The evaluation metrics CMR and F-measure of the SDGLOH are slightly better than other methods for most comparisons, emphasizing the algorithm's exceptional performance in the matching task. Especially when combined with our method, i.e., SMSER-SDGLOH, the highest CMR and F-measure are achieved in all image pairs. However, compared with other methods, our approach shows no significant difference in terms of RMSE, indicating that our method does not have a negative impact on registration accuracy. Fig. 15 depicts four representative scenarios from our selected image pairs, showcasing the successful registration accomplished by our proposed method.

In addition, it is evident that describing the entire region leads to greater stability and a higher correct matching rate, as opposed to solely describing the center point of the region. Specifically, we can observe and compare the use of SIFT-R, M-SURF-R, SIFT-P, and M-SURF-P under the same feature region detection method. This is attributed to the fact that the whole region is described to obtain more contextual and edge gradient information, which is more discriminative and feature-expression rich. While only describing the centroid neighborhood cannot capture all the characteristics of the region, it will also lose the uniqueness of the descriptor due to the homogeneity of the region, which may lead to instability of the matching and the reduction of the correct matching rate.

We observe that sometimes the point-neighborhood descriptor demonstrates superior registration accuracy compared with the region descriptor. This is because point-neighborhood descriptors focus solely on the region centroid's neighborhood, and the successful matches with point-neighborhood descriptors have minimal position deviation from the region centroid. On the contrary, the region descriptor must consider the gradient information of the entire region. The successful matching of the region descriptor with the center point can be influenced by the fitting of the region, leading to bias and potentially decreasing registration accuracy. This observation is also an indirect result of the fact that point features typically demonstrate higher registration accuracy than region features.

In summary, from the above analysis we can draw the following conclusions.

1) When the same descriptors are used, the feature region detected by SMSER demonstrates improved stability and correctness in matching.
2) When utilizing the same feature region detection method, SDGLOH achieves higher correct matching rates and F-measure values compared with other descriptors in most cases.

3) The whole region description matches better than the center of the region description.
4) In multiview and multitemporal remote sensing images, feature region matching, although slightly less accurate than feature point matching, can provide sufficient registration accuracy for most applications.

V. Discussion

In this article, we propose the robust region feature extraction method combining SMSER and SDGLOH. FT and NMS filtering are optimized first for MSER to ensure the saliency of the extracted feature regions, and then SDGLOH, which is more suitable for the region, is used to describe the feature regions. The performance of our method is validated on multiple pairs of multisatellite optical remote sensing images with different time phases and different perspectives. In the comparison of multiple sets of image registration experiments, our proposed method demonstrates high repeatability, correct matching rate, and ensures registration accuracy.

Our experimental method is more applicable to remote sensing images that contain a certain number of significant homogeneous regions, a property related to the nature of FT. The saliency map is applicable to images with salient regions. A saliency region is a part of an image that is striking and clearly different from its surroundings, and these regions may contain features, such as prominent objects, colors, textures, or shapes. In Fig. 15, the mosaic of our method-aligned images illustrate discernible homogeneous areas within the shared latitude and longitude region. These areas, such as vacant airports and homogeneous water bodies, are distinctly differentiated from the surrounding regions on grayscale. If the image lacks recognizable saliency regions, the FT algorithm's information enhancement will be of little use and will result in an image that lacks closed homogeneous region boundaries. For images lacking clear boundaries and saliency regions, the proposed method may encounter difficulties in extracting and matching features, resulting in unsatisfactory matching results. Therefore, our experimental method can be applied more effectively to the region matching of remote sensing images under specific conditions.

Meanwhile, our proposed improved region matching method provides crucial support for the construction of a stable lightweight feature database. The core advantage of this approach is that it maintains a high degree of feature stability while reducing the number of features and is more suitable for region description. This enables us to create and maintain lightweight feature databases with greater efficiency, thereby enhancing the performance of feature matching in scenarios with multiple temporal and view variations.

Afterward, we hope to continue to improve the feature region descriptors, not only focusing on SIFT- and GLOH-like descriptors, but also exploring the region contour descriptors to further enrich the representation of region features. We also recognize the need to further investigate the relationship between the feature area of the MSERs and the registration accuracy. This series of improvements and extensions will help further improve the performance of our method in complex and variable scenes and provide strong support for future image matching studies. In future studies, we will explore these issues more deeply.

VI. Conclusion

In this work, we propose a robust region feature extraction method that combines SMSER and SDGLOH. The core advantage of this method lies in its ability to reduce the number of features while maintaining a high level of stability and suitability for regional description. The performance of our method is validated on multiple pairs of multisatellite optical remote sensing images with different time phases and different perspectives. Meanwhile, our proposed enhanced region matching method significantly contributes to the construction of a stable lightweight feature database.

References

[1] A. Wong and D. A. Clausi, "ARRSI: Automatic registration of remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1483–1493, May 2007.
[2] D. Xiang, Y. Xie, J. Cheng, Y. Xu, H. Zhang, and Y. Zheng, "Optical and SAR image registration based on feature decoupling network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Oct. 2022, Art. no. 5235913.
[3] Y. Zhong, W. Liu, J. Zhao, and L. Zhang, "Change detection based on pulse-coupled neural networks and the NMI feature for high spatial resolution remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 537–541, Mar. 2015.
[4] Q. Xu, Y. Li, J. Nie, Q. Liu, and M. Guo, "UpanGAN: Unsupervised pansharpening based on the spectral and spatial loss constrained generative adversarial network," *Inf. Fusion*, vol. 91, pp. 31–46, Mar. 2023.
[5] Y. Li, Q. Xu, Z. He, and W. Li, "Progressive task-based universal network for raw infrared remote sensing imagery ship detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, May 2023, Art. no. 5610013.
[6] Q. H. Chen, X. G. Liu, W. Gao, and T. L. Liu, "An automatic ground control point matching based on GCP chip database for remote sensing images," in *Proc. Int. Conf. Image Anal. Signal Process.*, 2009, pp. 13–17.
[7] P. Tang et al., "Framework of remote sensing image automatic processing with "invariant feature point set" as control data set," *J. Remote Sens.*, vol. 20, no. 05, pp. 1126–1137, 2016.
[8] S. Ji, Y. Zhang, Y. Dong, and D. Fan, "Spaceborne lightweight image control points generation method," *Acta Geodaetica et Cartographica Sinica*, vol. 51, no. 03, pp. 413–425, 2022.
[9] R. Feng, Q. Du, X. Li, and H. Shen, "Robust registration for remote sensing images by combining and localizing feature- and area-based methods," *ISPRS J. Photogrammetry Remote Sens.*, vol. 151, pp. 15–26, May 2019.
[10] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
[11] D. Xiang, Y. Xu, J. Cheng, Y. Xie, and D. Guan, "Progressive keypoint detection with dense Siamese network for SAR image registration," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 5, pp. 5847–5858, Oct. 2023.
[12] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 404–417.
[13] A. Sedaghat, M. Mokhtarzade, and H. Ebadi, "Uniform robust scale-invariant feature matching for optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4516–4527, Nov. 2011.
[14] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, "KAZE features," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 214–227.
[15] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.
[16] E. Tola, V. Lepetit, and P. Fua, "DAISY: An efficient dense descriptor applied to wide-baseline stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 815–830, May 2010.
[17] A. Sedaghat and H. Ebadi, "Remote sensing image matching based on adaptive binning SIFT descriptor," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5283–5293, Oct. 2015.

[18] R. Feng, H. Shen, J. Bai, and X. Li, "Advances and opportunities in remote sensing image geometric registration: A systematic review of state-of-the-art approaches and future research directions," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 4, pp. 120–142, Dec. 2021.

[19] Y. Ye, J. Shan, L. Bruzzone, and L. Shen, "Robust registration of multimodal remote sensing images based on structural similarity," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2941–2958, May 2017.

[20] J. Fan, Y. Wu, M. Li, W. Liang, and Y. Cao, "SAR and optical image registration using nonlinear diffusion and phase congruency structural descriptor," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5368–5379, Sep. 2018.

[21] J. Li, Q. Hu, and M. Ai, "RIFT: Multi-modal image matching based on radiation-variation insensitive feature transform," *IEEE Trans. Image Process.*, vol. 29, pp. 3296–3310, Dec. 2020.

[22] D. G. Viswanathan, "Features from accelerated segment test (FAST)," in *Proc. 10th Workshop Image Anal. Multimedia Interactive Serv.*, 2009, pp. 6–8.

[23] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. Alvey Vis. Conf.*, 1988, vol. 15, pp. 147–151.

[24] C. Palmann, S. Mavromatis, and J. Sequeira, "A new geometric invariant to match regions within remote sensing images of different modalities," in *Proc. SPIE Image Signal Process. Remote Sens.*, 2011, vol. 8180, pp. 19–32.

[25] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: A survey," *Found. Trends Comput. Graph. Vis.*, vol. 3, no. 3, pp. 177–280, 2007.

[26] D. Jia, N. Zhu, N. Yang, S. Wu, Y. Li, and M. Zhao, "Image matching methods," *J. Image Graph.*, vol. 24, no. 5, pp. 677–699, 2019.

[27] T. Tuytelaars and L. van Gool, "Wide baseline stereo matching based on local, affinely invariant regions," in *Proc. Brit. Mach. Vis. Conf., Bristol: Brit. Mach. Vis. Assoc.*, 2000, pp. 38.1–38.14.

[28] T. Tuytelaars and L. Van Gool, "Matching widely separated views based on affine invariant regions," *Int. J. Comput. Vis.*, vol. 59, no. 1, pp. 61–85, Aug. 2004.

[29] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image Vis. Comput.*, vol. 22, no. 10, pp. 761–767, Sep. 2004.

[30] L. Liu, H. Y. Tuo, T. Xu, and Z. L. Jing, "Multi-spectral image registration and evaluation based on edge-enhanced MSER," *Imag. Sci. J.*, vol. 62, no. 4, pp. 228–235, May 2014.

[31] P. Martins, P. Carvalho, and C. Gatta, "On the completeness of feature-driven maximally stable extremal regions," *Pattern Recognit. Lett.*, vol. 74, pp. 9–16, Apr. 2016.

[32] L. Xin, J. Song, Y. Chen, and J. Hu, "Robust free-space detection in urban roads based on MSER extraction using gradient images," in *Proc. 37th Chin. Control Conf.*, 2018, pp. 4141–4146.

[33] P.-E. Forssen and D. G. Lowe, "Shape descriptors for maximally stable extremal regions," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, 2007, pp. 1–8.

[34] A. Śluzek, "Improving performances of MSER features in matching and retrieval tasks," in *Proc. Eur. Conf. Comput. Vis. Workshops*, 2016, vol. 9915, pp. 759–770.

[35] X. Liu, Y. Ai, J. Zhang, and Z. Wang, "A novel affine and contrast invariant descriptor for infrared and visible image registration," *Remote Sens.*, vol. 10, no. 4, Apr. 2018, Art. no. 658.

[36] R. Kimmel, C. Zhang, A. Bronstein, and M. Bronstein, "Are MSER features really interesting?," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 11, pp. 2316–2320, Nov. 2011.

[37] Á. Ordóñez, Á. Acción, F. Argüello, and D. B. Heras, "HSI-MSER: Hyperspectral image registration algorithm based on MSER and sift," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 12061–12072, Nov. 2021.

[38] L. Liu, H. Tuo, T. Xu, and Z. Jing, "Multi-sensor image registration using edge-enhanced maximally stable extremal region," in *Proc. 5th Int. Congr. Image Signal Process.*, 2012, pp. 901–905.

[39] J. Guo, H. Sun, and C. Zhu, "Multispectral remote sensing image registration based on maximally stable extremal regions," in *Proc. SPIE Conf. Multispectral Image Acquisition Process.*, 2009, vol. 7494, pp. 265–270.

[40] Q. Zhang, Y. Wang, and L. Wang, "Registration of images with affine geometric distortion based on maximally stable extremal regions and phase congruency," *Image Vis. Comput.*, vol. 36, pp. 23–39, Apr. 2015.

[41] L. Cheng, J. Gong, X. Yang, C. Fan, and P. Han, "Robust affine invariant feature extraction for image matching," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 246–250, Apr. 2008.

[42] Y. Zhang, Y. Guo, and Y. Gu, "Robust feature matching and selection methods for multisensor image registration," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2009, vol. 3, pp. III-255–III-258.

[43] P. Martins, P. Carvalho, and C. Gatta, "Stable salient shapes," in *Proc. Int. Conf. Digit. Image Comput. Techn. Appl.*, 2012, pp. 1–8.

[44] Y. Wu, W. Ma, M. Gong, L. Su, and L. Jiao, "A novel point-matching algorithm based on fast sample consensus for image registration," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 43–47, Jan. 2015.

[45] Y. Xiang, F. Wang, and H. You, "OS-SIFT: A robust sift-like algorithm for high-resolution optical-to-SAR image registration in suburban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3078–3090, Jun. 2018.

[46] H. Zhang et al., "Remote sensing image registration based on local affine constraint with circle descriptor," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Oct. 2022, Art. no. 8002205.

[47] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 1597–1604.

[48] H. Liu, D. Zhu, Y. Zhu, X. Xie, and H. Zhao, "Identification of flight area identification plate based on an improved MSER algorithm," *Int. J. Aerosp. Eng.*, vol. 2022, 2022, Art. no. 8374300.

[49] M. Agrawal, K. Konolige, and M. R. Blas, "Censure: Center surround extremas for realtime feature detection and matching," in *Proc. Eur. Conf. Comput. Vis*, 2008, pp. 102–115.