Robust Descriptor Algorithm Considering the Changing Gray Value Trends Inside Ground Objects for Heterogeneous Optical Image Matching

Li Xue, Yehua Sheng, and Ka Zhang

***Abstract*—Differences in sensor types, resolutions, and imaging conditions can lead to considerable spectral differences in heterogeneous optical remote sensing images and the similarity of scale-invariant feature transform (SIFT) or local self-similarities (LSS) feature descriptors of the same point can be poor. Consequently, we proposed a robust descriptor construction algorithm considering the changing gray values inside ground objects. The main contributions of this article include the following. First, based on the stability of the internal gray value changes of ground objects, we suggest that the change orientations and degrees of gray values of pixels can be used to express the stability of the same area of heterogeneous images, providing the basis for image matching; second, unlike many existing methods that use gradient information to calculate feature orientation and descriptors, the proposed algorithm uses change orientation and degree to calculate the feature orientation and descriptor, enabling it to obtain stable descriptors in image matching with large illumination changes. Experimental analysis of homologous and heterogeneous optical remote sensing images demonstrated the superior stability and capability of the proposed algorithm over commonly used algorithms, including the radiation-invariant feature transform, adaptive binning SIFT, gradient orientation modification SIFT, and LSS algorithms.**

***Index Terms*—Change trend of gray values, feature descriptor, heterogeneous remote sensing image, image matching.**

I. Introduction

AUTOMATIC matching of heterogeneous optical remote sensing images is a critical process in collaborative applications, such as change detection [1], image fusion [2], image mosaics [3], and target recognition [4]. It aims to find the same targets—that is, points, lines, or polygons—between different images, especially features (usually spots and corners) in the same position, and to obtain accurate correlation between images.

Matching algorithms are divided into two categories: traditional algorithms and algorithms based on deep learning. Traditional algorithms can be either pixel- or feature-based [5], [6], [7], [8], [9]. Feature-based matching algorithms are widely used in remote sensing image matching applications because of their strong robustness and small computational overhead [10], [11], [12]. However, due to differences in sensor types, spectral characteristics, and the imaging conditions of heterogeneous optical remote sensing images, the gray values of objects in the same area can be quite different—including a lot of nonlinear changes, which can decrease the robustness of descriptors such as the scale-invariant feature transform (SIFT) [13], speeded-up robust features [14], and local self-similarities (LSS) [15]. Although algorithms based on deep learning can achieve good results in heterogeneous image matching, these algorithms often need many training samples, and their matching performance depends on the quality of training samples [16].

Based on the fact that the changing gray value trends within the same ground object in heterogeneous optical images are mostly similar, we propose a stable descriptor construction algorithm to solve this problem—namely, the stable descriptor considering changing trends (SDCT) algorithm. The proposed algorithm can obtain stable descriptors in images with large spectral differences. First, the algorithm assumes that features of images have been extracted using feature-based methods such as the SIFT or FAST algorithms. Second, to improve the robustness of the descriptor, a cubic polynomial is used to fit the 8-neighbors' gray value of each pixel around the features, and the maximum and minimum values of a polynomial are used to calculate the change orientations and gray values of pixels. Edge information with serious degrees of change is removed to avoid its influence. Third, to solve the problem of image rotation, 2PI is divided into 36 parts (on average), and the cumulative change of each part's orientation after Gaussian weighting is calculated. The orientations with greater changes are taken as the orientation of the feature, and the pixels around the feature are rotated based on the feature orientations. Finally, the rotated

Manuscript received 28 May 2023; revised 5 August 2023 and 2 September 2023; accepted 18 September 2023. Date of publication 28 September 2023; date of current version 23 October 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 42271342 and Grant 42071364, in part by the Key Fund of the National Natural Science Foundation of China under Grant 41631175, and in part by the Natural Science Foundation of Jiangsu Province under Grant BK20201372. *(Corresponding authors: Yehua Sheng; Ka Zhang.)*

The authors are with the Key Laboratory of Virtual Geographic Environment, Ministry of Education, Nanjing Normal University, Nanjing 210023, China, also with the State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), Nanjing 210023, China, and also with the Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China (e-mail: 1151617653@qq.com; shengyehua@njnu.edu.cn; zhangka81@126.com).

Digital Object Identifier 10.1109/JSTARS.2023.3320552

images are divided into 4×4 grids, and the 2PI is divided into eight parts (on average). The cumulative change of each part in each unit after Gaussian weighting is calculated to form a 128-D vector, which can then be normalized to avoid the influence of noise.

The main contributions of this study include the following. First, based on the stability of the internal gray value changes of ground objects, we suggest that the change orientation and pixel gray values can be used to express the stability of the same area of heterogeneous images, providing the basis for image matching; second, unlike many existing methods that use gradient information to calculate feature orientation and descriptors, the proposed algorithm uses change orientation and degree to calculate the feature orientation and descriptor, enabling it to obtain stable descriptors in image matching with large illumination changes. Finally, experiments using homologous and heterogeneous images demonstrate the robustness of this algorithm to be higher than comparative algorithms.

The rest of this article is structured as follows. Section II briefly analyzes the improved algorithms related to feature descriptors. Section III introduces the premise behind the proposed algorithm, outlining its functionality in detail. Section IV summarizes the experiments and discusses the proposed algorithm, amongst others. Finally, Section V concludes the article.

II. Related Works

Researchers have conducted many studies in the field of image matching, proposing various algorithms to improve the stability of descriptors. Among them, the SIFT and LSS descriptors exhibit a certain adaptability to spectral changes, so many scholars have proposed related algorithms to improve the image matching effectiveness.

The SIFT algorithm uses the gradient orientation and values of pixels to construct descriptors, giving it a degree of stability when faced with illumination changes. However, the spectrum of remote sensing images from heterogeneous sources can be quite different (with nonlinear changes in many areas), leading to considerable differences in the gradient orientation and values in the same image areas [17]. Consequently, researchers improved the robustness of descriptors by reducing the orientation of features, changing the structure of descriptors, and preprocessing images. For example, Yi et al. [18] proposed the gradient orientation modification SIFT (GOM-SIFT) algorithm. To improve the stability of SIFT descriptors, the algorithm reduces the number of gradient orientations and uses scale restrictions to improve the precision of matching results. However, the scale restriction method is only applicable to initial matching pairs with a high outlier rate. Sedaghat and Ebadi [19] adopted a new descriptor structure to replace the rectangular structure in SIFT to overcome the problem of local viewpoint distortion of images. However, the descriptor still used gradient information for calculations, making it sensitive to the spectral changes of images. Since gradients are sensitive to spectral changes, the question is "Can other stable information be used to construct descriptors?" Calonder et al. [20] proposed the binary robust independent elementary features (BRIEF) algorithm by comparing the gray values of corresponding positions around features, thus reducing the influence of spectral changes on descriptors. However, this method does not have rotation invariance. To solve the problem, Rublee et al. [21] proposed the ORB (oriented FAST and rotated BRIEF) algorithm, which uses the intensity center to calculate the orientation of features. In recent years—because phase coherence information can effectively express the geometric structure information of images but contains less spectral information—it exhibits high stability to spectral changes. Consequently, several methods have been proposed to improve the stability of descriptors based on phase consistency information. For example, Ye and Shen [22] proposed the histogram of an oriented phase coherence algorithm, which uses phase information instead of gradient information. Moreover, Li et al. [23] extracted features by using phase coincidence information and proposed a maximum index graph to improve the stability of descriptors. However, it could take considerable time to calculate the phase coincidence information, which primarily contained the boundary information of ground objects, leading to the phenomenon that most of the matching pairs were concentrated at the edge of ground objects, affecting the distribution quality of the matching pairs; Gao et al. [24] proposed the partial main orientation map, which is used to improve the stability of descriptors to rotation and scale; Zhu et al. [25] proposed a rotation-invariant maximum index map to improve the invariance of descriptors to rotation. However, when the image contains many complex ground objects, the matching effect of descriptors based on MIM is not ideal; Zhou et al. [16] used multiscale convolutional gradient features to achieve image matching by a fast template scheme.

The LSS algorithm calculates the descriptor vector by selecting the largest correlation value in the local area, giving it high stability against color changes [15]. However, the LSS algorithm itself does not have rotation and scaling invariance abilities. Moreover, its recognition ability is low, being unable to reliably distinguish and match various features [7], [17]. Consequently, researchers have proposed several improvements. For example, Ye et al. [8] prematched images using the scale restriction SIFT (SR-SIFT) algorithm to obtain the rotation and scaling relationship, before matching features with LSS descriptors. However, the SR-SIFT algorithm is sensitive to spectral changes. Liu and Zeng [26] proposed a method for calculating the orientation of the LSS descriptor, which used correlation values instead of gradient information, as well as absolute values of difference instead of the square difference between pixel intensities and center intensities. However, the algorithm not only replaces statistical information but also reduces the stability of the LSS descriptor for spectrum changes. Sedaghat and Mohammadi [7] proposed a descriptor based on extended self-similarity, which improved the uniqueness of the descriptor by calculating self-similarity in multiple orientations. However, it used uniform robust SIFT to select features, which could lead to local aggregation of matching pairs.

Although researchers have proposed several improved algorithms based on the SIFT or LSS algorithms, the performance of these descriptors on heterogeneous images with different

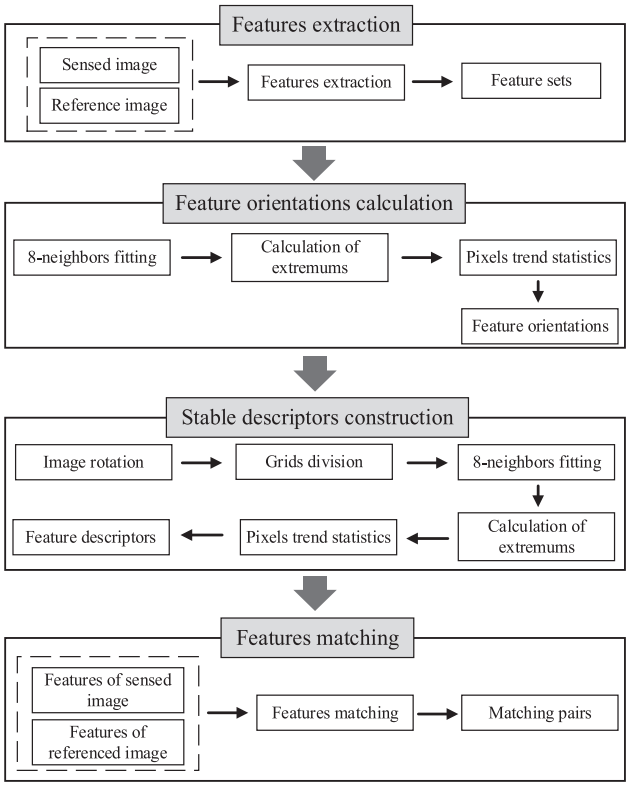


Fig. 1. Flowchart of the proposed algorithm.

spectral features is still far from satisfactory. Improved SIFT-related algorithms increase the stability of SIFT descriptors from an orientation and structure perspective, but these improved methods still use gradient information that is sensitive to illumination changes. Although some researchers have used frequency domain information instead of spatial domain information—which can achieve better results—the computational cost and memory requirements of frequency domain information are enormous. The improved LSS method improved the stability of the LSS descriptor by increasing the orientation of features, calculating related information in multiple orientations, and expanding the description subregion. However, it is still based on illumination intensity that can be sensitive to nonlinear illumination changes in heterogeneous images.

III. Methods: Construction of Stable Descriptors Considering Changing Gray Value Trends

A. Main Idea

This study proposes a robust descriptor construction algorithm considering the changing gray value trends inside ground objects to achieve reliable matching of heterogeneous optical remote sensing images. Fig. 1 shows the workflow of the proposed method—including feature orientation calculation and descriptor construction.

First, a feature extraction algorithm—such as SIFT or FAST—is used to extract the features of the sensed image and reference image. Second, a cubic polynomial is used to calculate

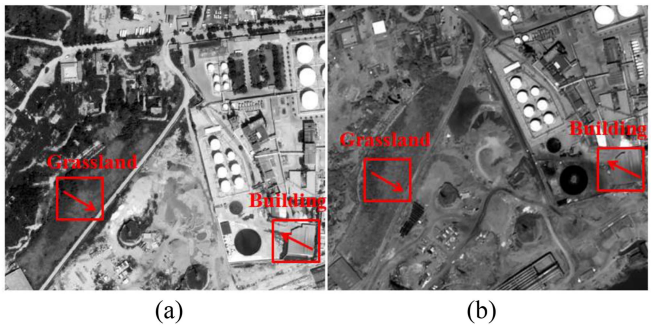


Fig. 2. Change trend of internal gray values of ground objects. (a) First band of QuickBird. (b) Panchromatic band of GF-2.

the change orientation and degree of each pixel's 8-neighbors around the feature, and the boundary pixels with serious change degrees are removed. The 2PI is divided into 36 parts (on average), and the cumulative change degree in each part being counted, and the orientations with larger cumulative change degrees are taken as the orientations of the features. Second, the pixels around the features are rotated based on the orientations of the features, and the rotated pixels are divided into 4×4 regular grids. A polynomial is used to calculate the change orientation and degree of each rotated pixel's gray value, and the edge pixels with serious change degrees are removed. The 2PI is divided into eight parts again (on average), and the cumulative change degree in each part in each grid area is calculated to form a 128-D vector that can be normalized to generate a descriptor. Finally, the nearest neighbor matching method is used to obtain matched pairs.

B. Feature Orientation Calculation

In heterogeneous optical remote sensing images, the gray values trends inside most of the same ground objects are stable, so the stable feature orientations can be calculated based on this characteristic. Fig. 2(a) and (b) shows QuickBird and GF-2 images of the same area, respectively. The two images have serious spectral differences and clear nonlinear spectral changes. However, the change trends of gray values in the same objects in the images are stable. As shown in the grassland and building areas denoted by the box in the images, their gray values gradually increase along the arrow orientation. Based on the above analysis, we proposed a method to calculate the feature direction, including three components—that is, calculating the change orientation and degree, removing the edge pixels, and calculating the orientations of features.

1) Calculating the Change Orientation and Degree: We used the orientation of the global minimum of 8-neighbors' gray values as the change orientation and the ratio of the global maximum and global minimum as the change degree. However, due to the influence of noise and image resolution, it is unstable to use the global maximum and global minimum to calculate the change orientation and degree, yet easy to obtain multiple global maxima and global minimums. Consequently, a polynomial is used to fit the 8-neighbors' gray value, after which the local

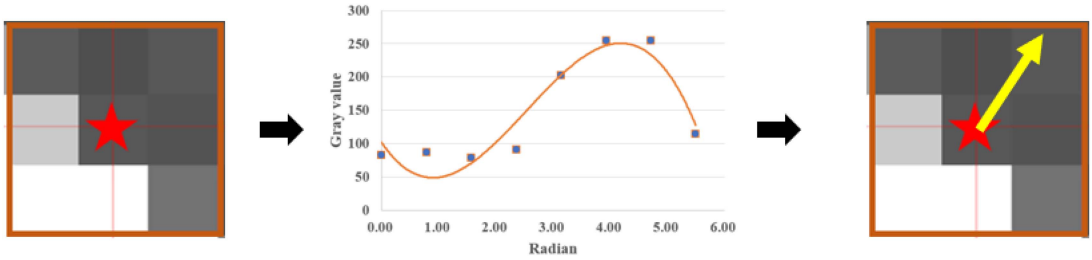


Fig. 3. Calculation flow of change orientation and degree.

minimum and local maximum of the polynomial are used—instead of the global minimum and global maximum—to obtain the accurate change orientation and degree, as shown in Fig. 3 (the pentagram denotes the detected pixel, the orientation of the yellow arrow represents the orientation of the local minimum, and the length of the yellow arrow represents the ratio of the local maximum and the local minimum). To obtain a more stable change orientation and degree, a polynomial is used on the 8-neighbors' gray value to obtain the overall change trend of the 8-neighbors' gray value. We use the orientation of the local minimum as the change orientation and the ratio of the local maximum and the local minimum as the change degree.

Experiments reveal that the stability of change orientation obtained by a cubic polynomial (1) is higher than that obtained by a quadratic polynomial (2). As shown in Fig. 4, Image A and Image B are different images of the same area. Although the gray values change considerably, the changing trend is similar, the gray value gradually decreasing from the bottom-left to the top-right. In Fig. 4, the blue dot represents the 8-neighbors' gray values of the central pixel of Image A, the blue line represents the fitting curve of the 8-neighbors' gray values of the central pixel of Image A, the brown dot represents the 8-neighbors' gray values of the central pixel of Image B, and the brown line represents the fitting curve of the 8-neighbors' gray values of the central pixel of Image B. Fig. 4 shows that the local minimum of the 8-neighbors' gray values of the central pixel of Image A is located in the top-right, and the local minimum of the 8-neighbors' gray values of the central pixel of Image B is located near the top. If accurate orientations are obtained by local quadratic polynomial fitting [see Fig. 4(c)], the orientations of the local minimums of the central pixels of Image A and Image B are 0.73 and 1.49, respectively, which are quite different from one another. If accurate orientations are obtained by cubic polynomial fitting [see Fig. 4(d)], the orientations of the local minimums of the central pixels of Image A and Image B are 0.74 and 0.90, respectively, which are somewhat closer. Consequently, we use a cubic polynomial to calculate the change orientations, as follows:

$$y = a_3x^3 + b_3x^2 + c_3x + d_3 \tag{1}$$

$$y = a_2x^2 + b_2x + c_2 \tag{2}$$

where x denotes the orientation and y denotes the gray value. Starting from the horizontal axis, the gray values of neighbors are obtained in a counterclockwise orientation.

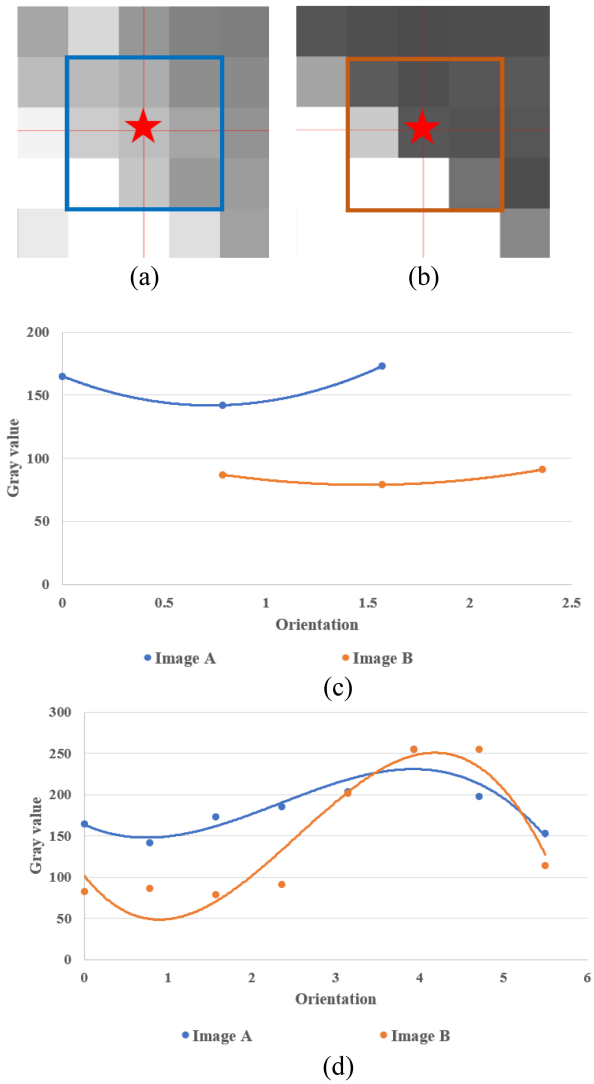


Fig. 4. Analysis of change orientations. (a) Image A. (b) Image B. (c) Quadratic polynomial fitting. (d) Cubic polynomial fitting.

If the 8-neighbors' pixel values are fitted using cubic polynomials based on only one order [see Fig. 5(a)], then only the local minimum or the local maximum can be obtained. Moreover, the obtained extreme value is unstable because the extreme values obtained in different orders differ. Consequently, the 8-neighbors' pixel values are fitted based on four sequences

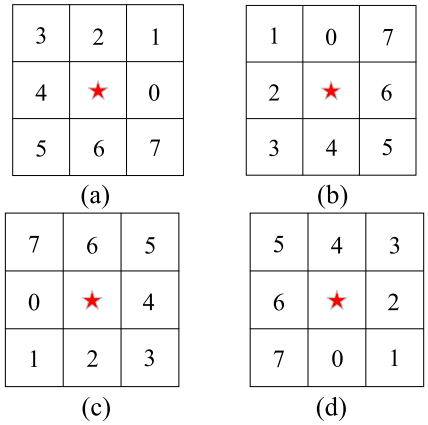


Fig. 5. Sequences of the 8-neighbors' values. (a) First sequence. (b) Second sequence. (c) Third sequence. (d) Fourth sequence.

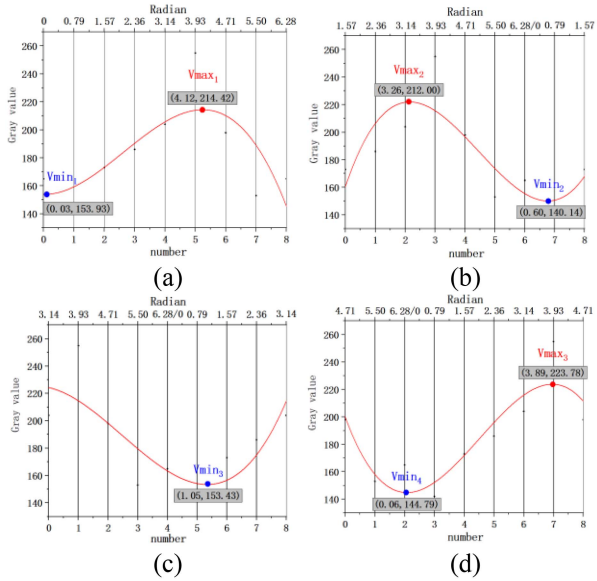


Fig. 6. Polynomial fitting results of four sequences. (a) Polynomial fitting results of the first sequence. (b) Polynomial fitting results of the second sequence. (c) Polynomial fitting results of the third sequence. (d) Polynomial fitting results of the fourth sequence.

[see Fig. 5(a)–(d)], and the extreme values of the four cubic polynomials are obtained. The specific steps of this process are as follows: first, based on the sequence shown in Fig. 5(a) (the red star in the figures represents the pixel to be detected), the gray values of the 8-neighbors are obtained as the ordinate value (y_i), and the corresponding orientations of 0, 0.79, 1.57, 2.36, …, and 5.50 are taken as the abscissa value (x_i) starting from the abscissa, in a counterclockwise direction. Second, a cubic polynomial is used for fitting, and the extreme values of the cubic polynomial are obtained in the interval $x\in[0,2PI]$. The extreme values of the remaining three sequences are obtained using the same steps. We can obtain multiple local minimums and local maxima. The question remains, "How can we use these extreme values to calculate the change orientation and degree?"

There are seven special values: the gray value of smallest local minimum (*Vs*), the gray value of largest local maximum (*Vl*), the gray value of local maximum corresponding to the smallest local minimum (*Vls*), the gray value of local minimum corresponding to the largest local maximum (*Vsl*), the average value of the orientations corresponding to local minimums (*Vsoa*), the average gray value of the local minimum (*Vsa*), and the average gray value of the local maxima (*Vla*). Fig. 6 shows the cubic polynomial fitting results of four sequences of 8-neighborhood pixels. The ordinate represents the gray value, the bottom abscissa represents the sequence number, and the top abscissa represents the corresponding radian. Fig. 6(a) shows that the polynomial fitting function of the first sequence has a local maximum $Vmax_1$ (4.12,214.42) in the interval [0,2π), where 4.12 is the orientation of $Vmax_1$, 214.42 is the gray value of $Vmax_1$, and a local minimum $Vmin_1$ (0.03,153.93). Fig. 6(b) shows that the polynomial fitting function of the second sequence has a local maximum $Vmax_2$ (3.26,212.00) and a local minimum $Vmin_2$ (0.60,140.14) in the interval [0,2π). Fig. 6(c) shows that the polynomial fitting function of the third sequence has only a local minimum $Vmin_3$ (1.05,153.43) in the interval [0,2π). Fig. 6(d) shows that the polynomial fitting function of the fourth sequence has a local maximum $Vmax_3$ (3.89,223.78) and a local minimum $Vmin_4$ (0.06,144.79) in the interval [0,2π).

Since the gray value (140.14) of $Vmin_2$ is the smallest, 140.14 is the gray value of the smallest local minimum (*Vs*). Since the gray value (223.78) of $Vmax_3$ is the largest, 223.78 is the gray value of the largest local maximum (*Vl*). The gray value *Vls* of the local maximum $Vmax_2$ corresponding to the smallest local minimum ($Vmin_2$) is 212.00 (If the smallest local minimum has no corresponding local maximum, then the gray value of the local maximum corresponding to the second smallest local minimum is taken as *Vls*. If the second smallest local minimum still has no corresponding maximum value, then the third smallest local minimum value is checked, and so on. If all the local minimums have no corresponding local maximums, then the 8-neighborhood pixel does not exist *Vls*). Since the largest local maximum ($Vmax_3$) has no corresponding local minimum, the gray value (153.93) of the local minimum ($Vmin_1$) corresponding to the second largest local maximum ($Vmax_1$) is taken as VSL (the searching method of VSL is like that of *Vls*). The orientations corresponding to all local minimums are 0.03, 0.60, 1.05, and 0.06, respectively, so the average value of the orientations (*Vsoa*) is 0.435. The gray values corresponding to all local minimums are 153.93, 140.14, 153.43, and 144.79, respectively, so the average gray value (*Vsa*) is 148.07. The gray values of all local maximums are 214.42, 212.00, and 223.78, respectively, so the average of gray values (*Vla*) is 216.73.

We consider the following four methods to calculate the change orientation and change degree.

1) The orientation corresponding to *Vs* is taken as the change orientation, and the ratio of *Vl* to *Vs* is taken as the change degree.
2) The orientation corresponding to *Vs* is taken as the change orientation, and the ratio of *Vls* to *Vs* is taken as the change degree.

TABLE I
COMPARISON RESULTS OF THE FOUR METHODS

No.	*P* (%)				*Correct Matches*			
	A	B	C	D	A	B	C	D
1	41.43	46.15	**67.50**	64.00	29	30	27	**32**
2	58.54	57.38	57.38	**78.35**	72	70	70	**76**
3	27.66	22.92	54.00	**70.00**	13	11	27	**49**
4	61.43	61.59	59.32	**70.59**	**86**	85	70	72

3) The orientation corresponding to *Vsl* is taken as the change orientation, and the ratio of *Vl* to *Vsl* is taken as the change degree.
4) *Vsoa* is taken as the change orientation, and the ratio of *Vla* to *Vsa* is taken as the change degree.

We used two homologous images and two heterogeneous images for our experiments. The experimental results are shown in Table I, in which the first and second pairs of images are homologous images, and the third and fourth pairs of images are heterogeneous images. Moreover, p denotes the accuracy, and *Correct Matches* denote the number of correct matching pairs, where A denotes the first method, B denotes the second method, C denotes the third method, and D denotes the fourth method. The bold fonts denote the highest accuracy or correct matching pairs. The table shows that the precision of the fourth method is the best overall—that is, 1.20–1.56 times that of the other three methods. The precision of the fourth method is 70.00, but the precision of the first, second, and third methods are low—that is, 27.66, 22.92, and 54.00, respectively—indicating the first, second, and third methods are unstable. Overall, the number of correct matching pairs of the fourth method is better than that of the other three methods. Consequently, the fourth method is used to calculate the change orientation and degree.

Gradient information is often used to calculate descriptors. Therefore, we use two images (I^1, I^2) with a largely spectral difference to compare gradient information with change trend information, to prove that change trend information exhibited higher similarity. Mutual information is often used to calculate the correlation between two sets or events, so this article uses mutual information to measure the similarity [27], [28]—that is, the higher the mutual information, the higher the similarity. First, calculate the gradient orientations (G_o^1, G_o^2), the gradient values (C_v^1, C_v^2), the change orientations (C_o^1, C_o^2), and the change degrees (G_v^1, G_v^2) of two images (I^1, I^2), as shown in Table II. We can then calculate the mutual information of C_o^1 and C_o^2, C_v^1 and C_v^2, G_o^1 andG_o^2, G_v^1 and G_v^2 ($MICo = 0.5679$, $MICv = 0.4717$, $MIGo = 0.3768$, $MIGv = 0.4063$), respectively. The mutual information of change orientations is 0.5769, which is considerably higher than that of the gradient orientation, 0.3768. Moreover, the mutual information of change degree is 0.4717, which is higher than that of the gradient value, 0.4063, revealing that change trend information has higher similarity than gradient information, so change trend information can be used to obtain descriptors of higher stability.

2) Removing Edge and Noise Information: The edge and noise information of ground objects in heterogeneous images is not stable, which can easily affect the stability of descriptors, making it necessary to remove the edge information. Experiments reveal that the change degree of the pixels inside the ground objects is small (usually less than or equal to 2.00), whereas the change degree of the edge and noise pixels with large gray value changes are large (usually more than 2.00). Consequently, based on the characteristic that the change degree of edge and noise pixels change considerably when the change degree of pixels is more than t_v, the pixels are considered as edge and noise information that should be removed, and the remaining pixels are considered as the internal pixels of the ground objects.

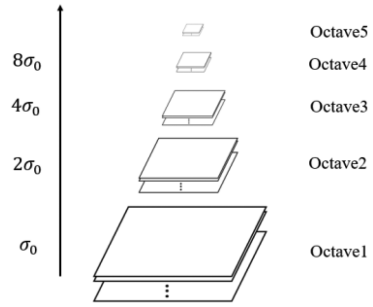


Fig. 7. Gaussian pyramid.

3) Calculating the Orientations of Features: The SIFT algorithm uses different standard deviations of the normal distribution to Gaussian blur the original image (3), and establishes a Gaussian pyramid of N_{oct} layers (Fig. 7, the default value of $\sigma 0$ is 1.6). In order to ensure the continuity of scales, the images of each layer of the pyramid are Gaussian blurred by using different degrees of standard deviation (5), (6). Each layer of the Gaussian pyramid has intvls+3($3 \leq$ intvls ≤ 5) Gaussian blurred images

$$w(x_i, y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-r^2/(2\sigma^2)} \tag{3}$$

$$N_{oct} = \log_2\{\min(H_{img}, W_{img})\} - t\, t \in [0, \log_2\{\min(H_{img}, W_{img})\}) \tag{4}$$

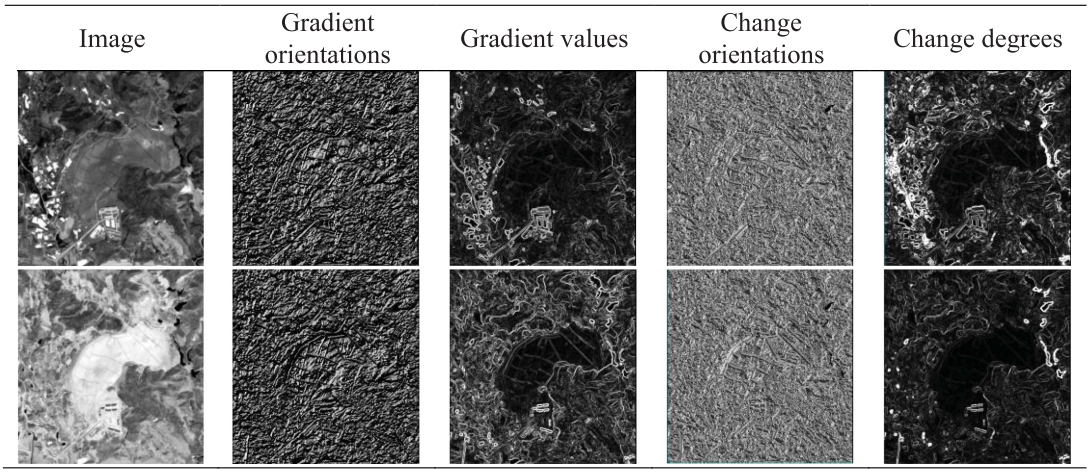
$$k = 2^{1/\text{intvls}} \tag{5}$$

$$sig[n] = \sqrt{(\sigma_0 \times k^{n-1} \times k)^2 - (\sigma_0 \times k^{n-1})^2} \tag{6}$$

where σ is the standard deviation of normal distribution, r is the blur radius, $w(x_i,y_i)$ is the Gaussian weight, H_{img} and W_{img} are the length and width of the image, and $sig[n]$ is the standard deviation of normal distribution of the nth image of each Gaussian pyramid.

The change trend information is used instead of the gradient information to calculate the orientations of features. First, we calculate the change orientation $o(x_i,y_i)$ (8), change degree $v(x_i,y_i)$ (9), and Gaussian weight $w(x_i,y_i)$ (3) of the $j_o \times j_o$ window pixels around the features (7), and calculate the product $vw(x_i,y_i)$ (11) of the change degrees less than the threshold (t_v) and Gaussian weights. Second, 2PI is evenly divided into 36 bins, and the weighted cumulative change degrees in each bin are counted (11) to form a histogram of the gray value change trend. Finally, we find the global maximum of the histogram,

TABLE II
GRADIENT INFORMATION AND CHANGE TREND INFORMATION OF IMAGES



calculate the orientations that are greater than or equal to 0.8 times the global maximum, and perform quadratic polynomial fitting on these orientations to obtain accurate orientations as the features orientations

$$j_o = 3\sigma_i \tag{7}$$

$$o(x_i, y_i) = Osa(f_1(x_i, y_i), f_2(x_i, y_i), f_3(x_i, y_i), f_4(x_i, y_i)) \tag{8}$$

$$v(x_i, y_i) = Vla(x_i, y_i)/Vsa(x_i, y_i) \tag{9}$$

$$vw(x_i, y_i) = v(x_i, y_i) \times w(x_i, y_i), v(x_i, y_i) < t \tag{10}$$

$$b_m = \begin{cases} vw(x_i, y_i), & \text{INT}(o(x_i, y_i)/(2PI/36)) = m \\ 0, & \text{other} \end{cases} \tag{11}$$

where σ_i denotes the scale of the image in the Gaussian pyramid corresponding to the feature (x_f,y_f), $f(x_i,y_i)$ denotes the cubic polynomial, $Osa(f_1(x_i,y_i),f_2(x_i,y_i),f_3(x_i,y_i),$ and $f_4(x_i,y_i))$ denote the average of orientations corresponding to local minimums of the four cubic polynomials, $Vla(x_i,y_i)$ denotes the average of local maxima of the four cubic polynomials, $Vsa(x_i,y_i)$ denotes the average of local minimums of the four cubic polynomials, r denotes the Euclidean distance between the pixel (x_i,y_i) and the feature (x_f,y_f) to be detected, σ denotes the standard deviation of normal distribution, b_m denotes the accumulated $vw(x_i,y_i)$ of the mth bin, and INT(A) denotes the integer of A.

In order to verify the stability of the proposed orientation, two images of the same area [see Fig. 8(a) and (b)] with large spectral differences are used for experiments. First, the SIFT feature algorithm is used to extract the features in the same position of the reference image and the matched image [see Fig. 8(c) and (d)]. Second, the SIFT orientations and the proposed orientations of features are calculated, respectively. Then, the number of features (N_t) whose direction difference is less than t is counted and the ratio (r) of N_t to the number of all features in the same position is calculated. Since the value of t should be small, t is selected from $\{1, 2, 3, 4\}$. The experimental results are shown in Table III. Table III shows that the ratio of the proposed orientation is higher than that of SIFT orientation, which indicates that the number of features with similar orientations is higher than that of SIFT algorithm. Therefore, the stability of the proposed orientation is better than SIFT orientation.

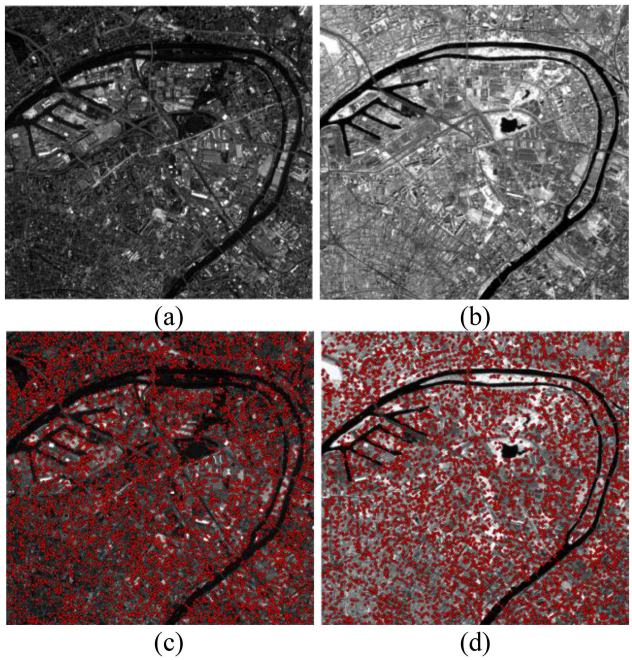


Fig. 8. Features of images. (a) Reference image. (b) Matched image. (c) Features of the reference image. (d) Features of the matched image.

TABLE III
RATIO OF FEATURES WITH SIMILAR ORIENTATIONS

	t=1°	t=2°	t=3°	t=4°
SIFT orientation	0.031	0.055	0.080	0.110
The proposed orientation	0.032	0.065	0.094	0.120

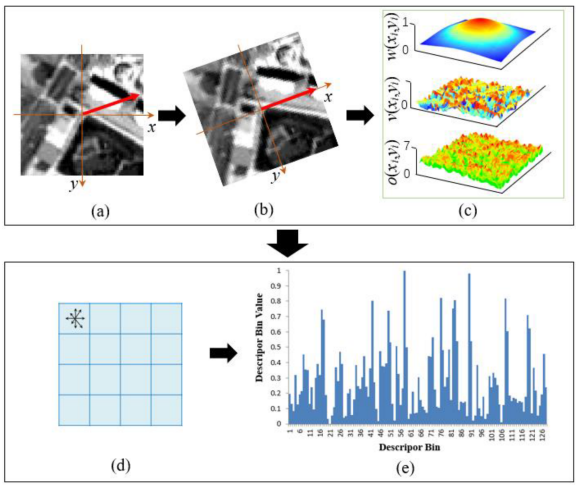


Fig. 9. Flowchart of descriptor calculations. (a) The pixels of $j_d \times j_d$ around the feature. (b) Rotated image according to feature orientation. (c) Gaussian weight, change degree, and change orientation. (d) 4 × 4 grids. (e) Stable descriptors.

C. Stable Descriptors Construction

Based on the characteristic that the changing gray value trend inside the ground objects in heterogeneous images is stable, we use the change orientations and degrees instead of the gradient orientations and values of the SIFT algorithm to construct descriptors. First, based on the orientations of the feature, the pixels of $j_d \times j_d$ (12) around the feature are rotated [see Fig. 9(b)] to ensure the rotation invariance of descriptors. Second, we calculate the change orientation ($o(x_i,y_i)$), change degree ($v(x_i,y_i)$), and Gaussian weight ($w(x_i,y_i)$) of each pixel [see Fig. 9(c)], and calculate the product of the change degree smaller than the threshold (t_v) and the Gaussian weight. Third, we divide the rotated pixels into 4 × 4 grids [see Fig. 9(d)], divide the orientation 2PI into eight bins (on average), and calculate the cumulative weighted change degree of each bin in each grid area to form an 8-D vector. Finally, after the 8-D vectors in each grid are normalized, the vectors in all grids are combined to form 128-D vectors, and the 128-D vectors are normalized—that is, the vector before normalization is h_i and the vector after normalization is l_i (14). Meanwhile, to prevent the influence of larger values in the vector, all values larger than the threshold (t_b) are assigned t_b, after which the 128-D vectors are normalized again to form stable descriptors [see Fig. 9(e)]

$$j_d = 15\sqrt{2}\sigma_i \tag{12}$$

$$l_i = \frac{h_i}{\sqrt{\sum h_i^2}} \tag{13}$$

$$l'_i = \begin{cases} l_i, & l_i \le t \\ 0.2, & l_i > t \end{cases}. \tag{14}$$

In order to verify the stability of the proposed descriptor, two images with large spectral differences are used for experiments (see Fig. 10), and the proposed descriptor is compared with LSS, SIFT, and AB-SIFT descriptors. Fig. 10 shows that the descriptors of the proposed algorithm have the highest similarity. SIFT and AB-SIFT descriptors are better than LSS descriptors.

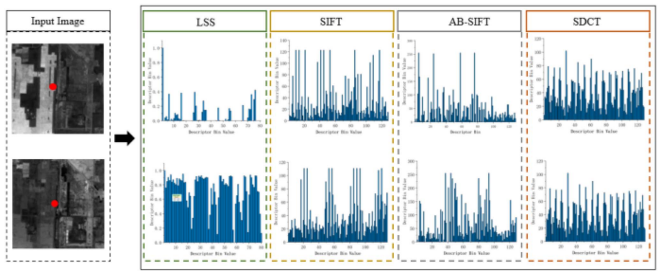


Fig. 10. Descriptor comparison result.

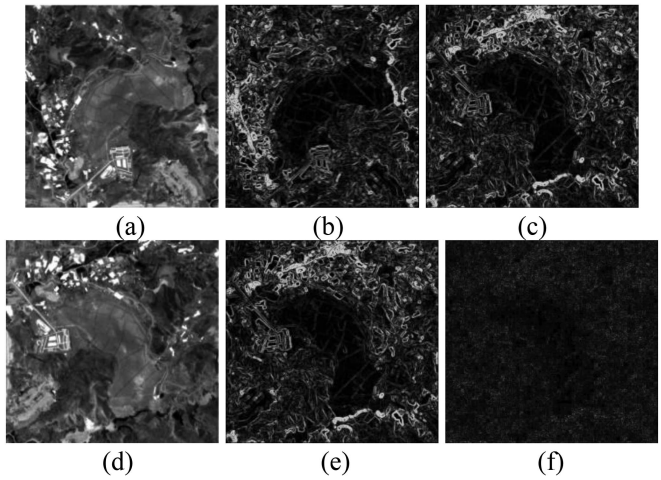


Fig. 11. Effect of rotation on the change degree image. (a) Input image. (b) Change degree image of image (a). (c) Rotated change degree image. (d) Rotated image. (e) Change degree image of image (d). (f) Error image.

LSS descriptors are extremely sensitive to spectral differences. The descriptors of LSS, SIFT, AB-SIFT, and the proposed algorithm are normalized to 0–1. The sum of variance of the two image descriptors is calculated, and the results are 37.94, 8.42, 8.68, and 4.37, respectively, which shows that compared with other algorithms, the proposed algorithm has higher stability to spectral differences.

Invariances of the proposed descriptor are as follows.

1) The proposed descriptor is calculated using the Gaussian pyramid of the images, which is obtained by scaling and blurring the original image. Therefore, the proposed descriptor is invariant to image scaling and blurring.
2) We use cubic polynomials to fit the gray values of 8-neighborhood pixels in four directions and then use the extremes of polynomials to calculate the descriptor (change orientation and degree). The cubic polynomial fitted to the 8-neighborhood pixels remains stable regardless of image rotation, thus the proposed descriptor is rotation invariant. As shown in Fig. 11 we rotate image (a) by 90° to obtain image (e), calculate the change degrees for image (a) and rotated image (e) to obtain images (b) and (e), respectively, and rotate image (b) by 90° to obtain image (c), and subtract image (c) from image (e) to obtain image (f). From image (f), we can see that most of the values are close to 0, indicating that the change degrees are invariant to image rotation.

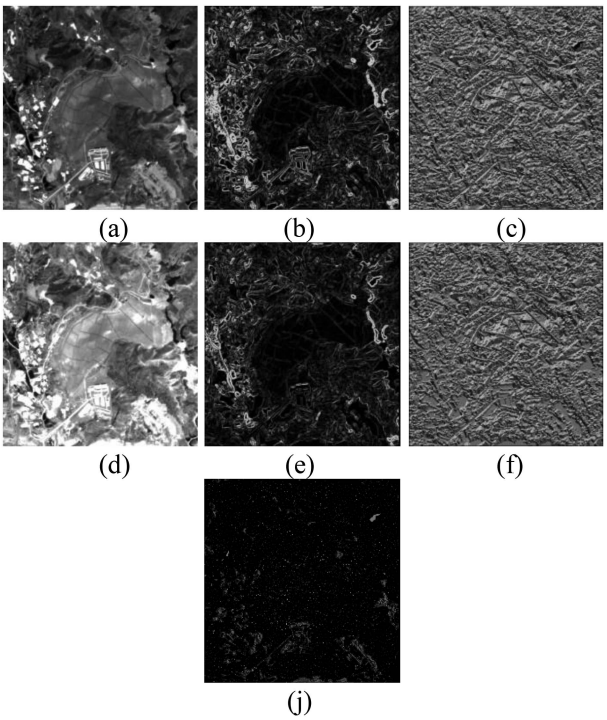


Fig. 12. Effect of stretching on trend information. (a) Input image. (b) Change degree image of image (a). (c) Change orientation image of image (a). (d) Stretched image. (e) Change degree image of image (d). (f) Change orientation image of image (d). (j) Error image.

3) The proposed descriptor uses local information rather than global information. Even if some areas of the image exhibit significant nonlinear spectral changes, the proposed descriptor can still match regions with small spectral changes, resulting in higher applicability of the proposed descriptor.
4) The proposed descriptor is calculated using the change trends of gray values, which mainly describes the orientation and degree of grayscale variation in local regions. If change trends of gray values in the local region of the reference and matched images are stable, the proposed descriptor can be used. The change trends of gray values in images with linear spectral changes are stable, thus the proposed descriptor is applicable to images with linear spectral changes (the proposed descriptor is not applicable to images with grayscale inversion changes, as the grayscale change trend has changed).

As shown in Fig. 12, we linearly change the gray values of the input image (a) to obtain the image (d), and calculate the change degrees [image (b) and image (e)] and change orientations [image (c) and image (f)] of images (a) and (d). Because the gray values change, the change degrees will also change [there is a difference between the gray values of image (b) and image (e)], but in a local area, the values with a large degree of change are still large values, and the small values are still small values. Although the gray value changes, the change orientations of the gray values in the local area are relatively stable (because the image is stretched, the larger and the smaller gray values in the original image will be the same after gray stretching, resulting in the change orientations, but the change orientations in most areas are unchanged), so the images (c) and (f) are similar. Subtract images (c) and (f) to obtain image (j). From the image (j), we can see that most of the values are close to 0. The above information indicates that the trends of gray values are stable to the linear change of the spectrum.

IV. Experiments and Discussion

To verify the reliability of the proposed algorithm, several parts of homologous or heterogeneous remote sensing images are used for experiments. The matching performance of the SDCT algorithm is compared with that of the LSS [15], GOM-SIFT [18], AB-SIFT [19], and RIFT [23] algorithms, all algorithms adopting the default parameters in the literature. To ensure the fairness of comparison, all algorithms use the nearest neighbor method to match features, the threshold of the nearest neighbor method being 0.8. The experiment has been conducted on a computer comprising a 3.6 GHz Intel Core i7-7700 CPU, 16 GB RAM, and a C# platform. The following section discusses the datasets, evaluation criteria, study parameters, experimental results, and stability analysis.

A. Datasets

The datasets are divided into two categories—that is, homologous and heterogeneous optical image pairs (the proposed algorithm is suitable for matching between optical and optical images, but not for matching between SAR and optical images). The overlapping rate of these image pairs ranges from 30% to 90%, the acquisition times differ, and they include images taken in cities and suburbs. Consequently, they have different texture features, obvious spectral differences, and changes in ground objects. The first type of dataset comprises images from the same sensor in different bands, taken at different times, and the second type of dataset comprises images of different sensors, the resolution difference between images is of the order of 1–4 times. The thumbnails of the image pairs are shown in Fig. 13.

B. Evaluation Criteria

Precision (p) and recall (r) are common criteria to measure the matching results, so we use these two criteria to evaluate the proposed and comparison algorithms

$$p = \frac{TP}{TP + FP} \tag{15}$$

$$r = \frac{TP}{TP + FN} \tag{16}$$

where TP denotes the number of correct matching pairs that are detected as correctly matching pairs, FP denotes the number of incorrectly matching pairs that are detected as correct matching pairs, and FN denotes the number of correctly matching pairs that are detected as incorrectly matching pairs.

The evenly distributed correlations selected by skilled operators are used to evaluate the positional accuracy of the matched pairs. The SIFT algorithm is used to extract the features of the

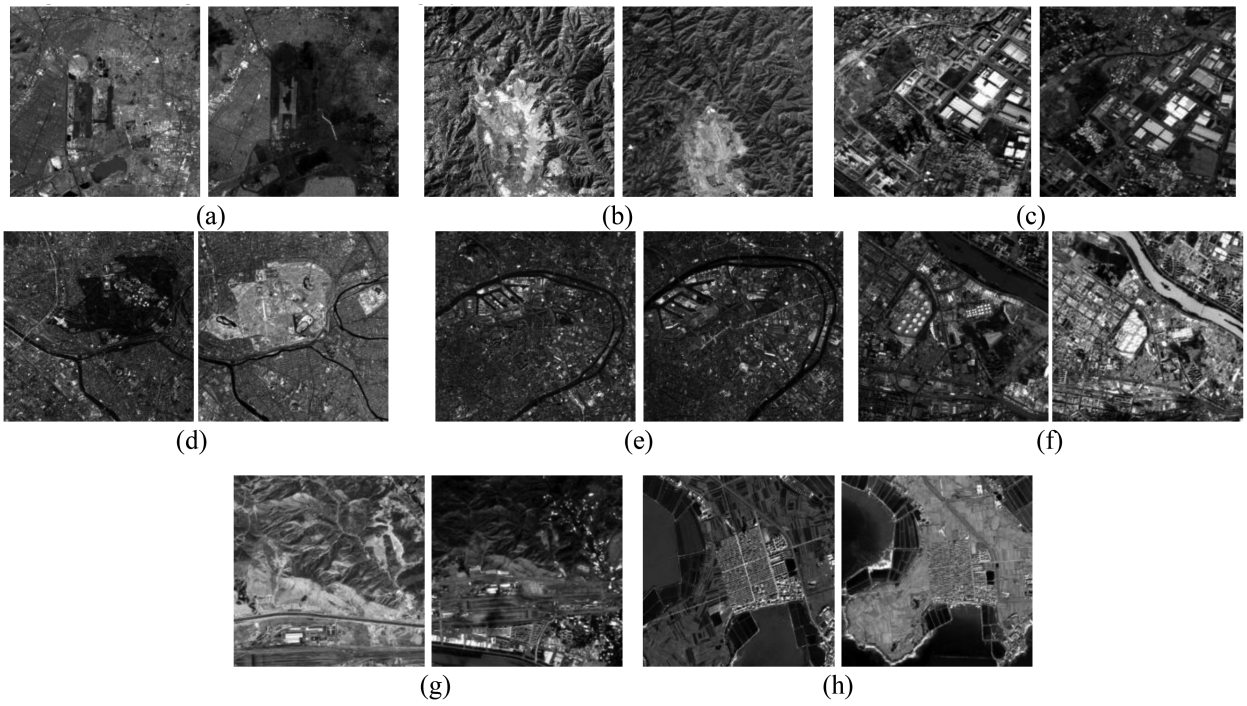


Fig. 13. Thumbnails of the image pairs. (a) First image pair. (b) Second image pair. (c) Third image pair. (d) Fourth image pair. (e) Fifth image pair. (f) Sixth image pair. (g) Seventh image pair. (h) Eighth image pair.

sensed image. Skill operators select 100–200 evenly distributed target points from these features, and their conjugate positions are found on the reference image. All target point pairs are used to fit the projection model to calculate the residuals of the matched pairs. Matching pairs with residuals less than the threshold are regarded as correct matching pairs.

C. Parameters Study

There are two parameters of the proposed algorithm, namely t_v and t_b. Consequently, t_v is used to judge whether a pixel belongs to an edge area so that when the change degree of a pixel is greater than t_v, the pixel is considered as an edge pixel or noise. The smaller the t_v value, the more pixels are considered as edge pixels or noise. Conversely, the larger the t_v value, the fewer pixels are considered as edge pixels or noise. In addition, t_b is used to limit the larger value in the descriptor vector, and the default value (0.2) of SIFT descriptor is used in the proposed algorithm.

Two homologous image pairs (the first and third image pairs shown in Table IV) and two heterogeneous image pairs (the fifth and seventh image pairs shown in Table IV) are used to determine the threshold (t_v). First, the SIFT algorithm is used to extract image features. Second, the orientations and descriptors of the features are calculated using different thresholds (t_v). Finally, the nearest neighbor method is used to calculate the similarity of descriptors, and the precision (p) and the number of correctly matching pairs are obtained (see Table V).

To obtain more regions with large change degrees, the threshold (t_v) is selected from the set {2,3,4,5,6}. Table V shows that when the threshold (t_v) ranges from 2 to 6, the precision and the number of correctly matching pairs are similar. When the threshold (t_v) is greater than or equal to 5, the precision and the number of correctly matching pairs tend to be stable. When the threshold (t_v) is equal to 2, 3, or 4, the precision may be higher than that when the threshold (t_v) is greater than or equal to 5. However, when t_v is equal to 2, the overall effect of the number of correctly matching pairs is at its worst. When the threshold (t_v) is equal to 4, the precision of the first, third, and seventh image pairs is better than when the threshold (t_v) is equal to 3, and the number of correctly matching pairs of the first, third, fifth, and seventh image pairs is equal to or greater than when the t_v is equal to 3. Therefore, on the whole, when the threshold (t_v) is equal to 4, the precision and the number of correct matching pairs of the four image pairs are the best, so 4 can be considered to be the default value of t_v.

D. Experimental Results and Stability Analysis

Table VI shows the precision and recall of the matching results of the proposed algorithm, as well as the AB-SIFT, GOM-SIFT, and LSS algorithms. The bold fonts in the table represent the highest precision and recall values. Figs. 14–16 show the matching results of the algorithms. The blue lines in the figures represent the correctly matching pairs and the red lines represent the incorrectly matching pairs. Table VI and Figs. 14–16 reveal the following.

1) Among all the matching image pair results, the precisions of the proposed algorithm are the best overall, with a small

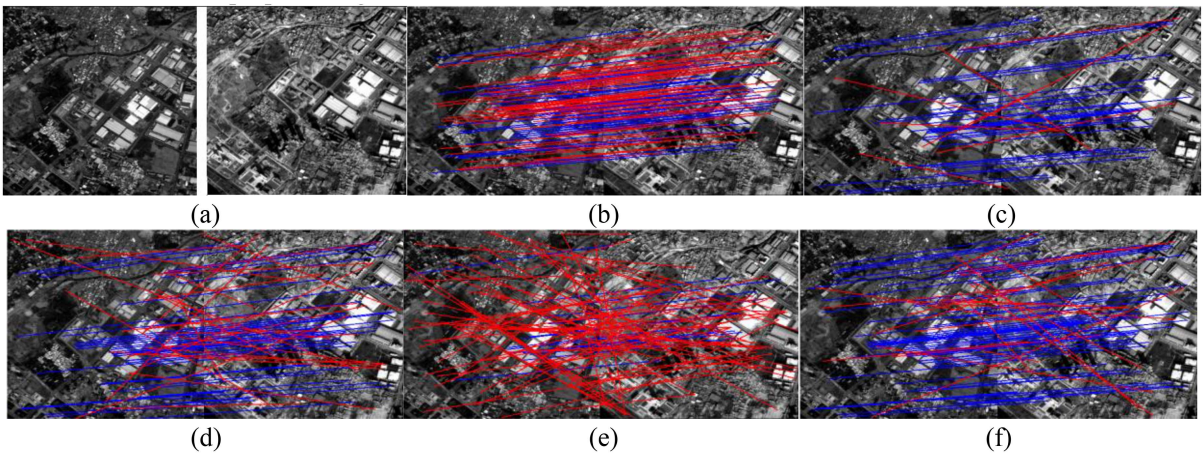


Fig. 14. Comparison results of the third image image pair. (a) Input image. (b) RIFT. (c) AB-SIFT. (d) GOM-SIFT. (e) LSS. (f) SDCT.

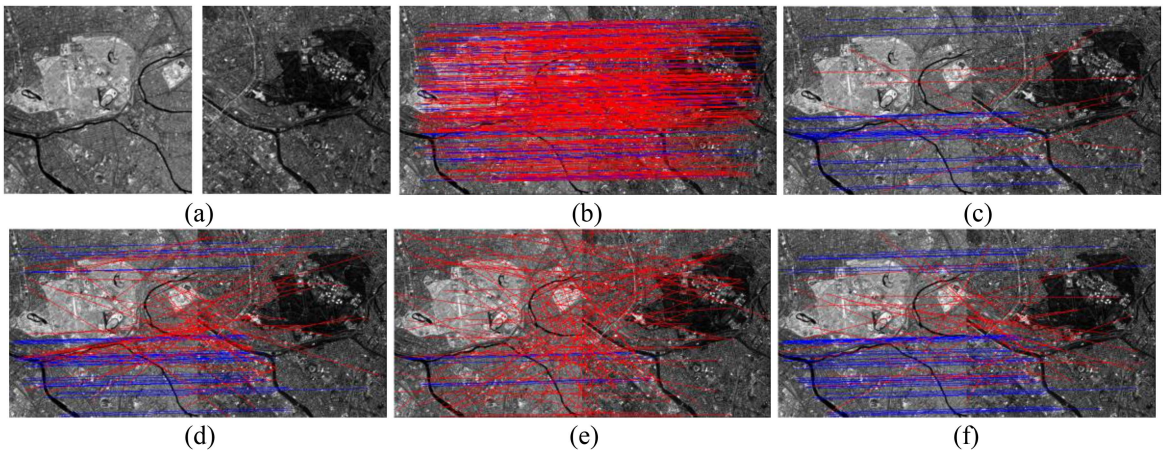


Fig. 15. Comparison results of the fourth image pair. (a) Input image. (b) RIFT. (c) AB-SIFT. (d) GOM-SIFT. (e) LSS. (f) SDCT.

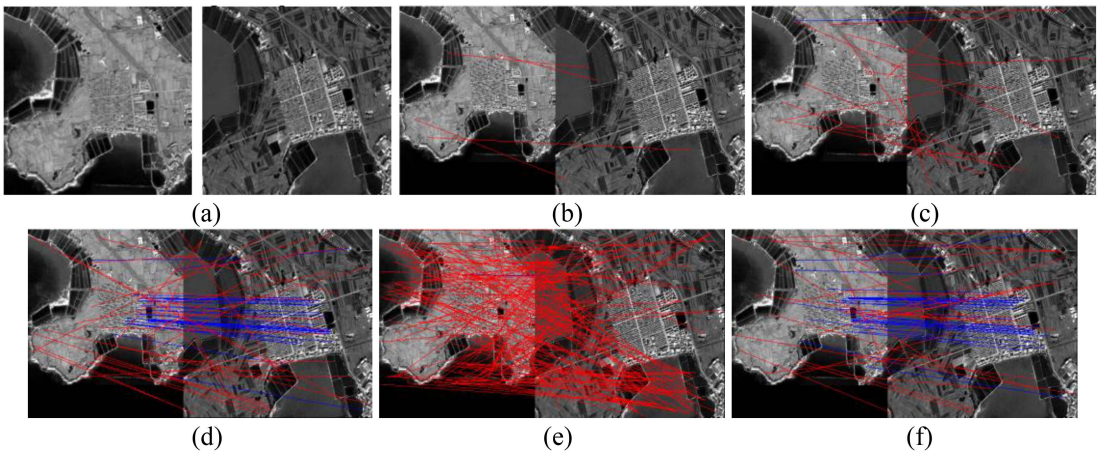


Fig. 16. Comparison results of the eighth image pair. (a) Input image. (b) RIFT. (c) AB-SIFT. (d) GOM-SIFT. (e) LSS. (f) SDCT.

TABLE IV
DATASETS

Category	No.	Image Pair	Size，Resolution	Date	Location
Homologous images	1	Landsat 8: B2 Landsat 8: B5	700 × 700, 30 m 700 × 700, 30 m	2019.01 2018.02	USA-Mexico, Urban
	2	GF-1: B1 GF-1: Pan	700 × 700, 8 m 2000 × 2000, 2 m	2017.08 2014.07	China-Lanzhou, Rural
	3	GF-2: B1 GF-2: B3	424 × 411, 4 m 432 × 406, 4 m	2015.08 2015.01	China-Lanzhou, Rural
	4	AVNIR-2: B3 AVNIR-2: B1	700 × 700, 10 m 700 × 700, 10 m	2007.03 2011.04	France-Paris, Urban
Heterogeneous images	5	AVNIR-2: B1 Landsat 8: Pan	610 × 581, 10 m 488 × 464, 15 m	2011.04 2019.02	France-Pair, Urban
	6	GF-1: B3 ZY-3: B1	583 × 581, 8 m 600 × 600, 5.8 m	2017.08 2016.09	China-Lanzhou, Urban
	7	GF-1: B1 GF-2: B3	500 × 500, 8 m 700 × 700, 4 m	2017.08 2017.05	China-Lanzhou, Rural
	8	GoogleEarth: B1 SPOT5: Pan	600 × 600, 3.85 m 729 × 726, 2.5 m	2002.12 2003.12	China-Shandong, Rural

TABLE V
MATCHING RESULTS OF DIFFERENT VALUE OF T_V

t_v / No.	*P* (%)					*Correct Matches*				
	2	**3**	**4**	**5**	**6**	**2**	**3**	**4**	**5**	**6**
1	**67.35**	64.00	65.31	65.31	65.31	**33**	32	32	32	32
3	**79.57**	78.35	78.57	78.79	78.79	74	76	77	**78**	**78**
5	71.43	**77.14**	73.33	73.33	74.32	50	54	**55**	**55**	**55**
7	68.93	70.59	**70.87**	69.90	69.90	71	72	**73**	72	72

TABLE VI
PRECISION AND RECALL OF COMPARISON ALGORITHMS

No.	*p*(%)					*r*(%)				
	RIFT	**AB-SIFT**	**GOM-SIFT**	**LSS**	**SDCT**	**RIFT**	**AB-SIFT**	**GOM-SIFT**	**LSS**	**SDCT**
1	42.47	51.16	41.30	5.94	**64.00**	**10.06**	1.91	1.65	1.13	2.78
2	0.00	40.91	39.47	3.28	**57.78**	0.00	0.29	0.49	0.03	**0.84**
3	47.66	**82.61**	66.28	17.17	78.35	4.71	11.73	17.59	5.25	**23.46**
4	25.32	**69.23**	38.24	6.86	61.04	**13.34**	1.44	2.07	0.37	2.50
5	0.00	17.39	46.73	1.42	**70.00**	0.00	0.31	**3.86**	0.15	3.78
6	0.00	14.29	70.95	2.30	**81.37**	0.00	0.07	4.38	0.07	**7.38**
7	0.00	56.25	70.30	10.81	**70.59**	0.00	2.27	8.95	1.51	**9.08**
8	0.00	5.56	46.88	0.65	**51.61**	0.00	0.07	2.12	0.07	**2.26**

fluctuation range (51.61–81.37%) and an average precision (66.84%), because the proposed algorithm constructs the descriptor using stably changing trend information. Compared with gradient information—the AB-SIFT and GOM-SIFT algorithms use gradient information—and correlation value—the LSS algorithm uses correlation value—the changing trend information is more robust to spectral changes. In addition, the recall rate of the proposed algorithm is better than that of the compared algorithms. Although the recall rate of the fifth image (3.78%) of the proposed algorithm is slightly lower than that of the GOM-SIFT algorithm (3.86%), the precision of the proposed algorithm (70.00%) is 1.49 times that of the GOM-SIFT algorithm (46.73%). The precision of the proposed algorithm for the first, third, and fourth images are better than that of the RIFT algorithm, and the average precision of the proposed algorithm for these three images (67.8%) is about 1.76 times that of the RIFT algorithm (38.48%) (since the RIFT algorithm is invariant in scale [23], this article only compares the image pairs with the same resolution). The precisions of the proposed algorithm of the third and fourth image pairs are slightly

lower (78.35%, 61.04%) than that of the AB-SIFT algorithm (82.61%, 69.23%), because the spectral difference between the third and fourth image pairs is small, and the similarity of the AB-SIFT descriptors based on gradient information is higher, allowing it to obtain higher precision. Fig. 14 shows that the spectral difference of the third image pair is small, the spectrum changing linearly. Although the spectrum change in the middle area of the fourth image pair is large (see Fig. 15), the spectrum change in most areas of the images is small. The precision similarity of the third and fourth image pairs using this algorithm is lower than that of the AB-SIFT algorithm, but because the robustness of the proposed algorithm is higher, the recall rates (22.46%, 2.50%) are approximately twice than that of the AB-SIFT descriptor (11.73%, 1.44%).

2) Because the RIFT algorithm has no scale invariance, the algorithm can only match the first, third, and fourth image pairs, but cannot match other images. Therefore, we only analyze the image pairs with the same resolution. RIFT algorithm matches images according to the edge information of ground objects, which can improve the robustness of the algorithm to heterogeneous images, but it also makes the algorithm only deal with a certain type of image. When the edge information is less or more complex, the matching effect of this algorithm is not ideal. When the images have the same resolution and contain a few simple features, such as the first and fourth pairs of images, the recall rate of RIFT algorithm is better than other algorithms, but the precision is lower than the proposed algorithm and AB-SIFT. This is because the RIFT algorithm uses the corners obtained from frequency domain information as feature points, whereas the proposed algorithm and AB-SIFT use the corners and spots extracted from SIFT as feature points. That is to say, compared with the proposed algorithm and AB-SIFT algorithm, RIFT algorithm has selected feature points, so the recall rate will be better than the proposed algorithm and AB-SIFT algorithm. The descriptors of the proposed algorithm, AB-SIFT algorithm, and GOM-SIFT algorithm are more robust than RIFT algorithm in homologous images with little spectral change. When the resolution of the images is the same, but they contain a lot of complicated ground objects, such as the third pair of images, the precision (47.66%) and recall (4.71%) of RIFT are lower than those of the proposed algorithm, AB-SIFT, and GOM-SIFT (the precision is 78.35%, 82.61%, and 66.28%, respectively, and the recall is 23.48%, respectively).

3) The AB-SIFT algorithm is suitable for image pairs with small or linear spectral changes (the first, second, third, fourth, and seventh image pairs), its precision is greater than 40.00%. However, when the spectral changes are large and nonlinear (the fifth, sixth, and eighth image pairs), its precision is lower (less than 20.00%). In particular, the eighth image pair (see Fig. 16) exhibits a considerable spectral change, containing a large number of nonlinear change areas, resulting in the precision of the AB-SIFT algorithm being only 5.56%. Furthermore, the recall rates of the AB-SIFT algorithm are generally lower than those of the GOM-SIFT and proposed algorithms, because it directly uses gradient information to calculate descriptors, gradient information being sensitive to illumination changes.

4) The stability of the GOM-SIFT algorithm is higher than that of the AB-SIFT algorithm, the precision of the GOM-SIFT algorithm ranging from 38.24% to 70.95%, whereas that of the AB-SIFT algorithm fluctuated considerably, ranging from 5.56% to 82.61%. In addition, the recall rates of the GOM-SIFT algorithm are generally better than those of the AB-SIFT algorithm. The precisions of the GOM-SIFT algorithm are lower than that of the AB-SIFT algorithm for homologous image pairs (the first, second, third, and fourth image pairs) with small spectral changes. However, the precision of the GOM-SIFT algorithm is vastly better than that of the AB-SIFT algorithm, being approximately 3–8 times that of the AB-SIFT algorithm. For example, the precisions of the GOM-SIFT algorithm in the fifth, sixth, and eighth image pairs are 46.73%, 70.95%, and 46.88%, respectively, whereas those of the AB-SIFT algorithm are 17.39%, 14.29%, and 5.56%, respectively, as the GOM-SIFT algorithm modifies the gradient orientation, reducing the sensitivity of descriptors to spectral changes, as well as the uniqueness of descriptors to some extent. This resulted in lower precision than that of the AB-SIFT algorithm in image pairs with small spectral changes, but considerably better precision than that of the AB-SIFT algorithm in image pairs with large spectral differences. The matching effect of the LSS algorithm is the worst, indicating that the LSS algorithm is not suitable for remote sensing images with spectral changes.

The above analysis shows that the proposed algorithm exhibits high stability, and can be applied to image pairs with small spectral differences and still obtain better results for images with large spectral differences, making it a better solution than the AB-SIFT, GOM-SIFT, and LSS algorithms in general.

V. Conclusion

To solve the problem of the low descriptor similarity caused by large spectral differences of heterogeneous optical remote sensing images, we propose a stable descriptor construction algorithm that considered the changing gray value trends inside ground objects. The proposed algorithm calculates the orientations and descriptors of features using stable change orientations and degrees. Consequently, the proposed algorithm can solve the problem of the low descriptor similarity of corresponding features. In addition, the cubic polynomial is used to fit the 8-neighbors' gray values to obtain accurate and stable change orientations and degrees, reducing the influence of noise and image resolution. The threshold (t_v) is used to eliminate edge pixels and noise with large spectral changes, thus further improving the stability of the descriptor. Experiments with homologous and heterogeneous optical remote sensing images show that the proposed descriptor exhibits high stability to spectral changes, and its precision and recall rates are better than those of comparative descriptors. (Demo codes are available at https://orcid.org/0000--0002--3535--1770.)

REFERENCES

[1] J. Zheng et al., "Growing status observation for oil palm trees using unmanned aerial vehicle (UAV) images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 173, pp. 95–121, 2021.

[2] M. Kakooei and Y. Baleghi, "Fusion of satellite, aircraft, and UAV data for automatic disaster damage assessment," *Int. J. Remote Sens.*, vol. 38, no. 8–10, pp. 2511–2534, 2017.

[3] Z. Chen et al., "A new image mosaic of Greenland using Landsat-8 OLI images," *Sci. Bull.*, vol. 65, no. 7, pp. 522–524, 2020.

[4] J. Zhang et al., "Cloud detection in high-resolution remote sensing images using multi-features of ground objects," *J. Geovis. Spatial Anal.*, vol. 3, no. 2, pp. 1–9, 2019.

[5] W. Li and H. Leung, "A maximum likelihood approach for image registration using control point and intensity," *IEEE Trans. Image Process.*, vol. 13, no. 8, pp. 1115–1127, Aug. 2004.

[6] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *Int. J. Comput. Vis.*, vol. 60, no. 1, pp. 63–86, 2004.

[7] A. Sedaghat and N. Mohammadi, "Uniform competency-based local feature extraction for remote sensing images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 135, pp. 142–157, 2018.

[8] Y. Ye et al., "A local phase based invariant feature for remote sensing image matching," *ISPRS J. Photogrammetry Remote Sens.*, vol. 142, pp. 205–221, 2018.

[9] L. Xue, S. Yang, and Y. Liu, "Remote sensing image matching featured by the optimal entropy classification," *J. Appl. Remote Sens.*, vol. 15, no. 4, 2021, Art. no. 044515.

[10] A. Sedaghat and H. Ebadi, "Distinctive order based self-similarity descriptor for multi-sensor remote sensing image matching," *ISPRS J. Photogrammetry Remote Sens.*, vol. 108, pp. 62–71, 2015.

[11] 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.

[12] Y. Xiang, N. Jiao, F. Wang, and H. You, "A robust two-stage registration algorithm for large optical and SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Dec. 8, 2021, Art. no. 5218615, doi: 10.1109/TGRS.2021.3133863.

[13] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[14] H. Bay et al., "Speeded-up robust features (SURF)," *Comput. Vis. Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[15] E. Shechtman and M. Irani, "Matching local self-similarities across images and videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2007, pp. 1–8.

[16] L. Zhou, Y. Ye, T. Tang, K. Nan, and Y. Qin, "Robust matching for SAR and optical images using multiscale convolutional gradient features," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Sep. 20, 2021, Art. no. 4017605, doi: 10.1109/LGRS.2021.3105567.

[17] 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.

[18] Z. Yi, C. Zhiguo, and X. Yang, "Multi-spectral remote image registration based on SIFT," *Electron. Lett.*, vol. 44, no. 2, pp. 107–108, 2008.

[19] 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.

[20] M. Calonder et al., "BRIEF: Binary robust independent elementary features," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 778–792.

[21] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 2564–2571.

[22] Y. Ye and L. Shen, "HOPC: A novel similarity metric based on geometric structural properties for multi-modal remote sensing image matching," *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 3, pp. 9–16, 2016.

[23] 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. 17, 2019, doi: 10.1109/TIP.2019.2959244.

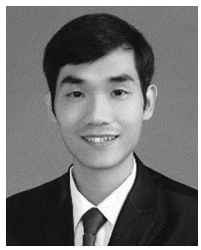
[24] C. Gao, W. Li, R. Tao, and Q. Du, "MS-HLMO: Multiscale histogram of local main orientation for remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jul. 2022, Art. no. 5626714.

[25] B. Zhu, C. Yang, J. Dai, J. Fan, Y. Qin, and Y. Ye, "R2FD2: Fast and robust matching of multimodal remote sensing images via repeatable feature detector and rotation-invariant feature descriptor," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Apr. 2023, Art. no. 5606115.

[26] J. Liu and G. Zeng, "Description of interest regions with oriented local self-similarity," *Opt. Commun.*, vol. 285, no. 10/11, pp. 2549–2557, 2012.

[27] P. Viola and W. M. Wells III, "Alignment by maximization of mutual information," *Int. J. Comput. Vis.*, vol. 24, no. 2, pp. 137–154, 1997.

[28] J. Li, W. Xu, P. Shi, Y. Zhang, and Q. Hu, "LNIFT: Locally normalized image for rotation invariant multimodal feature matching," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Apr. 2022, Art. no. 5621314.



Li Xue received the B.S. degree in geographic information system from North China University of Water Resources and Electric Power, Zhengzhou, China, in 2016, and the M.Sc. degree in cartography and geographic information system from Lanzhou Jiaotong University, Lanzhou, China, in 2019. He is currently working toward the Ph.D. degree in cartography and geographic information system with Nanjing Normal University, Nanjing, China.

His research interests include the image matching and shadow extraction.

Yehua Sheng received the Ph.D. degree in engineering measurement from the China University of Mining and Technology, Jiangsu, China, in 1997.

He is currently a Full Professor in cartography and geographic information system with Nanjing Normal University, Nanjing, China. His research interests include geospatial data collection and processing, multidimensional spatiotemporal data model, virtual geographic environment, and digital photogrammetry.

Ka Zhang received the bachelor's degree in surveying and mapping engineering and the master's degree in photogrammetry and remote sensing from China University of Mining and Technology, Beijing, China, in 2002 and 2005, respectively, and the Ph.D. degree in cartography and geographic information system from Nanjing Normal University, Nanjing, China, in 2008.

Since 2008, he has been with the Key Laboratory of Virtual Geographic Environment, Ministry of Education and School of Geography, Nanjing Normal University, where he is currently a Full Professor. He has authored or coauthored more than 20 academic papers in SCI/EI indexed journals. His research interests include image processing and digital photogrammetry.

Dr. Zhang was the recipient of several provincial awards of technology progress.