



Article

Edge Consistency Feature Extraction Method for Multi-Source Image Registration

Yang Zhou ^{1,2}, Zhen Han ^{3,*} , Zeng Dou ⁴, Chengbin Huang ⁴, Li Cong ⁴, Ning Lv ⁵ and Chen Chen ³ ¹ School of Computer Science and Technology, Xidian University, Xi'an 710071, China; zhy@mwr.gov.cn² Ministry of Water Resources, Beijing 101400, China³ School of Telecommunication Engineering, Xidian University, Xi'an 710071, China; cc2000@mail.xidian.edu.cn⁴ State Grid Jilin Province Electric Power Company Limited Information Communication Company, Changchun 130021, China; douzeng@jl.sgcc.com.cn (Z.D.); huangchb@jl.sgcc.com.cn (C.H.); congli@jl.sgcc.com.cn (L.C.)⁵ School of Electronic Engineering, Xidian University, Xi'an 710071, China; nlv@mail.xidian.edu.cn

* Correspondence: hanzhen@stu.xidian.edu.cn

Abstract: Multi-source image registration has often suffered from great radiation and geometric differences. Specifically, grayscale and texture from similar landforms in different source images often show significantly different visual features, and these differences disturb the corresponding point extraction in the following image registration process. Considering that edges between heterogeneous images can provide homogeneous information and more consistent features can be extracted based on image edges, an edge consistency radiation-change insensitive feature transform (EC-RIFT) method is proposed in this paper. Firstly, the noise and texture interference are reduced by preprocessing according to the image characteristics. Secondly, image edges are extracted based on phase congruency, and an orthogonal Log-Gabor filter is performed to replace the global algorithm. Finally, the descriptors are built with logarithmic partition of the feature point neighborhood, which improves the robustness of the descriptors. Comparative experiments on datasets containing multi-source remote sensing image pairs show that the proposed EC-RIFT method outperforms other registration methods in terms of precision and effectiveness.

Keywords: image registration; edge consistency features; multi-source remote sensing images



Citation: Zhou, Y.; Han, Z.; Dou, Z.; Huang, C.; Cong, L.; Lv, N.; Chen, C. Edge Consistency Feature Extraction Method for Multi-Source Image Registration. *Remote Sens.* **2023**, *15*, 5051. <https://doi.org/10.3390/rs15205051>

Academic Editor: Salah Bourennane

Received: 16 July 2023

Revised: 24 August 2023

Accepted: 25 August 2023

Published: 21 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the continuous breakthrough of remote sensing technology, the ways to obtain information through multi-source satellites, unmanned aerial vehicles, and other sensors have gradually increased, and the space-air-ground integrated system based on multi-sensors has gradually improved. Due to the different methods and principles of data acquisition by different sensors, the acquired Earth observation data have their own information advantages, and edge intelligence greatly improves multi-source data transmission and processing speed [1]. In order to integrate the multi-source remote sensing data to achieve complementary advantages and improve the data availability, information fusion of remote sensing images is usually required. Multi-source image registration refers to the technology of matching images in the same scene obtained by different image sensors and establishing the spatial geometric transformation relationship between them. Its accuracy and efficiency will directly impact the subsequent tasks, such as remote sensing interpretation in land-use surveillance [2]. Although the remote sensing data are commonly coarse-registered with the spatial reference, there are still slight translation or rotation differences between the images. The fine-registration task is particularly important to realize the precise alignment of objects.

However, due to the large differences in the operating mechanisms and conditions of imaging sensors, multi-source remote sensing images often show significant variations in

the geometric and radiometric features. These obvious differences between multi-source remote sensing images are reflected in the registration process as the same landform or the target in the image presents distinct image features, especially internal texture features. Therefore, algorithms such as SIFT proposed for homologous registration often fail [3]. Traditional multi-source image registration methods can be divided into two categories: region-based and feature-based methods. Region-based methods often use similarity measures and optimization methods to accurately estimate transformation parameters. A typical solution is to evaluate similarity with MI [4], which is generally limited by high computational complexity. Another approach is to improve processing efficiency through the domain transfer techniques such as fast Fourier Transform (FFT) [5]. However, facing large-scale remote sensing images, region-based methods are sensitive to severe noise caused by imaging sensors and atmosphere, making it difficult to be applied widely. Therefore, feature-based methods are more commonly used in the remote sensing field. These methods usually detect significant features between images (such as points, lines, and regions) and then identify correspondence by describing the detected features. The PSO-SIFT algorithm optimizes the calculation of image gradients based on SIFT and exploits the intrinsic information of feature points to increase the number of accurate matching points for the problem of multi-source image matching [6]. The applicatory multi-source data include optical, SAR, infrared, and LiDAR images, and the SAR and optical images are the most widely used and representative ones. Therefore, many studies focus on SAR–optical registration. OS-SIFT is mainly focused on SAR images and optical images matching and uses the multiscale exponential weighted average ratio and the multiscale Sobel operator to calculate the SAR image and optical image gradients to improve the robustness [7]. The multiscale histogram of local main orientation (MS-HLMO) [8] method develops a basic feature map for local feature extraction and a matching strategy based on Gaussian scale space. The development of deep learning technology has accelerated the research of computer vision, and more and more learning-based methods have emerged in the registration field [9]. A common strategy is to generate deep and advanced image representations, such as using the network to train feature detectors [10], descriptors [11], or similarity measures, instead of traditional registration steps. There are also many methods based on Siamese networks or pseudo-Siamese networks [12], which obtain the deep feature representation by training the network [13] and then calculate the similarity of the output feature map to achieve template matching. As the basis of the deep learning method, large datasets are critical to its performance. However, there are few available multi-source remote sensing datasets and the acquisition cost is high; labeled examples are even rarer, resulting in the difficulty of wide application to multi-source data, which greatly limits the development of learning-based methods.

These early feature-based methods are mainly based on gradient information, but the application in multi-source image registration is often limited due to the sensitivity to illumination differences, contrast differences, etc. In recent years, as the concept of phase congruency (PC) [14] has been introduced into image registration, more and more registration methods based on it show superior performance. The histogram of orientated phase congruency (HOPC) feature descriptor achieves matching between multiple modal images by constructing descriptors from phase congruency intensity and orientation information [15]. Li et al. proposed a radiation change insensitive feature transformation (RIFT) algorithm to detect significant corner and edge feature points on the phase congruency map and constructed the maximum index map (MIM) descriptor based on the Log-Gabor convolution image sequences [16], realizing the insensitivity to multi-modal image radiation changes. Yao et al. proposed a histogram of absolute phase consistency gradients (HAPCG) algorithm, which extended the phase consistency model, established absolute phase consistency directional gradients, and built the HAPCG descriptors [17] achieving robust matching between different source images with large illumination and contrast difference. Fan et al. proposed a 3MRS method based on a 2D phase consistency model to construct a new template feature based on Log-Gabor convolutional image sequences

and used 3D phase correlation as a similarity metric [18]. In general, although these phase-congruency-based registration methods demonstrate excellent performance in multi-source image matching, there are still two problems for remote sensing images registration: (1) the interference of noise and texture on feature extraction cannot be avoided; (2) the computation of phase congruency involves Log-Gabor transforms at multiple scales and directions, which is heavy in computation and results in the extension of feature detection time.

The key to automatic registration of multi-source images lies in how to extract and match feature points between heterogeneous images. Since edges are relatively stable features in multi-source images, which can maintain stability when the imaging conditions or mode changes [19], feature extraction based on edge contour information can greatly enhance the registration performance. As shown in Figure 1, for the same area, terrain selected by the red box exhibits significant differences in multi-source images, while the edge marked with the green box tends to be more consistent. Therefore, a multi-source remote sensing image registration method based on homogeneous features is proposed in this paper, namely, edge consistency radiation-variation insensitive feature transform (EC-RIFT). Firstly, in order to suppress speckle noise without destroying edge integrity, a non-local mean filter is performed on SAR images, and edge enhancement is applied to visible images with abundant texture to help extract coherent features. Secondly, image edge information is extracted based on phase congruency, and the orthogonal Log-Gabor filter (OLG) is chosen instead of the global filter algorithm for feature dimension reduction, speeding up the feature detection. Finally, the EC-RIFT algorithm constructs descriptors based on the sector neighborhood of edge feature points, which is more sensitive to nearby points and improves the robustness of the descriptors, thus improving the registration accuracy. The main contributions of our paper are as follows:

1. To capture the similarity of geometric structures and morphological features, phase congruency is used to extract image edges, and the differences between multi-source remote sensing images are suppressed by eliminating noise and weakening texture.
2. To reduce computing costs, phase congruency models are constructed using Log-Gabor filtering in orthogonal directions.
3. To increase the dependability of descriptor features with richer description information, sector descriptors are built based on edge consistency features.

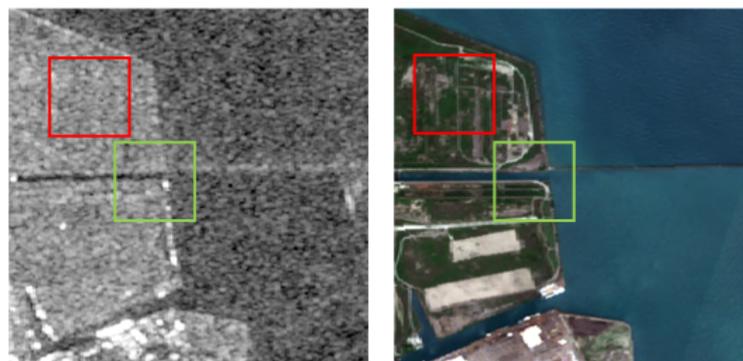


Figure 1. Texture and edge differences between multi-source images. The red box represents textured terrain, and the green box contains significant edges.

The rest of the paper is organized as follows: Section 2 introduces the proposed method in detail. Section 3 evaluates the proposed method by comparing experimental results with other representative methods. Section 4 discusses several important aspects related to the proposed method. Finally, the paper is concluded in Section 5.

2. Materials and Methods

In this section, we first give a brief introduction of RIFT. Firstly, the PC is calculated for each orientation of the Log-Gabor convolution, and the features on the maximum moment

map are extracted by the FAST detector [20]. Then the amplitudes of the convolution results at each scale in the same orientation are summed to obtain the MIM. A convolution sequence ring structure is applied to generate multiple MIMs for each feature in the reference image. Based on the MIM, the descriptors are constructed with feature coding technology like the SIFT algorithm [3]. Finally, the nearest neighbor matching strategy is used on descriptors to obtain the initial correspondence, and the outliers are removed by fast sample consensus (FSC) algorithm [21].

The EC-RIFT algorithm process can be divided into five steps: image preprocessing, feature point detection, descriptor construction, feature matching, and outlier removal. The main process of the proposed framework is shown in Figure 2. The blue boxes indicate the focus of the research in this paper. The following subsections provide a detailed description. Moreover, the important symbols in the paper are explained in Table 1.

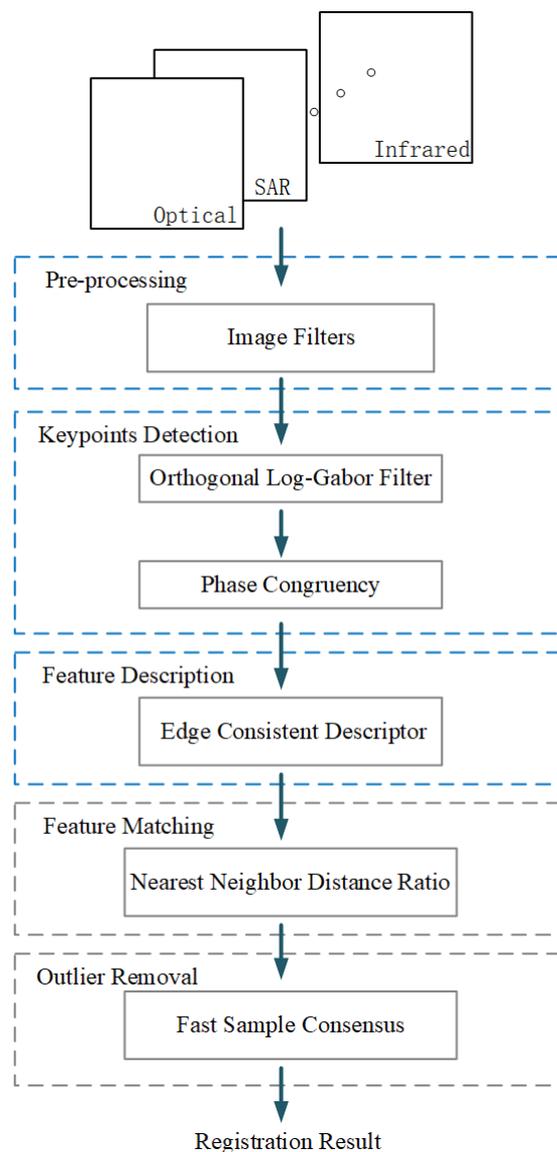


Figure 2. Implementation flow chart of EC-RIFT algorithm.

Table 1. Explanations of important symbols.

Symbol	Explanation	Symbol	Explanation
FI	input of filter	A	amplitude
FO	output of filter	ϕ	phase
w	weight	s	scale
N	number	o	orientation
ρ	polar radius	σ	bandwidth
θ	polar angle		

2.1. Multi-Source Image Preprocessing

The interference of noises and textures on infrared or LiDAR images is not as obvious as that on SAR images in terms of feature extraction. The regular denoise filters, such as median filters, are efficient for infrared images. The design of filters for SAR and optical images is emphasized in this section.

2.1.1. Non-Local Mean Filtering

SAR images are always corrupted by speckle noise, leading to the challenge of subsequent tasks such as feature detection, so it is necessary to suppress the speckle noise first. The non-local mean filtering algorithm (NLM) defines similar pixels as the same neighborhood pattern and uses the information within a fixed-size window around the pixel to represent it. The loss of image structure information during the noise reduction process can be avoided, and it thus performs well in maintaining image edge and structure information [22]. The NLM filtering is calculated as (1):

$$FO_{NLM}(x) = \sum w(x,y)FI_{NLM}(y) \quad (1)$$

where FO_{NLM} and FI_{NLM} represent the output and input images of NLM filters, respectively, x and y represent the indices of pixels, and the weight $w(x,y)$ represents the similarity between pixels x and y , which is determined by the distance $\|V(x) - V(y)\|$ between rectangular regions $V(x)$, $V(y)$, as shown in (2)–(4).

$$w(x,y) = \frac{1}{Z(x)} \exp\left(-\frac{\|Vx - Vy\|^2}{h^2}\right) \quad (2)$$

$$\|Vx - Vy\|^2 = \frac{1}{d^2} \sum_{z \in S_d} \|FI_{NLM}(x+z) - FI_{NLM}(y+z)\|^2 \quad (3)$$

$$Z(x) = \sum_y \exp\left(-\frac{\|Vx - Vy\|^2}{h^2}\right) \quad (4)$$

where $Z(x)$ is the normalization parameter, h is the smoothing parameter, S_d denotes the search window region that regulates the degree of Gaussian function's decay, d represents the window side, and z is the point within S_d .

2.1.2. Co-Occurrence Filtering

To take full advantage of the edge information of image, the edges need to be enhanced while weakening the texture. The co-occurrence filter (CoF) is an edge-preserving filter in which pixel values that appear more frequently in the image are weighted higher in the co-occurrence matrix, so the image texture can be smooth without considering the grayscale difference [23]. Pixel values that rarely appear at the same time are weighted

lower in the co-occurrence matrix, which can better maintain the boundaries. The CoF is defined according to (5):

$$FO_{CoF}(p) = \frac{\sum_{q \in N(p)} G_{\sigma_s}(p, q) \cdot M(FI_{CoF}(p), FI_{CoF}(q)) \cdot FI_{CoF}(q)}{\sum_{q \in N(p)} G_{\sigma_s}(p, q) \cdot M(FI_{CoF}(p), FI_{CoF}(q))} \quad (5)$$

where FO_{CoF} and FI_{CoF} represent the output and input images of CoF filters, respectively, p and q are pixel indices, G denotes the Gaussian filter, σ_s is the standard deviation of G , and M is calculated from the co-occurrence matrix as (6)

$$M(FI_{CoF}(p), FI_{CoF}(q)) = \frac{C(FI_{CoF}(p), FI_{CoF}(q))}{h(FI_{CoF}(p))h(FI_{CoF}(q))} \quad (6)$$

where h denotes the histogram of pixels, indicating the occurrence frequency of p and q . The co-occurrence matrix C is computed as (7).

$$C(FI_{CoF}(p), FI_{CoF}(q)) = \sum_{p, q} \exp\left(-\frac{d(p, q)^2}{2\sigma^2}\right) \quad (7)$$

In Equation (7), d denotes the Euclidean distance, σ^2 is a fixed value, $\sigma^2 = 2\sqrt{5} + 1$.

The comparison of images with/without preprocessing is shown in Figure 3. There are significant geometric and radiometric differences between the original images, making it more difficult to be matched with high precision. It is clear that the speckle noises in SAR image Figure 3a are filtered in Figure 3b, and the textures in optical image Figure 3c are suppressed in Figure 3d with complete edge maintained. The edge information left in the two images has a stronger consistency.

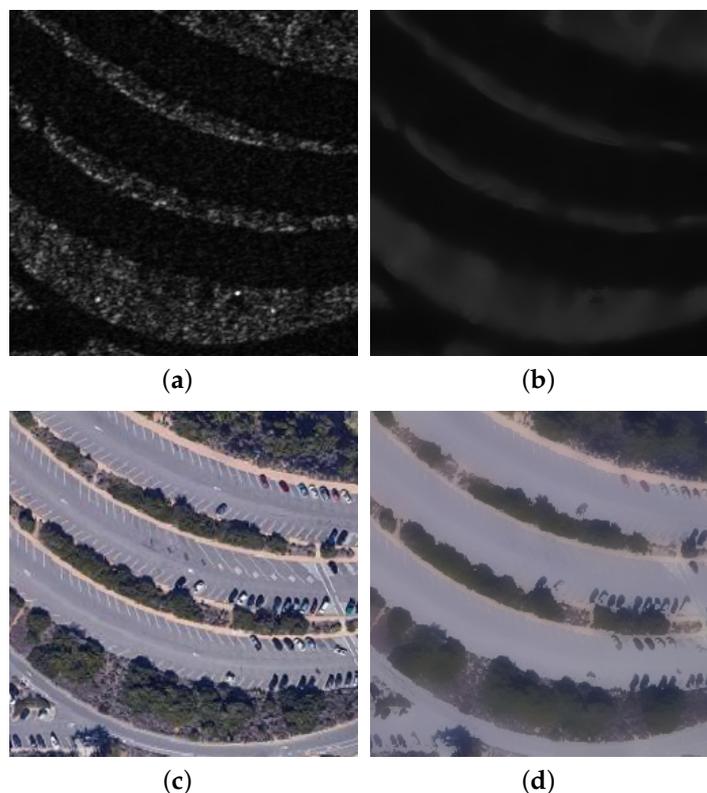


Figure 3. Results of preprocessing. (a) Original SAR image. (b) Denoised SAR image. (c) Original optical image. (d) Enhanced optical image.

2.2. Edge Feature Detection

Edge information can largely reflect the structural features of objects and is more stable in heterogeneous images, so feature detection based on complete edges can provide a more reliable foundation for registration. The phase congruency model can overcome the effects of contrast and brightness changes and extract more complete feature edges than gradients. The Log-Gabor filter is able to maintain zero direct current components in an even symmetric filter, and the bandwidth is arbitrary. Therefore, the Log-Gabor filter is efficient to calculate the phase congruency. However, the feature extraction based on it is time-consuming because of the high dimensions. The number of scales and orientations are set to four and six, respectively. Therefore, in this paper, a 2D orthogonal Log-Gabor filter (OLG) is used instead of the global filter to achieve feature dimension reduction [24].

The geometric structure of images can be well preserved by phase congruency. Oppenheim first identified the ability of phase to extract significant information from images [25], and later Morrone and Owens proposed a model of phase congruency [26]. Since the degree of phase congruency is independent of the overall amplitude of signal, it has better robustness and gives more detailed edge information than the gradient operator when employed for edge detection. Kovessi simplified the calculation of phase congruency with Fourier phase components and extended the model to 2D [14]. Kovessi employed Log-Gabor filters in the computation, which were demonstrated to be more natural for human eyesight.

The intensity value of phase congruency is calculated as Equation (8):

$$PC(x, y) = \frac{\sum_s \sum_o \omega_o(x, y) [A_{so}(x, y) \Delta \Phi_{so}(x, y) - T]}{\sum_s \sum_o A_{so}(x, y) + \zeta} \quad (8)$$

where s and o respectively denote the filter scale and orientation, ω_o denotes the weight, and ζ is a minimal constant preventing the denominator from being 0. A_{so} and ϕ_{so} denote the amplitude and phase of the PC, respectively, and can be derived from Equations (9) and (10):

$$A_{so}(x, y) = \sqrt{E_{so}(x, y)^2 + O_{so}(x, y)^2} \quad (9)$$

$$\phi_{so}(x, y) = \arctan(O_{so}(x, y) / E_{so}(x, y)) \quad (10)$$

E_{so} and O_{so} represent the results of even-symmetric filtering and odd-symmetric filtering convolved with image I , respectively. The calculation formula is shown in (11):

$$[E_{so}(x, y), O_{so}(x, y)] = [I(x, y) * L^{even}(x, y, s, o), I(x, y) * L^{odd}(x, y, s, o)] \quad (11)$$

L is a 2D Log-Gabor filtering function, defined as (12):

$$L(\rho, \theta, s, o) = \exp\left(\frac{-(\rho - \rho_s)^2}{2\sigma_\rho^2}\right) \exp\left(\frac{-(\theta - \theta_{so})^2}{2\sigma_\theta^2}\right) \quad (12)$$

where σ represents the bandwidth, and ρ_s and θ_{so} are the center frequencies of Log-Gabor filtering.

The edges extracted based on phase congruency are often used for edge detection because of their ability to be unaffected by local light variations in the image, and they can contain various information, especially information at low edge contrast.

A 2D OLG is formed with two Log-Gabor filters with mutually orthogonal directions, which can be expressed by (13). The OLG proposed reduces the amount of computation while avoiding the redundancy of features, and thus can improve the running speed under the premise of ensuring the integrity of features.

$$OL(\rho, \theta, s, o) = |L(\rho, \theta, s, o) - L\left(\rho, \theta \pm \frac{\pi}{2}, s, o\right)| \quad (13)$$

To validate the the effect of the OLG algorithm on removing redundant features, the FAST algorithm is applied based on the phase congruency representations processed in the two strategies, respectively, and the results are shown in Figure 4. Subfigures (a)–(d) denote features detected based on LG algorithm, and (e)–(h) are based on the OLG algorithm. To illustrate the effectiveness of the proposed algorithm, image pairs SAR1/OPT1 and SAR2/OPT2 of different landforms are selected for comparison. It can be observed that the OLG algorithm performs better in depicting the image edges, and some of the isolated points were removed compared with the results of LG. The feature points obtained by the orthogonal method tend to be more recognizable, which makes it more likely to obtain the correct match in the subsequent processing.

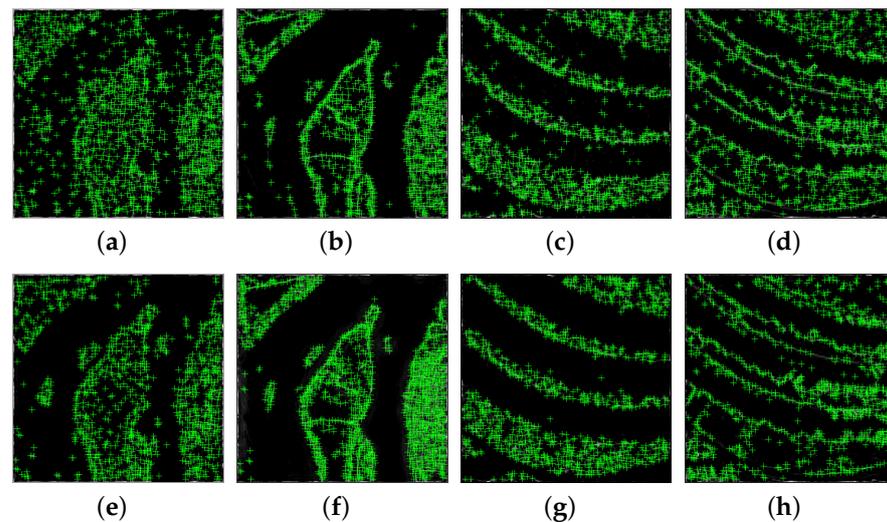


Figure 4. Features detected with LG/OLG. (a) LG-SAR1. (b) LG-OPT1. (c) LG-SAR2. (d) LG-OPT2. (e) OLG-SAR1. (f) OLG-OPT1. (g) OLG-SAR2. (h) OLG-OPT2.

2.3. Sector Descriptor Construction

Referring to the RIFT algorithm, the descriptors construction of EC-RIFT is performed on the MIM. It accumulates the amplitudes $A_{s_0}(x, y)$ in the same direction at N_s scales first to obtain $A_o(x, y)$, and then the MIM is constituted with the direction index value ω , which represents the maximum values of these pixels.

To enhance the robustness and uniqueness of the descriptors, the GLOH-like approach is chosen to construct the descriptors in this paper [27]. For the obtained features based on edge consistency, the statistical histogram in log-polar coordinates enables the descriptors to be more sensitive to near points than distant points. As shown in Figure 5, the pixels in the neighborhood of keypoints are divided into log-polar coordinate concentric circles with the radius set to $0.25 r$ and $0.75 r$, supposing that the radius of neighborhood is r . If the area of the divided sector is too small, the features are not sufficient, while the larger section will increase the descriptor dimension and the burden of computation. The pixels are divided into eight equal parts in the angular, each equal part being $\pi/4$, so that 16 sectors and 1 circle are formed, for a total of 17 image sub-blocks. The points on the same ring are close to each other, so the feature points are equally discriminated. Each block computes a histogram of six directions, resulting in a descriptor vector of $17 \times 6 = 102$ dimensions. To simplify the calculation, if some descriptor component exceeds the threshold value, the component is set to the threshold size, and the threshold value is empirically set as 0.2 in this paper. Finally, the descriptor vector is normalized.

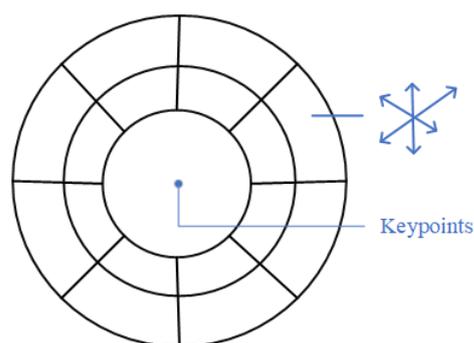


Figure 5. The descriptors of EC-RIFT method.

2.4. Feature Matching and Outlier Removal

An effective algorithm to match feature points is the Nearest Neighbor Distance Ratio (NNDR) method, which is proposed by Lowe. It screens out local features by comparing the distances of nearest neighbor and second nearest neighbor features, and filters out features with low discrimination. A feature point is more likely to be correctly matched if the point most similar to its descriptor is close, and the point less similar to it is farther away.

After the feature matching, there are still some mismatched point pairs, so the EC-RIFT algorithm chooses the FSC algorithm to remove the outliers and estimate the parameters of the affine transformation model.

3. Experiments and Results

In this section, we first introduce the datasets and evaluation metrics. Subsequently, we investigate the effectiveness of the proposed method in stages. Finally, aiming at proving the superiority of the EC-RIFT, the results are presented and evaluated by comparing them with four representative multi-source image registration methods.

3.1. Datasets

Two types of datasets are used to evaluate the registration performance of the EC-RIFT algorithm in this paper. The first is formed from seven pairs of SAR images and optical images with various sizes and sources. The image pairs in other modalities are also employed to assess the generalization ability of proposed algorithm. Both datasets are described in detail.

The SAR and optical images are obtained from different sensors, such as Terra SAR-X, Gaofen satellites, Google Earth, and Landsat satellites and aerial vehicles, covering various resolutions and sizes. The multi-source image dataset consists of infrared–optical, day–night, depth–optical pairs, widely employed in remote sensing image registration. All image pairs capture various terrains, including urban areas, fluvial deposits, rivers, and farmland, and the images exhibit significant geometric and radiometric differences, posing challenges for registration. The data are shown in Figure 6.

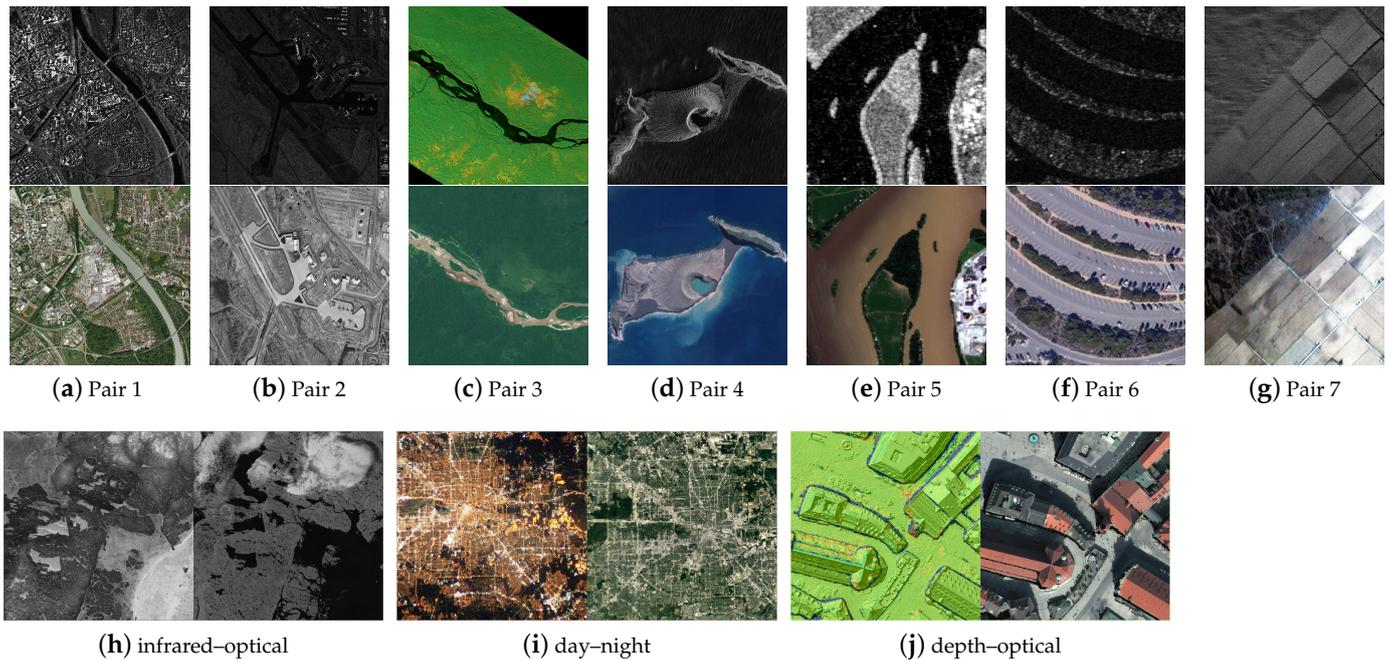


Figure 6. Dataset presentation. The (a–g) denote sar-optical image pairs of different landforms. And the (h–j) represent infrared-optical, day-night, depth-optical landforms, respectively.

3.2. Evaluation Criterion

For quantitative evaluation, the repeatability [28], root mean square error (RMSE), and running time are selected to evaluate the performance of the proposed method.

Repeatability [28] is often used as a metric to evaluate the quality of detected features, which is defined to represent the rate of potential corresponding points in all detected points. The repeatability can be calculated as Equation (14):

$$\text{Repeatability} = \frac{N_c}{(N_1 + N_2)/2} \quad (14)$$

where N_c represents the number of potential corresponding points, and N_1 and N_2 represent the total number of feature points detected in the reference image and moving image, respectively. If a feature point meets the conditions shown in Formula (15), it is considered as a potential corresponding point. Here, i represents the index of keypoint, (x_{1i}, y_{1i}) is the keypoint coordinates of the reference image, and (x_{2i}, y_{2i}) is the keypoint coordinates obtained after affine transformation of the moving image.

$$\sqrt{(x_{2i} - x_{1i})^2 + (y_{2i} - y_{1i})^2} < 2 \quad (15)$$

The root mean square error (RMSE) is used to evaluate the precision of the registration results. RMSE is defined as (16):

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (x_{2i} - x_{1i})^2 + (y_{2i} - y_{1i})^2}{n}} \quad (16)$$

where i is the pixel index, n is the number of matching pairs, (x_{1i}, y_{1i}) is the keypoint coordinates of the reference image, and (x_{2i}, y_{2i}) is the keypoint coordinates obtained after affine transformation of the image to be registered. The smaller the value of RMSE, the higher the precision of registration.

3.3. Registration Results and Analysis

To demonstrate the effect of each processing section on the registration results, experiments are designed to evaluate the registration results. The comparison experiments are conducted with HAPCG [17], OS-SIFT [7], LNIFT [29], and RIFT [16] algorithms to prove that the proposed method has better performance in registration. The experiments are implemented using MATLAB 2021b on a platform with Advanced Micro Devices, Inc. (AMD)-Ryzen-5-3400G with Radeon Vega Graphics 3.70-GHz CPU, and 16-GB RAM.

3.3.1. Comparison Experiment of Preprocessing Algorithms

The preprocessing step can help in realizing the registration of multi-source images with large radiometric differences. To demonstrate the effectiveness of the algorithm applied in this paper, Pair 6 is chosen for the experiment, whose similarity to the original data is too low to be registered by all algorithms. Figure 7 depicts the registration results obtained with/without the denoising and enhancement process. Subfigure (a) represents the matching result of original data, which is obviously unable to be registered. The SAR and optical images in (b) and (c) are preprocessed, respectively. Even if one of the images is processed, the image pairs failed to register. For the denoised SAR image and enhanced optical image, the corresponding points are matched correctly, as shown in (d). It is clear that the noise in the original SAR image and the texture in the optical image have a strong impact on registration, and more correct matching pairs are obtained after preprocessing.

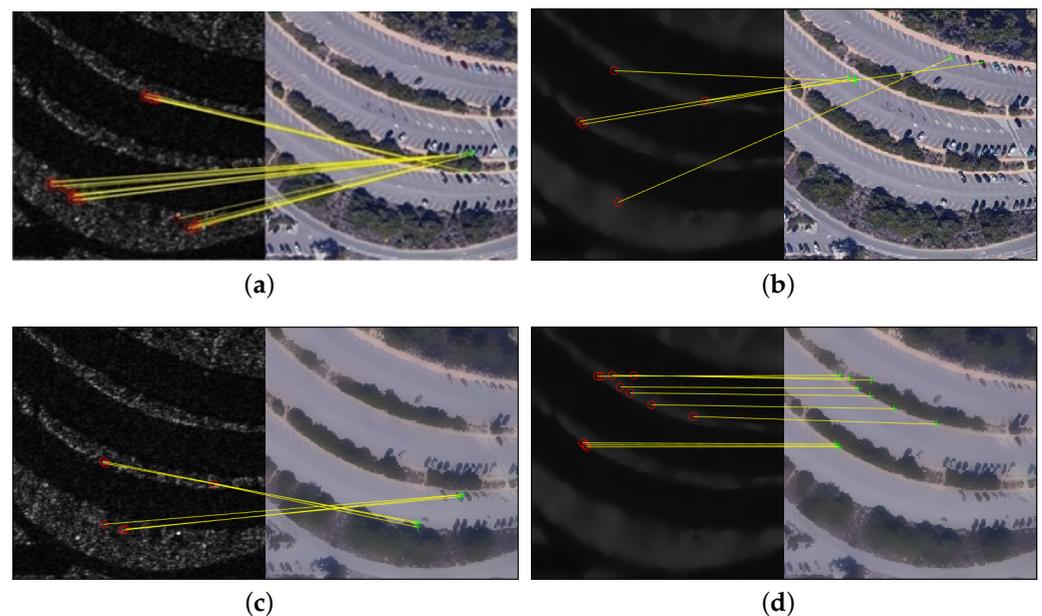


Figure 7. Registration results with/without preprocessing. (a) Without preprocessing. (b) SAR image is denoised. (c) Optical image is enhanced. (d) Both images are preprocessed.

In addition, comparison experiments are conducted on images preprocessed with aforementioned steps. As listed in Table 2, our method performs the best among the four methods in terms of RMSE and running time.

It should be noted that this step is optional for remote sensing image registration. There is serious radiation and geometric distortion in the data for this experiment, which brings greater challenges to registration, and the proposed steps contribute to extracting the corresponding features.

3.3.2. Comparison Experiment of OLG and LG

Since the OLG algorithm can reduce redundancy in feature detection, the repeatability is chosen to evaluate its effectiveness. A higher repeatability indicates a higher probability of extracting the corresponding features. The results are shown in Figure 8.

Table 2. Comparison of preprocessed data registration results.

Data	Method	RMSE/px	Running Time/s
Pair 6	HAPCG	0.88	2.8
	OS-SIFT	× ¹	×
	LNIFT	1.39	15.45
	RIFT	0.99	2.26
	EC-RIFT	0.95	1.99
Pair 7	HAPCG	1.59	2.56
	OS-SIFT	×	×
	LNIFT	1.22	12.61
	RIFT	1.11	1.79
	EC-RIFT	0.92	1.76

¹ The × means that the algorithm fails.

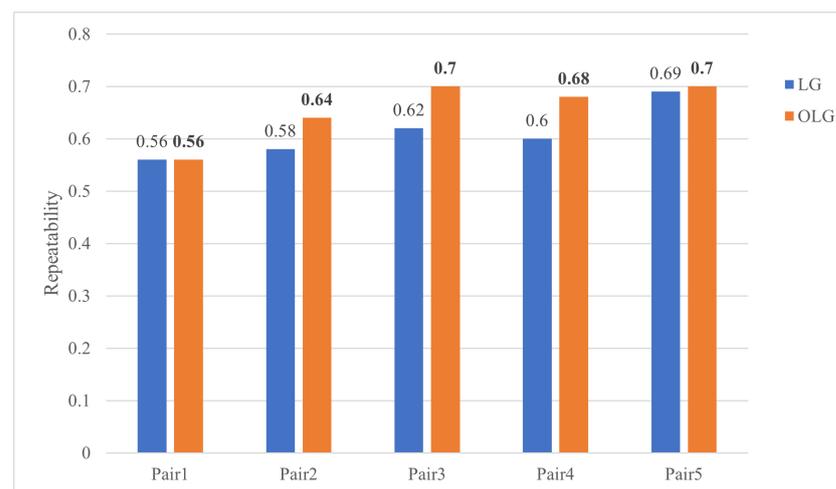


Figure 8. Repeatability of OLG and LG. The bold numbers represent the higher scores.

The OLG algorithm achieves scores higher than or equal to LG on all images, demonstrating the contribution to improving the registration accuracy.

In addition, the feature detection time and registration precision of OLG are tested in this experiment, and the results are listed in Table 3. The bold numbers in the table denote the best results. As can be seen from Table 3, for all image pairs, OLG can reduce the feature detection time by nearly half, and the accuracy is improved on all data.

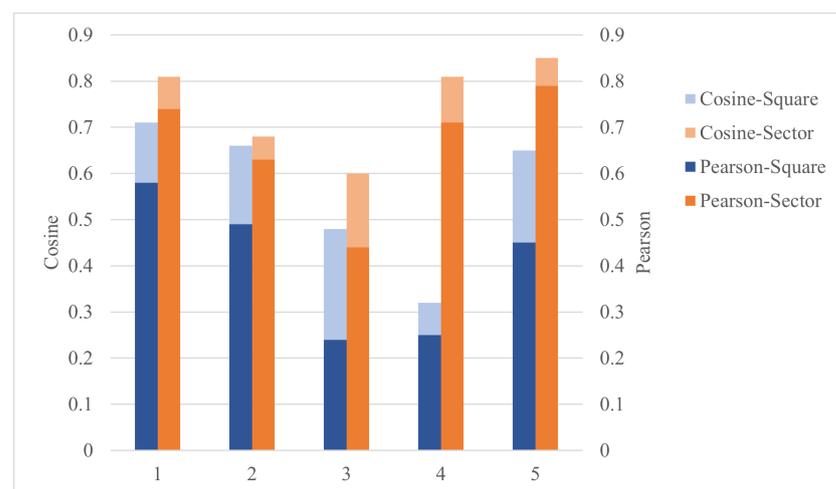
Table 3. Registration results of OLG and LG.

Data	Filter	Detection Time/s	RMSE/px
Pair 1	LG	0.99	1.41
	OLG	0.58	1.35
Pair 2	LG	3.20	1.36
	OLG	1.81	1.32
Pair 3	LG	1.50	1.32
	OLG	1.04	1.27
Pair 4	LG	0.99	1.34
	OLG	0.65	1.33
Pair 5	LG	0.23	1.40
	OLG	0.15	1.29

3.3.3. Comparison Experiment of Square Descriptor and Sector Descriptor

The purpose of this experiment is to compare the performance of square descriptors and the sector descriptors used in this paper. The similarity between the descriptor vectors on paired features determines the success rate of matching. In order to illustrate that the proposed descriptors are better capable of extracting the common information, we chose two metrics to measure the similarity of the paired descriptors. Cosine similarity and Pearson correlation coefficient are commonly used methods to measure the similarity between two vectors, both in the range of $[-1, 1]$. A value of 1 means that the vectors are completely similar, and -1 indicates the opposite.

Five pairs of points are selected randomly from Pair 1 for this experiment, then the descriptors are built in square and sector methods, respectively. The similarity of vectors are shown in the Figure 9. The bars in light color represent cosine similarity, and those in dark color represent Pearson correlation coefficient of two methods. The sector descriptor (the orange bars) outperforms the square descriptor (the blue bars) on all points selected. It thus proves that descriptors based on sector neighborhood fit the corresponding points matching better.

**Figure 9.** The similarity of MIM.

Furthermore, the comparison experiment of registration results is shown in Figure 10. For the reliability of the experiment, all parts of the method are kept consistent except for the descriptor construction step. This result demonstrates that our descriptors have a higher robustness and are thus conducive to multi-source image registration.

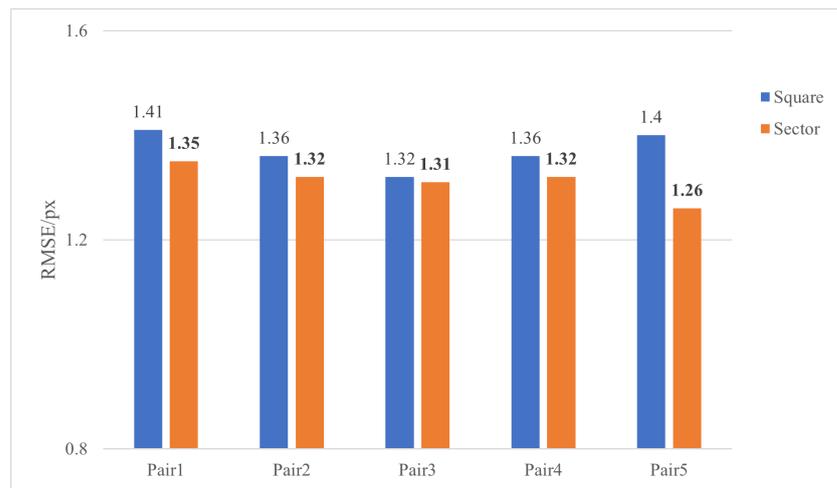


Figure 10. Comparison of OLG and LG. The bold numbers represent the best results.

3.3.4. Comparative Results on SAR–Optical Data

To analyze the performance of the proposed method, we firstly conduct experiments on SAR and optical dataset, which is most commonly used in practice. For the fairness of the experiment, all methods use the same matching method and parameters are set under the recommendations of their authors.

Figure 11 represents the matching results of the EC-RIFT algorithm. It is clear that the sufficient features are matched correctly, confirming the effectiveness of the proposed method for SAR and optical image registration. Table 4 shows the quantitative results of the five methods. The EC-RIFT method is capable of steadily registering image pairs with lower error and time.

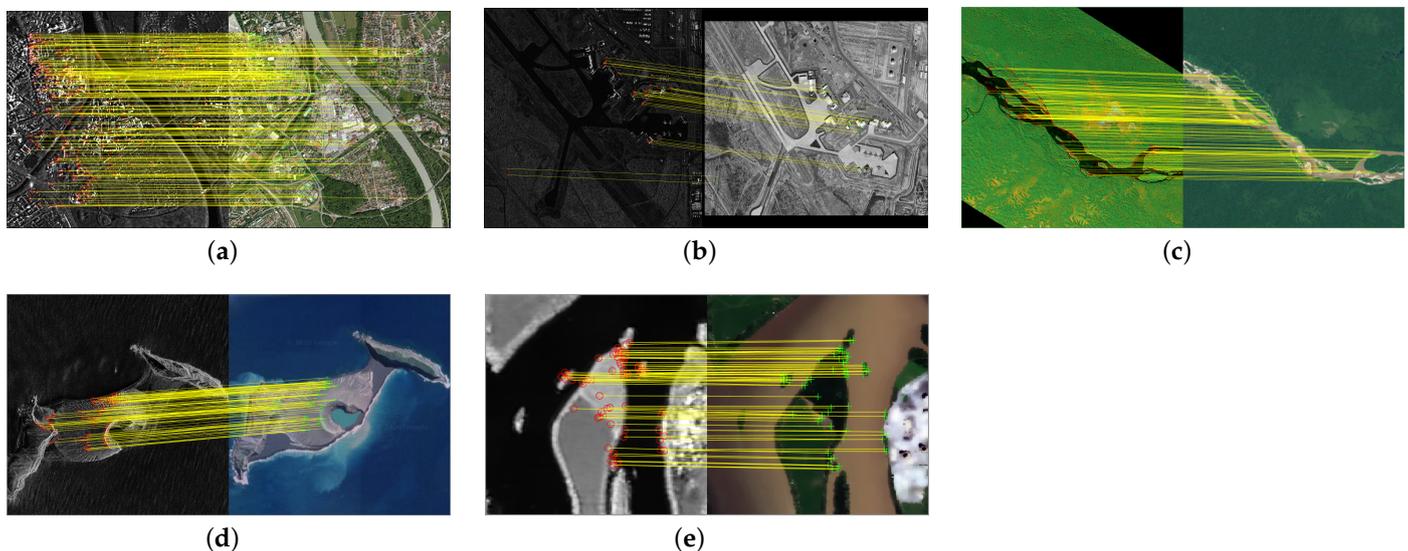


Figure 11. Matching results for SAR–optical images. (a) Pair 1. (b) Pair 2. (c) Pair 3. (d) Pair 4. (e) Pair 5.

Table 4. Comparison with other methods. The bold numbers represent the best performance.

Data	Method	RMSE/px	Running Time/s
Pair 1	HAPCG	1.95	9.98
	OS-SIFT	1.38	9.15
	LNIFT	1.35	54.64
	RIFT	1.37	10.45
	EC-RIFT	1.35	8.81
Pair 2	HAPCG	1.95	26.13
	OS-SIFT	× ¹	×
	LNIFT	×	×
	RIFT	1.34	14.26
	EC-RIFT	1.19	11.01
Pair 3	HAPCG	1.93	13.64
	OS-SIFT	NAN ²	NAN
	LNIFT	1.29	62.64
	RIFT	1.32	6.41
	EC-RIFT	1.28	5.06
Pair 4	HAPCG	1.97	7.98
	OS-SIFT	×	×
	LNIFT	×	×
	RIFT	1.35	9.36
	EC-RIFT	1.24	9.27
Pair 5	HAPCG	1.76	2.12
	OS-SIFT	NAN	NAN
	LNIFT	1.33	10.01
	RIFT	1.36	3.95
	EC-RIFT	1.31	2.39

¹ The × means that the algorithm fails. ² The NAN represents the number of correct matching is less than five.

The performance of OS-SIFT is the most vulnerable. For Pairs 2–5, there are not enough correct matching pairs obtained by OS-SIFT, which we find unconvincing. The poor performance of the OS-SIFT algorithm is due to the lack of edge retention capacity, which reduces the accuracy of feature detecting and matching, and thus fails on the data with significant radiation variations. In terms of RMSE, the proposed method outperforms other methods in all image pairs. For Pair 1, the proposed EC-RIFT obtains the same value of RMSE as LNIFT in Pair 1, but it consumes less time. For Pair 2, the results obtained by our method are 38.97% and 11.19% better than those obtained by HAPCG and RIFT, respectively, achieving significant accuracy improvement. The LNIFT method fails on Pair 2 and Pair 4, and the RMSE of other pairs are slightly higher than the method proposed. The running times of the HAPCG method for Pair 4 and Pair 5 are lower than our method, but as shown in Figure 12, the average time for all image pairs of our method is 7.31 s, which is still the shortest among these methods. It should be noted that the OS-SIFT is annotated as *NAN* because it obtains valid results only on Pair 1 and takes longer than our algorithm on it. Taken together, the proposed method outperforms these popular methods in terms of robustness.

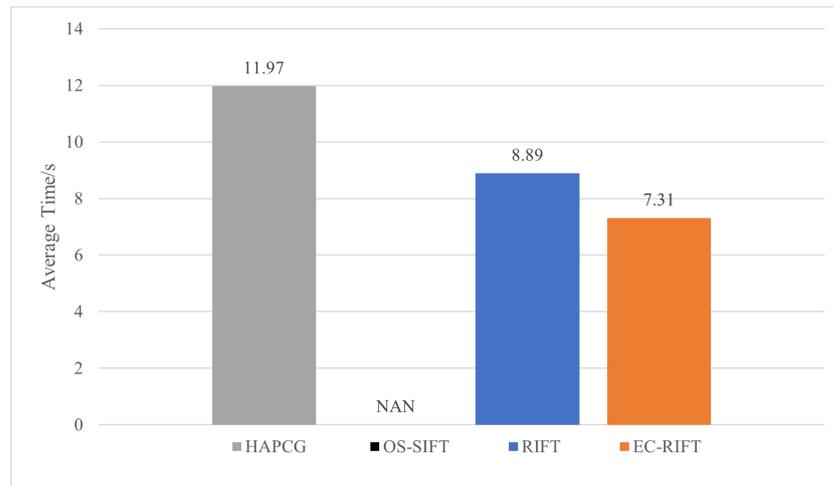


Figure 12. Average running time of registration methods.

3.4. Comparative Results on Multi-Source Data

The experimental results on SAR–optical images demonstrate the effectiveness and superiority of our method. As an extension, we focus on the registration performance of other multi-source images in this subsection. It should be noted that the noise of multi-source data used in this experiment is weak, so there is no additional processing performed in the pre-processing step. The visual results are shown in Figure 13. The EC-RIFT method successfully matched a large number of points on these image pairs.

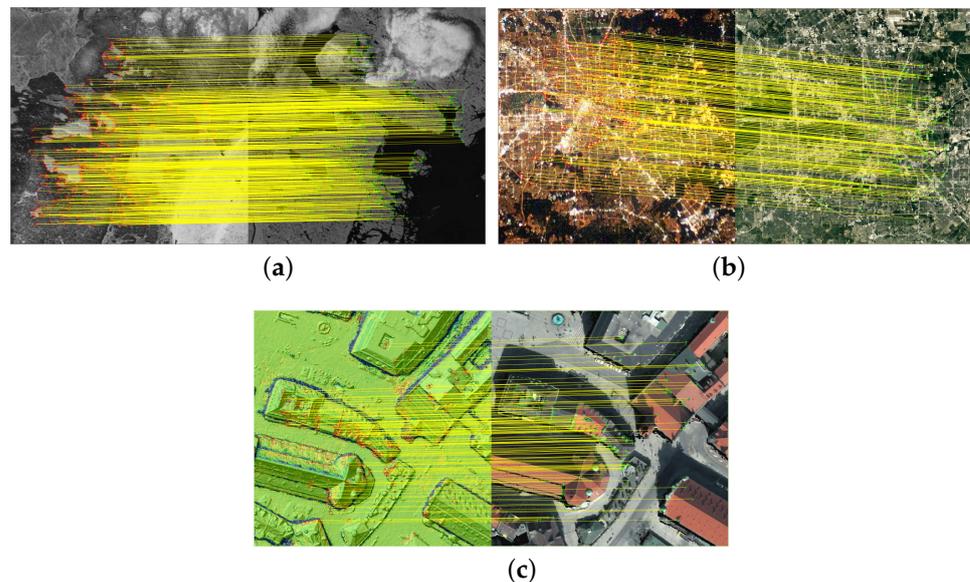


Figure 13. Matching results for multi-source images. (a) Infrared–optical. (b) Day–night. (c) Depth–optical.

Table 5 lists the comparative registration results on multi-source images. The LNIFT and OS-SIFT fail on day–night data, and our method obtains robust registration results. The RMSE of the proposed method is the same as that of OS-SIFT and LNIFT on infrared–optical data, but the running time is shorter. For the day–night and depth–optical data, the EC-RIFT method achieves the highest accuracy.

Table 5. Quantitative results for multi-source images. The bold numbers denote the best results.

Data	Method	RMSE/px	Running Time/s
Infrared–optical	HAPCG	1.65	18.66
	OS-SIFT	1.16	16.71
	LNIFT	1.29	64.73
	RIFT	1.16	14.70
	EC-RIFT	1.16	13.13
Day–night	HAPCG	1.83	9.79
	OS-SIFT	× ¹	×
	LNIFT	×	×
	RIFT	1.27	12.37
	EC-RIFT	1.26	10.68
Depth–optical	HAPCG	1.95	8.90
	OS-SIFT	1.41	8.80
	LNIFT	1.36	55.06
	RIFT	1.36	11.98
	EC-RIFT	1.33	11.50

¹ The × means that the algorithm fails.

4. Discussion

Taking both the quantitative and the qualitative results into account, the experimental results in Section 3 illustrate that the proposed method achieves better registration performance for multi-source images. In this section, we first validate the idea that smoothing textures and noises can enhance the edge features, which are more stable between multi-source images. Additionally, the rotation and scale invariances of the method are analyzed, leading to the limitations and future work expectations.

4.1. The Effect of Noises on MIM

Since image descriptors are built based on the histogram statistics of MIM, the distribution in the neighborhood of feature points on MIM determines the probability of corresponding points being matched. Therefore, we compare the MIMs of original images and processed images to research the effect of this step.

As shown in Figure 14, the influence of speckle noises are significant in the SAR image, and the irregular interference directly disturbs the robustness of descriptors. The textures which are not present in the SAR image also exacerbate the difference between MIMs. The MIMs of processed image pairs are much more similar, leading to higher registration accuracy.

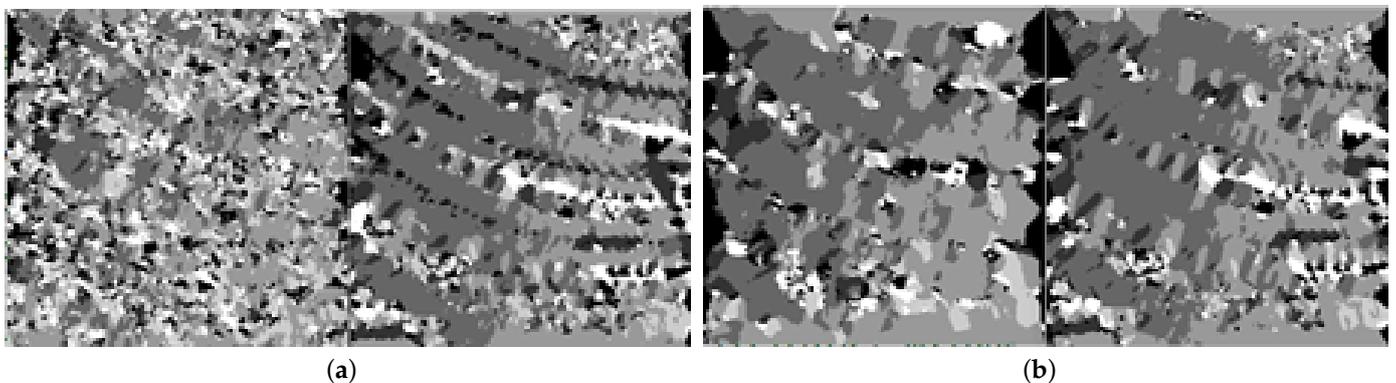


Figure 14. MIMs of images. (a) MIM of original image pairs. (b) MIM of processed image pairs.

Note that the preprocessing methods in this paper are selected according to the characteristics of the images. For images of other modalities, the idea of eliminating noise

and texture differences between image pairs still holds, but the specific algorithms need to be redesigned.

4.2. Fine-Registration and Considerable Difference

The imaging spatial resolution difference between satellite and drone is great, and the imaging level is uneven between different satellites. The scale is different depending on the height, the rotation is different depending on the orbit, and the distortion is different depending on the modality. Therefore, the multi-sensor data for situational awareness needs to be fine-registered first in the space–air–ground integrated network [30].

Our research is based on the assumption that remote sensing images have been registered with geographic information, which means that there are mainly translation differences between the input image pairs. Actually, there is only a tiny scale difference on the coarse-registered images calibrated by the location or attitude information, and the rotation angle is subtle, in which case our method works well. The feature matching results of image pairs with scale and rotation differences are shown in Figure 15. The resolution of the SAR image in (a) is reduced from $500 \text{ px} \times 492 \text{ px}$ to $400 \text{ px} \times 394 \text{ px}$, while the resolution of the optical image remains the original. The SAR images of two sizes are manually rotated clockwise by 5 degrees in (b) and (c), respectively. For indistinctive differences in scale and rotation, whether one or both exist in the image pairs, the proposed method can obtain enough matching pairs.

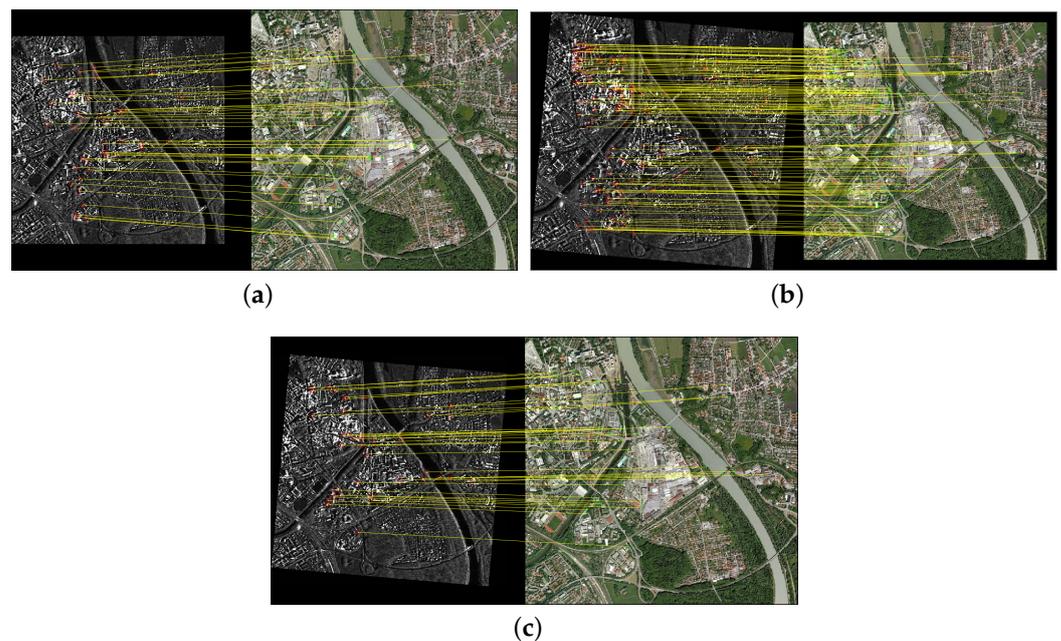


Figure 15. Matching results with scale and rotation differences. (a) Scale difference. (b) Rotation difference. (c) Rotation difference.

However, our method is sensitive to significant scale differences and rotation transformations, so it may fail in these scenes. In terms of rotation-invariance, the solution has been proposed in [31], but it is computationally heavy and time-consuming. In terms of scale-invariance, methods based on the image pyramid [32] may help to improve the registration performance. In future work, we will focus on proposing a more efficient approach to improve the feature invariance of significant scale and rotation.

5. Conclusions

Aiming at handling the challenge of extracting consistent features from multi-source images with severe geometric and radiometric distortions, this paper proposes a robust registration method based on the edge consistency features. We first design the denois-

ing and enhancing algorithms according to the characteristics of remote sensing images to reduce the interference of redundant information on registration. Then the orthogonal method is adopted for feature extraction to improve computational efficiency. Finally, the sector-neighborhood descriptor is constructed to obtain a more robust representation. Compared with other representative registration methods, the proposed method obtains higher speed and accuracy, which can effectively realize the registration task of SAR and optical images. Due to the potential of real-time performance, the improved algorithm is expected to be applied in edge intelligence analysis [33] and on-orbit satellite platforms [34]. Furthermore, the proposed method is proven to have a superior generalization ability to be applied to images in other modalities. However, the proposed method may not work properly in the case of large rotation and scale deviation in the images, and we will explore the technology of geometric deformation invariance in the future work.

Author Contributions: Conceptualization, Y.Z.; methodology, Y.Z. and Z.H.; validation, Y.Z.; investigation, Z.D. and C.H.; writing—original draft preparation, Z.H. and N.L.; writing—review and editing, N.L. and C.C.; supervision and suggestions, L.C., Z.D., and C.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program of China (2020YFB1807500), the National Natural Science Foundation of China (62072360, 62001357, 62172438, 61901367), the key research and development plan of Shaanxi province (2021ZDLGY02-09, 2023-GHZD-44, 2023-ZDLGY-54), the Natural Science Foundation of Guangdong Province of China (2022A1515010988), Key Project on Artificial Intelligence of Xi'an Science and Technology Plan (23ZDCYJSGG0021-2022, 23ZDCYYYCJ0008, 23ZDCYJSGG0002-2023), Xi'an Science and Technology Plan (20RGZN0005) and the Proof-of-concept fund from Hangzhou Research Institute of Xidian University (GNYZZ2023QC0201).

Data Availability Statement: Not applicable.

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

Abbreviations

The following abbreviations are used in this manuscript:

SAR	Synthetic Aperture Radar
EC-RIFT	Edge Consistency Radiation-variation Insensitive Feature Transform
NLM	Non-local Mean
CoF	Co-occurrence Filter
OLG	Orthogonal Log-Gabor
RMSE	Root Mean Square Error

References

- Chen, C.; Wang, C.; Liu, B.; He, C.; Cong, L.; Wan, S. Edge Intelligence Empowered Vehicle Detection and Image Segmentation for Autonomous Vehicles. *IEEE Trans. Intell. Transp. Syst.* **2023**, 1–12. [[CrossRef](#)]
- Lv, N.; Zhang, Z.; Li, C. A hybrid-attention semantic segmentation network for remote sensing interpretation in land-use surveillance. *Int. J. Mach. Learn. Cybern.* **2023**, *14*, 395–406. [[CrossRef](#)]
- Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [[CrossRef](#)]
- Chen, H.M.; Varshney, P.K.; Arora, M.K. Performance of Mutual Information Similarity Measure for Registration of Multitemporal Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 2445–2454. [[CrossRef](#)]
- Ye, Y.; Bruzzone, L.; Shan, J.; Bovolo, F.; Zhu, Q. Fast and Robust Matching for Multimodal Remote Sensing Image Registration. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9059–9070. [[CrossRef](#)]
- Ma, W.; Wen, Z.; Wu, Y.; Jiao, L.; Gong, M.; Zheng, Y.; Liu, L. Remote Sensing Image Registration with Modified SIFT and Enhanced Feature Matching. *IEEE Geosci. Remote Sens. Lett.* **2016**, *14*, 3–7. [[CrossRef](#)]
- Xiang, Y.; Wang, F.; You, H. OS-SIFT: A Robust SIFT-Like Algorithm for High-Resolution Optical-to-SAR Image Registration in Suburban Areas. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 3078–3090. [[CrossRef](#)]
- Gao, C.; Li, W.; Tao, R.; Du, Q. MS-HLMO: Multiscale Histogram of Local Main Orientation for Remote Sensing Image Registration. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–14. [[CrossRef](#)]
- Ma, W.; Zhang, J.; Wu, Y.; Jiao, L.; Zhu, H.; Zhao, W. A Novel Two-Step Registration Method for Remote Sensing Images Based on Deep and Local Features. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4834–4843. [[CrossRef](#)]

10. Zhou, L.; Ye, Y.; Tang, T.; Nan, K.; Qin, Y. Robust Matching for SAR and Optical Images Using Multiscale Convolutional Gradient Features. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [[CrossRef](#)]
11. Zhuoqian, Y.; Tingting, D.; Yang, Y. Multi-temporal Remote Sensing Image Registration Using Deep Convolutional Features. *IEEE Access* **2018**, *6*, 38544–38555.
12. Zhang, H.; Ni, W.; Yan, W.; Xiang, D.; Bian, H. Registration of Multimodal Remote Sensing Image Based on Deep Fully Convolutional Neural Network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 3028–3042. [[CrossRef](#)]
13. Hughes, L.; Marcos, D.; Lobry, S.; Tuia, D.; Schmitt, M. A deep learning framework for matching of SAR and optical imagery. *ISPRS J. Photogramm. Remote Sens.* **2020**, *169*, 166–179. [[CrossRef](#)]
14. Kovesi, P. Phase congruency: A low-level image invariant. *Psychol. Res.* **2000**, *64*, 136–148. [[CrossRef](#)]
15. Ye, Y.; Shan, J.; Bruzzone, L.; Shen, L. Robust Registration of Multimodal Remote Sensing Images Based on Structural Similarity. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2941–2958. [[CrossRef](#)]
16. Li, J.; Hu, Q.; Ai, M. RIFT: Multi-Modal Image Matching Based on Radiation-Variation Insensitive Feature Transform. *IEEE Trans. Image Process.* **2020**, *29*, 3296–3310. [[CrossRef](#)]
17. Yao, Y.; Zhang, Y.; Wan, Y.; Liu, X.; Guo, H. Heterologous Images Matching Considering Anisotropic Weighted Moment and Absolute Phase Orientation. *Geomat. Inf. Sci. Wuhan Univ.* **2021**, *46*, 1727.
18. Fan, Z.; Liu, Y.; Liu, Y.; Zhang, L.; Zhang, J.; Sun, Y.; Ai, H. 3MRS: An Effective Coarse-to-Fine Matching Method for Multimodal Remote Sensing Imagery. *Remote Sens.* **2022**, *14*, 478. [[CrossRef](#)]
19. Sui, H.; Liu, C.; Gan, Z. Overview of multi-modal remote sensing image matching methods. *J. Geod. Geoinf. Sci.* **2022**, *51*, 1848.
20. Rosten, E.; Drummond, T. Machine learning for high-speed corner detection. In Proceedings of the ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, 7–13 May 2006; pp. 430–443.
21. Wu, Y.; Ma, W.; Gong, M.; Su, L.; Jiao, L. A Novel Point-Matching Algorithm Based on Fast Sample Consensus for Image Registration. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 43–47. [[CrossRef](#)]
22. Buades, A.; Coll, B.; Morel, J. A non-local algorithm for image denoising. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 2, pp. 60–65.
23. Jevnisek, R.J.; Avidan, S. Co-occurrence Filter. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 3816–3824.
24. Yang, H.; Fu, Y.; Zeng, J. Face recognition algorithm based on orthogonal Log-Gabor filter binary mode. *J. Intell. Syst.* **2019**, *14*, 330–337.
25. Oppenheim, A.; Lim, J. The importance of phases in signal. *IEEE Trans. Comput. Sci.* **1981**, *69*, 333–382. [[CrossRef](#)]
26. Morrone, M.C.; Owens, R.A. Feature detection from local energy. *Pattern Recognit. Lett.* **1987**, *6*, 303–313. [[CrossRef](#)]
27. Mikolajczyk, K.; Schmid, C. A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1615–1630. [[CrossRef](#)] [[PubMed](#)]
28. Mikolajczyk, K.; Schmid, C. An Affine Invariant Interest Point Detector. In Proceedings of the ECCV 2002, 7th European Conference on Computer Vision, Copenhagen, Denmark, 28–31 May 2002; Volume LNCS 2350.
29. Li, J.; Xu, W.; Shi, P.; Zhang, Y.; Hu, Q. LNIFT: Locally Normalized Image for Rotation Invariant Multimodal Feature Matching. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–14. [[CrossRef](#)]
30. Zhang, Y.; Chen, C.; Liu, L.; Lan, D.; Jiang, H.; Wan, S. Aerial Edge Computing on Orbit: A Task Offloading and Allocation Scheme. *IEEE Trans. Netw. Sci. Eng.* **2023**, *10*, 275–285. [[CrossRef](#)]
31. Han, C. RI-LPOH: Rotation-Invariant Local Phase Orientation Histogram for Multi-Modal Image Matching. *Remote Sens.* **2022**, *14*, 4228.
32. Gao, C.; Li, W. Multi-scale PIIFD for Registration of Multi-source Remote Sensing Images. *J. Beijing Inst. Technol.* **2021**, *30*, 12.
33. Chen, C.; Yao, G.; Liu, L.; Pei, Q.; Song, H.; Dustdar, S. A Cooperative Vehicle-Infrastructure System for Road Hazards Detection with Edge Intelligence. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 5186–5198. [[CrossRef](#)]
34. Hnatushenko, V.; Kogut, P.; Uvarov, M. Variational approach for rigid co-registration of optical/SAR satellite images in agricultural areas. *J. Comput. Appl. Math.* **2022**, *400*, 113742. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.