remote sensing

MDPI

Article

RTV-SIFT: Harnessing Structure Information for Robust Optical and SAR Image Registration

Siqi Pang [1], Junyao Ge [1], Lei Hu [1], Kaitai Guo [1], Yang Zheng [1], Changli Zheng [2], Wei Zhang [2] and Jimin Liang [1,*]

[1] Key Laboratory of Collaborative Intelligence Systems, Ministry of Education of China, School of Electronic Engineering, Xidian University, Xi'an 710071, China; sqpang@stu.xidian.edu.cn (S.P.); 21021110370@stu.xidian.edu.cn (J.G.); hulei@xidian.edu.cn (L.H.); ktguo@xidian.edu.cn (K.G.); zhengy@xidian.edu.cn (Y.Z.)

[2] Science and Technology on Electronic Information Control Laboratory, Southwest China Research Institute of Electronic Equipment, Chengdu 610036, China; zszcl@163.com (C.Z.); rruun@126.com (W.Z.)

* Correspondence: jimleung@mail.xidian.edu.cn

Abstract: Registration of optical and synthetic aperture radar (SAR) images is challenging because extracting located identically and unique features on both images are tricky. This paper proposes a novel optical and SAR image registration method based on relative total variation (RTV) and scale-invariant feature transform (SIFT), named RTV-SIFT, to extract feature points on the edges of structures and construct structural edge descriptors to improve the registration accuracy. First, a novel RTV-Harris feature point detection method by combining the RTV and the multiscale Harris algorithm is proposed to extract feature points on both images' significant structures. This ensures a high repetition rate of the feature points. Second, the feature point descriptors are constructed on enhanced phase congruency edge (EPCE), which combines the Sobel operator and maximum moment of phase congruency (PC) to extract edges from structured images that enhance robustness to nonlinear intensity differences and speckle noise. Finally, after coarse registration, the position and orientation Euclidean distance (POED) between feature points is utilized to achieve fine feature point matching to improve the registration accuracy. The experimental results demonstrate the superiority of the proposed RTV-SIFT method in different scenes and image capture conditions, indicating its robustness and effectiveness in optical and SAR image registration.

Keywords: image registration; relative total variation (RTV); structure extraction; phase congruency (PC); optical and synthetic aperture radar (SAR) images

Citation: Pang, S.; Ge, J.; Hu, L.; Guo, K.; Zheng, Y.; Zheng, C.; Zhang, W.; Liang, J. RTV-SIFT: Harnessing Structure Information for Robust Optical and SAR Image Registration. *Remote Sens.* **2023**, *15*, 4476. https://doi.org/10.3390/rs15184476

Academic Editors: Jie Feng, Gui-Song Xia, Xiangrong Zhang, Gong Cheng, Lichao Mou and Dusan Gleich

Received: 30 July 2023
Revised: 3 September 2023
Accepted: 8 September 2023
Published: 12 September 2023

1. Introduction

Optical and synthetic aperture radar (SAR) image registration serves as the basis for many tasks in remote sensing image analysis [1–3]. Multimodal remote sensing image registration has been extensively studied, with intensity-based, feature-based, and learning-based methods being the main categories [4]. Intensity-based methods use various similarity metrics, such as mutual information [5], normalized cross-correlation coefficient [6], and cross-cumulative residual entropy [7], to match image patches. However, these methods are often limited in performance when applied to multimodal images with large radiation differences, such as optical and SAR Images.

With the development of deep learning technology, some learning-based methods have been proposed [8–12]) to perform registration on multi-modal images by learning image features. However, due to the difficulty in obtaining sufficient multi-modal remote sensing images and ground truth for training and testing [13], and the difficulty in achieving end-to-end registration implementation, these methods are not widely used. In addition, learning-based methods have poor adaptability to different remote sensing images and their training performance is highly dependent on computer hardware [14], so they have not been widely applied. In contrast, feature-based methods are faster and more flexible.

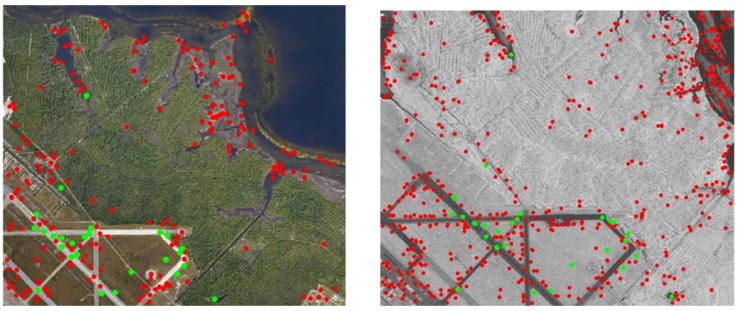
In this work, the feature-based approach is adopted to tackle the problem of optical and SAR Image registration, which faces two fundamental challenges:

1. How to find homology and effective feature points on optical and SAR images that have significantly different inherent natures?
2. How to overcome nonlinear intensity differences and construct feature descriptors that are similar at corresponding points but distinguishable at non-corresponding points?

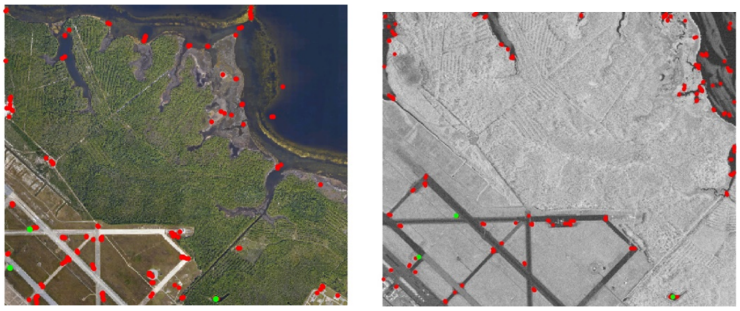
These two issues must be considered together in order to obtain a good registration performance. To clarify this point, Figure 1 shows the feature points detected and matched by the PSO-SIFT [15], SAR-SIFT [16], and our proposed RTV-SIFT method in a pair of optical and Gaofen-3 (GF-3) SAR images. The matched points are marked in green and the unmatched points are marked in red. The PSO-SIFT method directly adopts the scale-invariant feature transform (SIFT) detector [17], which extracts many unrepeatable feature points (the large number of red points in Figure 1a) in both optical and SAR images because the second-order partial operation in the Difference of Gaussian (DoG) [18] is sensitive to noise and details [16]. This will not only increase the computational burden of the following feature matching task but also will increase the incorrect matching rate. To overcome this problem, Zhang et al. [19] improved the anti-noise capability of SIFT algorithm by using a Canny edge detector to remove the wrong candidate points on the edges. Fan et al. [20] combined the spatial relationship of feature points to filter out unrepeatable feature points. Xie et al. [21] masked the complex regions to avoid extracting the interference points. Radiation-variation insensitive feature transform (RIFT) [22] proposed to detect corner and edge points on phase congruency (PC) maps that are robust to noise.

Forero et al. [23] and Sharma et al. [24] compared and analyzed various improved feature point detectors, many of which still rely on second-order derivatives as in SIFT or methods that take extremes in the neighborhood, which can make the feature point detectors sensitive to noise. Harris-Laplace detector [25] is more sensitive to corner points and has a degree of robustness to speckle noise [26]. Hence, it is widely used for feature point detection in remote sensing images. Chen et al. [27] employed a Harris detector to detect feature points on multimodal retinal images. Fan et al. [28] designed a uniform nonlinear diffusion-based Harris (UND-Harris) feature extraction method, which reduces the effect of speckle noise and obtains more uniformly distributed feature points. The SAR-SIFT method improved the traditional multiscale Harris detector by a new gradient calculation method, such that, as illustrated in Figure 1b, the detected feature points are mainly distributed on the corners of significant structures, effectively avoiding the influence of noise and texture details. However, the SAR-SIFT was originally proposed for SAR image registration. If the nonlinear intensity differences between optical and SAR images are not considered to construct distinctive feature descriptors, only a few feature points can be matched when the SAR-SIFT is directly applied to the optical-SAR image registration task.

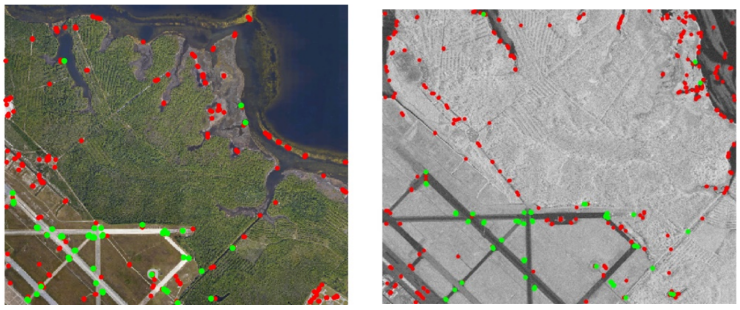
To avoid the effect of nonlinear intensity difference on the descriptors, Chen et al. [27] proposed a partial intensity invariant feature descriptor (PIIFD) that uses the averaging squared gradients in place of the traditional gradient to restrict the gradient direction within $[0, \pi)$ for solving the gradient reversal phenomenon. LNIFT [13] proposes a local normalization filter, which first transforms the original image into a normalized image, and then detects and describes feature points on the normalized image, in order to reduce the nonlinear intensity differences between multi-modal images. PSO-SIFT suggested calculating the gradient direction and gradient magnitude of the feature descriptors on the image boundaries. Shuai et al. [29] and Yu et al. [30] constructed descriptors by combining phase consistency [31–33] and gradient amplitude to address the gradient direction inconsistency in multimodal images. Fan et al. [28] built phase congruency structural descriptor (PCSD) on PC structure images which can get discriminable and robust structural edge descriptors.



(**a**) PSO-SIFT: mCMR = 66.84%



(**b**) SAR-SIFT: mCMR = 19.32%



(**c**) Ours: mCMR = 95.49%

Figure 1. Feature points extracted and matched in an optical image (**left**) and a SAR image (**right**) by (**a**) PSO-SIFT, (**b**) SAR-SIFT, and (**c**) our RTV-SIFT method. The points shown in the images are the detected feature points, with the green points indicating the correctly matched points. mCMR is the mean correct matching ratio (CMR, defined in Section 3).

Xie et al. [21] found that the unavailable features are mainly concentrated in complex regions, so they masked these regions to avoid extracting the interference points. We found that complex regions are texture-rich regions [34] such as mountainous areas, vegetation-covered areas, and dense urban areas. And the speckle noise in SAR images can also be regarded as a texture feature. The image inherent nature of optical and SAR images differ significantly in these texture regions, affecting the repeatability of feature points and descriptors' uniqueness. Smoothing the images can reduce the feature points extracted in the texture regions. However, many studies have shown that Gaussian smoothing destroys the natural edges of the image so that both details and noise are smoothed to the same extent, which reduces the localization accuracy of feature points and Description accuracy [16,28]. Therefore, how suppressing the texture while maintaining the sharpness of the edges of the image structure becomes the crux of the problem.

The relative total variation (RTV) method [34] has proven to be effective in smoothing texture while preserving the edges. Building upon this, we propose a novel RTV-Harris feature point detection method by combining the RTV smoothing and multiscale Harris detector. Our approach allows for more accurate extraction of feature points focusing on the structure edges, where the feature points have a higher matching potential. To account for the nonlinear differences in intensity between optical and SAR images, we propose an

enhanced phase congruency edge detector to construct the structural feature descriptor. Finally, a coarse-to-fine matching strategy is adopted to increase the correct matching rate. Experimental results demonstrate that our method yields higher feature point matching rates, as illustrated in Figure 1c. For ease of description, we name the proposed method as RTV-SIFT.

The contributions of this paper are as follows.

1. Based on the RTV theory and multi-scale Harris detector, an RTV-Harris feature point detector is proposed so that the detected feature points are distributed at the structural edges with higher matching potential.
2. To mitigate the effect of nonlinear intensity difference between optical and SAR images, an enhanced phase consistency edge (EPCE) descriptor is proposed for the structural feature description of the feature points.
3. A coarse-to-fine matching strategy based on feature point position and orientation Euclidean distance (POED) is introduced to improve the registration precision.

The paper is organized subsequently as follows. Section 2 describes the proposed registration method, including the RTV-Harris feature point detector, the EPCE feature descriptor, and the POED based matching method. The experimental setup, procedure, and result analysis are provided in Section 3. Section 5 gives the conclusion of this paper.

2. Proposed Method

The schematic diagram of the proposed RTV-SIFT method is shown in Figure 2 and the modules are described as follows.

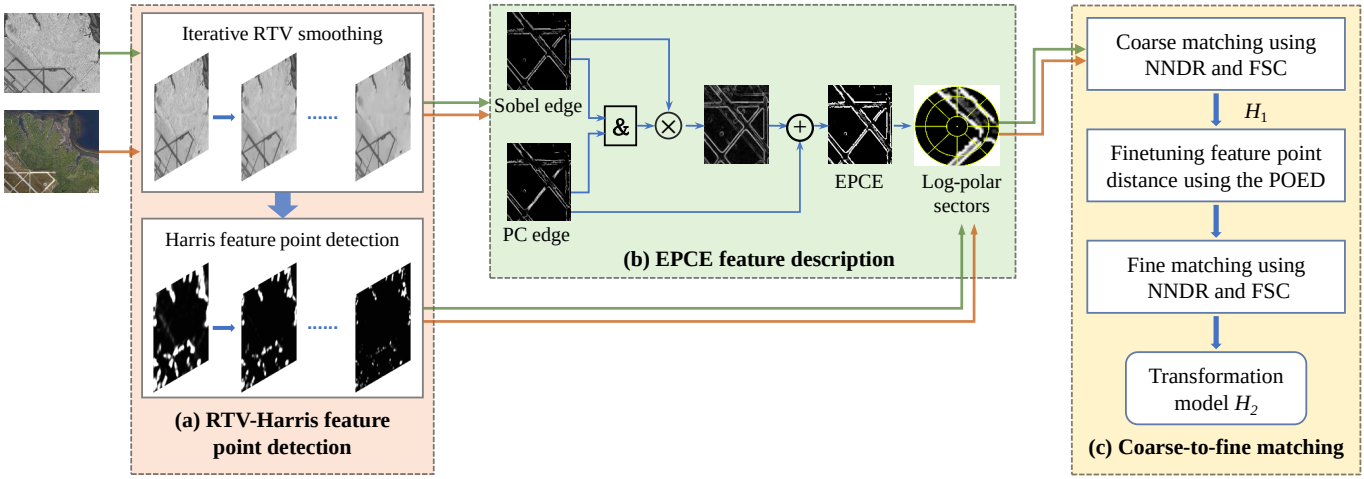


Figure 2. Diagram of the RTV-SIFT method. Abbreviations: RTV: Relative Total Variation; EPCE: Enhanced Phase Consistent Edge; NNDR: Nearest Neighbor Distance Ratio; FSC: Fast Sample consensus; POED: Position Orientation Euclidean Distance. "&", "x", "+" denote the "and", "multiply" and "add" operations for the corresponding pixel values, respectively.

2.1. Iterative Structure Preserving Smoothing Using RTV

As discussed in Section 1, it should be avoided to extract feature points for registration in the texture region. Therefore, the original optical and SAR images are first smoothed using the RTV method [34] to remove the texture while preserving the structural information in the images. This is achieved by optimizing the following objective function:

$$\arg\min_{S} \sum_{p} \left\{ (S(p) - I(p))^2 + \lambda \cdot \left(\frac{WTV_x(p)}{WIV_x(p)} + \frac{WTV_y(p)}{WIV_y(p)} \right) \right\}, \quad (1)$$

where I is the original optical or SAR image, S is the estimated structural image, $p = (x, y)$ is the pixel coordinate, and λ is a weight coefficient. WTV and WIV are the windowed total variation and windowed inherent variation measures, respectively, defined as

$$\begin{aligned} WTV_x(p) &= G_\sigma \cdot \left|\frac{\partial S(p)}{\partial x}\right|, \quad WTV_y(p) = G_\sigma \cdot \left|\frac{\partial S(p)}{\partial y}\right|, \\ WIV_x(p) &= \left|G_\sigma \cdot \frac{\partial S(p)}{\partial x}\right|, \quad WIV_y(p) = \left|G_\sigma \cdot \frac{\partial S(p)}{\partial y}\right|, \end{aligned} \tag{2}$$

where G_σ is a two-dimensional Gaussian kernel with variance of σ. Both the texture and structural regions will yield a large WTV response, while the WIV response is smaller in pure texture regions than in the structural regions.

The optimization problem in Equation (1) can be implemented iteratively [34], expressed as

$$v_S^{n+1} = (\mathbb{1} + \lambda L^n)^{-1} \cdot v_I, \tag{3}$$

where v_S and v_I are vector representations of S and I, respectively; L is the weight matrix computed based on the structural vector v_S, $\mathbb{1}$ is an identity matrix, and n is the iteration index.

For an optical or SAR image, a series of images $\{S^n, n = 1, \cdots, N\}$ with progressively suppressed texture information can be generated using the Equation (3), which is hereafter named the RTV iteration space. The choice of the maximum number of iterations N, referred to as the number of layers of the RTV iteration space, is discussed in Section 3.3.

2.2. Multiscale Feature Point Detection

The feature points of optical and SAR images are detected by Harris operators [25] with different scales in their RTV iteration spaces, respectively. Specifically, a gradient covariance matrix is first computed for each pixel in each structural image S^n in the RTV iteration space. The covariance matrix is then smoothed with a Gaussian convolution kernel with standard deviation $\sqrt{2}\sigma_n$ and multiplied by a scale factor σ_n^2. The resulting covariance matrix is written as

$$C(p) = \sigma_n^2 \cdot G_{\sqrt{2}\sigma_n} * \begin{bmatrix} \left(\frac{\partial S^n(p)}{\partial x}\right)^2 & \frac{\partial S^n(p)}{\partial x} \cdot \frac{\partial S^n(p)}{\partial y} \\ \frac{\partial S^n(p)}{\partial x} \cdot \frac{\partial S^n(p)}{\partial y} & \left(\frac{\partial S^n(p)}{\partial y}\right)^2 \end{bmatrix}, \tag{4}$$

where we set $\sigma_n = \sigma_0 \cdot 2^{-n/3}$ with σ_0 the initial Gaussian kernel variance.

Finally, the feature points in S^n are obtained by thresholding the Harris response score

$$R(p) = \det(C(p)) - d \cdot \mathrm{Tr}^2(C(p)), \tag{5}$$

where d is a constant sensitivity factor, "det" and "Tr" denote the determinant and trace of a matrix, respectively.

2.3. Feature Point Description

In order to alleviate the impact of nonlinear intensity differences between optical and SAR images on registration accuracy and to improve the discriminability of feature point representation, feature descriptors can be constructed by utilizing the image edge information. Existing methods usually use the Sobel operator (e.g., in [15]) or phase congruency (e.g., in [29]) to extract the image edges. However, the Sobel operator is sensitive to noise and prone to multi-pixel width, while the phase congruency maximum moment cannot reflect the contrast of edges.

In this study, we propose an enhanced phase consistent edge (EPCE) detector that combines the advantages of phase congruency and Sobel operator. More specifically,

for each structural image S^n in the RTV iteration space, its maximum moment of phase congruency edge map M^n and the Sobel edge map H^n are first computed. Figure 3 shows the Sobel and phase congruency edge maps of an example optical image. Due to the sensitivity of the Sobel operator to image details, the Sobel edge map looks quite noisy. To remove the noisy edge fragments from the Sobel edge map, H^n is filtered by its intersection with the phase congruency edge map M^n, which is formulated as

$$\mathbb{H}^n(p) = (M^n(p) \& H^n(p)) \cdot H^n(p), \tag{6}$$

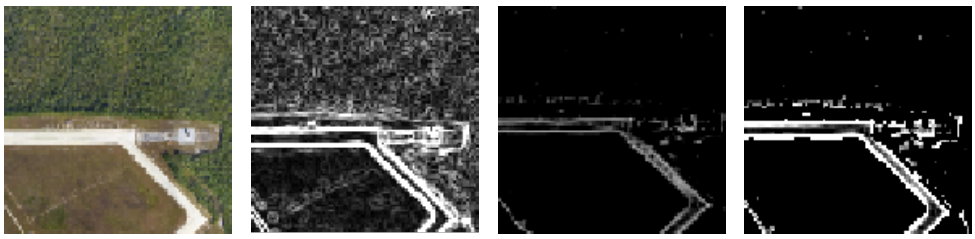
where the symbol "&" denotes the pixel-wise "AND" operation. The results obtained are logical 0 or 1. Finally, the filtered Sobel edge map is summed with the phase congruency edge map M^n to obtain an enhanced EPCE map M^n_{en}, written as

$$M^n_{en}(p) = \mathbb{H}^n(p) + M^n(p). \tag{7}$$

Figure 3d illustrates that the EPCE detector is effective in extracting image contours and filtering out the noisy edge fragments. The gradient amplitude and orientation of the EPCE map M^n_{en} are calculated as

$$\begin{aligned} G^n(p) &= \sqrt{(\frac{\partial M^n_{en}}{\partial x})^2 + (\frac{\partial M^n_{en}}{\partial y})^2}, \\ R^n(p) &= \arctan(\frac{\partial M^n_{en}}{\partial y} / \frac{\partial M^n_{en}}{\partial x}). \end{aligned} \tag{8}$$

The domain orientations of feature points and the construction of descriptors are computed using histograms of gradient orientation, similar to SIFT. Instead of using a square neighborhood for feature points in SIFT, we use log-polar sectors neighborhoods to compute the descriptor as in PSO-SIFT, which generates 17 bins as the Log-polar sectors in Figure 2b. And the gradient orientations are quantized in eight bins, so that EPCE yields a 136-dimensional feature descriptor.



(**a**) (**b**) (**c**) (**d**)

Figure 3. Example edge maps of (**a**) an optical image extracted by (**b**) the Sobel operator, (**c**) the phase congruency, and (**d**) the proposed enhanced phase congruency edge detector.

2.4. Coarse-to-Fine Feature Point Matching

In the feature point matching process, we first employ the Nearest Neighbor Distance Ratio (NNDR) [17] to coarsely match the feature points of the optical and SAR images, and use the Fast Sample Consensus (FSC) algorithm [35] to reject incorrect corresponding points and obtain the coarse affine transformation model H_1 between the optical and SAR images.

To further improve the matching accuracy, Ma et al. [15] employed the position, scale, and main orientation information of the feature points to recalculate the Euclidean distance between the feature points, which is named PSOED and expressed as

$$\begin{aligned} PSOED(p_i, p_j) &= err(p_i, p_j) ED(p_i, p_j), \\ err(p_i, p_j) &= (1 + e_p(p_i, p_j))(1 + e_s(p_i, p_j))(1 + e_o(p_i, p_j)), \end{aligned} \tag{9}$$

where p_i and p_j are the feature points on the reference image and sensed image, respectively; $ED(p_i, p_j)$ represents the Euclidean distance between the feature descriptors of p_i and p_j, $e_p(p_i, p_j)$, $e_s(p_i, p_j)$ and $e_o(p_i, p_j)$ respectively denote the position error, scale error and main orientation error between p_i and p_j, which are defined as

$$\begin{aligned} e_p(p_i, p_j) &= ||p_i - H_1(p_j)||_2, \\ e_s(p_i, p_j) &= |1 - r^* \cdot \frac{s_j}{s_i}|, \\ e_o(p_i, p_j) &= |\Delta\theta_{i,j} - \Delta\theta^*|, \end{aligned} \tag{10}$$

where H_1 denotes the coarse transformation model, $||\cdot||_2$ represents the Euclidean distance calculation, s_i and s_j are the scales of p_i and p_j respectively, $\Delta\theta_{i,j}$ is the main orientation difference between p_i and p_j, r^* and $\Delta\theta^*$ are the statistical maxima of the scale ratio and the major orientation differences between the matched pairs obtained from the initial matching, respectively.

Due to the different resolution and edge diffusion effects of different scale images in Gaussian scale space, it is difficult to correctly match the feature points with large-scale disparity. Whereas in the RTV-Harris scale space, the images have consistent resolution and clear edges, thus minimizing the impact of scale difference on our method. To reduce the computational burden, we discard the scale error term in the PSOED method and refer to the new Euclidean distance function as POED, expressed as

$$POED(p_i, p_j) = (1 + e_p(p_i, p_j))(1 + e_o(p_i, p_j))ED(p_i, p_j). \tag{11}$$

After coarse registration and obtaining the coarse transform matrix H_1, the Euclidean distances between descriptors of the feature points are fine-tuned using Equation (11). Then NNDR is used again to match the feature points with the fine-tuned distance between descriptors and the FSC algorithm is used to obtain a more accurate transformation matrix H_2.

3. Experiments and Results

In this section, we evaluate the performance of the proposed RTV-SIFT and compare it with several state-of-the-art registration algorithms. The hardware device for our experiments is a computer equipped with AMD Ryzen 9 3900XT CPU and 32GB memory, and the software platform is MATLAB R2019a.

3.1. Evaluation Metrics

The metrics for evaluating the registration method are as follows

Repeatability rate. repeatability rate [36] is a criterion for evaluating the stability of the feature point detector in detecting the homonymous point on two modal images. Let $\tilde{x_1}$ and $\tilde{x_i}$ be points that lie on the common part of the images of I_1 and I_i, and they are defined as

$$\{\tilde{x_1}\} = \{x_1 | H_{1i}x_1 \in I_i\}, \{\tilde{x_i}\} = \{x_i | H_{i1}x_i \in I_1\}. \tag{12}$$

where H_{ij} denotes the homography between images I_i and I_j. Then, a neighborhood size ϵ is determined, and the ϵ-repetition rate is defined as

$$R_i(\epsilon) = \{(\tilde{x}_1, \tilde{x}_i) | dist(H_{1i}\tilde{x}_1, \tilde{x}_i) < \epsilon\}. \tag{13}$$

where $dist()$ denotes The Euclidean distance between two points. Thus, the repeatability rate for image I_i is defined as

$$r_i(\epsilon) = \frac{|R_i(\epsilon)|}{min(n_1, n_i)}. \tag{14}$$

where $n_1 = |\{\tilde{x_1}\}|$ and $n_i = |\{\tilde{x_i}\}|$ are the number of points detected in the common part of images I_1 and I_i respectively.

Correct Matching Number (CMN). CMNis an important indicator of the robustness of registration algorithms. In our experiments, we manually selected more than 20 pairs of uniformly distributed and reliable corresponding points in each pair of the test images to estimate a standard transformation matrix H_o. After applying the H_o transformation, matches with position errors of less than 3 pixels are considered correct matches.

Correct Matching Ratio (CMR). CMR is the ratio of the CMN to the total number of matches N_m, written as

$$\mathrm{CMR} = \frac{\mathrm{CMN}}{\mathrm{N_m}}. \tag{15}$$

Root Mean Square Error (RMSE). RMSE is an important indicator of registration accuracy and the formula is written as

$$\mathrm{RMSE} = \left(\frac{1}{\mathrm{N_m}} \sum_{\mathrm{i}=1}^{\mathrm{N_m}} ||(\mathrm{x_r^i}, \mathrm{y_r^i}) - \mathrm{H_o}(\mathrm{x_s^i}, \mathrm{y_s^i})||_2 \right)^{1/2}, \tag{16}$$

where (x_r^i, y_r^i) and (x_s^i, y_s^i) are the coordinates of the ith matched point pair on the reference image and the sensed image, respectively.

Distribution of the matched points (S_{cat}). Goncalves et al. [37] points out that the correct corresponding points should be distributed over the whole image as much as possible, and proposes a quantitative metric S_{cat} to measure this distribution. First, given that $x = \{x_1, x_2, \cdots, x_N\}$ denotes the correct corresponding feature points in the reference image. Then, the Euclidean distances between every feature point and all of the feature points are calculated as

$$D_i = \{dist(x_i, x_1), dist(x_i, x_2), \cdots, dist(x_i, x_N)\}. \tag{17}$$

where $i \in \{1, 2, 3, \cdots, N\}$. Finally, S_{cat} is calculated as

$$S_{cat} = \frac{\sum_{i=1}^{N} med(D_i)}{N}. \tag{18}$$

where $med(D_i)$ denotes the medians of D_i. On the same order of magnitude, the more uniform the distribution of feature points is, the larger the S_{cat} value is. Note that the coordinates of feature points used in the calculation of S_{cat} are normalized.

Time(s). The time required for two images to be successfully matched from input to output was calculated.

3.2. Test Images

This paper uses two sets of test images. The first one is a subset of the OS (Optical-SAR) dataset [38]. Since this dataset contains aligned optical and SAR Images, we use it to verify the repeatability of the proposed RTV-Harris feature point detector. We randomly selected 8 pairs of aligned optical and SAR images of different scenes (size 512X512, resolution 1 m) from the OS dataset [38], as shown in Figure 4.

Another test dataset contains 11 pairs of unaligned large-scene optical and SAR images. The details of the images are shown in Table 1. These test images are used to evaluate the performance of the overall registration algorithm, shown in Figure 5.

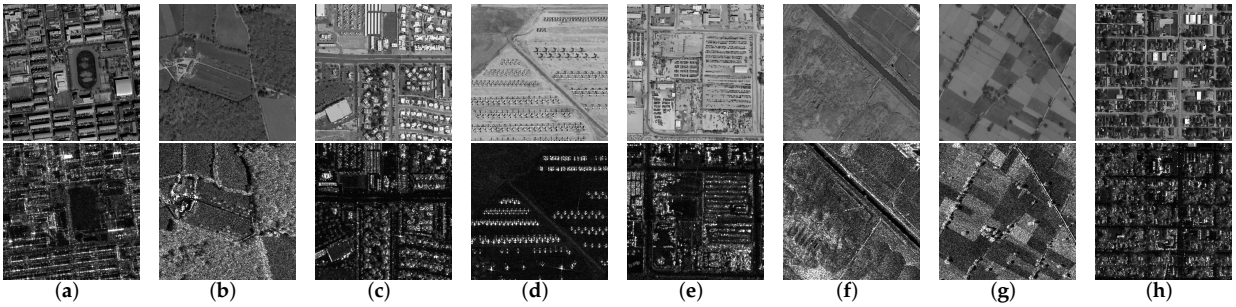


Figure 4. Randomly selected 8 pairs of aligned optical (top) and SAR (bottom) images from the OS dataset [38]. where, (**a**,**c**,**h**) is urban area, (**b**,**f**,**g**) is farmland area, (**e**) is suburban area, and (**d**) is airport area.

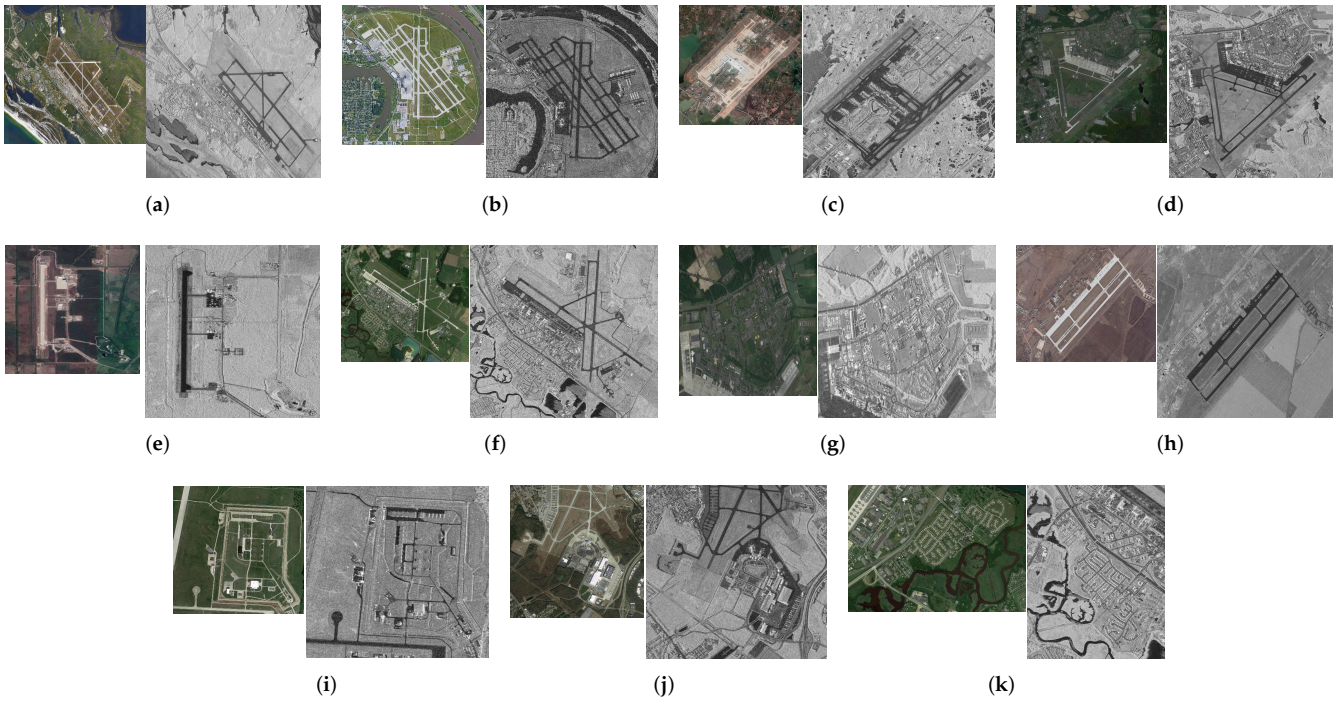


Figure 5. 11 pairs of unaligned large-scene optical and SAR images. (**a**–**k**) are image pairs 1–11, respectively.

3.3. Parameter Settings

The variance σ of the Gaussian kernel in Equation (2) plays a crucial role in feature point detection. We set σ to 2 pixels for the optical images and 4 pixels for the SAR images in the experiments. It can be fine-tuned according to the texture size.

Under the guidance of Xu et al. [34], the parameter λ in Equation (3) is set to 0.004. In the traditional multi-scale Harris algorithm, as the Gaussian smoothing scale increases, the image edges gradually spread, so a gradually increasing scale window is needed to detect feature points. On the contrary, in the RTV-Harris scale space, the size of the spatial scale window σ_n is gradually decreasing in order to preserve the sharpness of the edges. We set the initial scale σ_0 of the RTV-Harris detector to 6 pixels and the ratio of adjacent scales to $2^{-\frac{1}{3}}$. The constant d in the Harris response score function is set to 0.04. The score function threshold for feature point detection is set to 0.1. The ratio threshold of NNDR is set to 0.9.

Table 1. Information about the test images.

	Type	Image Pairs										
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Size	SAR	1025 × 800	600 × 675	735 × 768	788 × 888	350 × 675	975 × 925	875 × 552	975 × 800	501 × 494	900 × 898	705 × 878
	Optical	832 × 640	496 × 544	468 × 528	720 × 704	224 × 416	656 × 624	677 × 482	864 × 608	356 × 366	665 × 689	919 × 643
Resolution	SAR	1 m	1 m	1 m	1 m	1 m	1 m	1 m	1 m	1 m	1 m	1 m
	Optical	2 m	2 m	2 m	2 m	2 m	2 m	2 m	2 m	2 m	2 m	2 m
Source	SAR	GF-3	GF-3	GF-3	GF-3	GF-3	GF-3	GF-3	GF-3	GF-3	GF-3	GF-3
	Optical	Google Earth	Google Earth	Google Earth	Google Earth	Google Earth	Google Earth	Google Earth	Google Earth	Google Earth	Google Earth	Google Earth
Scene		Airport	Airport	Airport	Airport	Airport	Airport	Dense urban area	Airport	Airport	Airport	Suburb with large rotation angle

Note: GF-3 is the Chinese satellite Gaofen-3.

In addition to the above parameters, the number of layers N of the RTV iteration space, a key parameter of the RTV-SIFT method, was determined experimentally. We evaluated the average CMN and the average consumption time of the RTV-SIFT method on the test images with N ranging from 1 to 11, while other parameters were kept the same as above. The results in Figure 6 show that the average CMN does not always increase with the number of layers. When N is greater than 8, the growth of average CMN is no longer significant and even starts to decrease after 10 layers. However, the average consumption time of the algorithm keeps increasing with the number of layers. The results suggest that increasing the number of layers in the RTV iteration space does not necessarily lead to better registration performance. As the number of feature points increases, more redundant feature points are involved in matching, which ultimately leads to a decrease in registration performance. Therefore, to strike a balance between computational efficiency and algorithmic robustness, we advocate setting N = 8 for the RTV iteration space in all the following experiments.

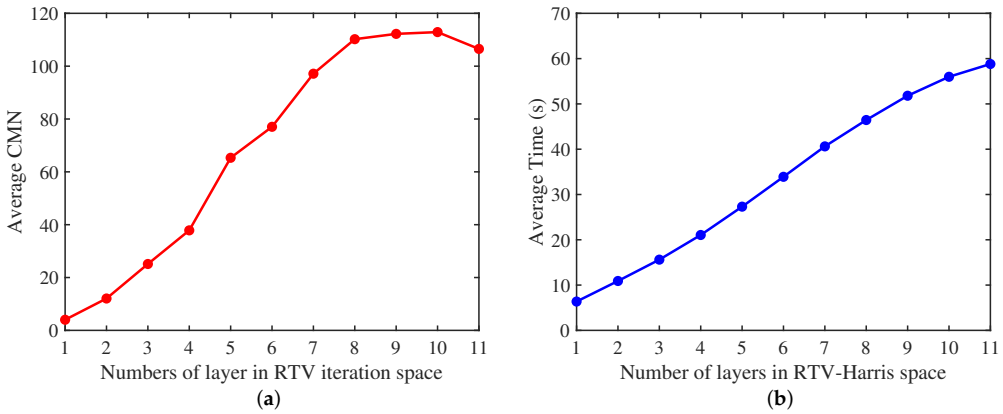


Figure 6. The average CMN (**a**) and the consumption time (**b**) of the RTV-SIFT method with the number of layers N in the RTV iteration space varying from 1 to 11.

3.4. Performance of RTV-Harris Feature Point Detector

The repeatability rate [36] with different thresholds was calculated for the 8 pairs of aligned optical and SAR images from the OS (Optical-SAR) dataset. The comparison results with the DoG method [17], the SAR-Harris method in SAR-SIFT [16], and the modified multi-scale Harris (multiscale-Harris with refinement) method in OS-SIFT are shown in Figure 7.

Among the four methods, the RTV-Harris detector achieves the highest repeatability rate of feature points for different senses and localization errors, as shown in Figure 7. In particular, for scenes with more small targets, such as the image pair (d), the advantage of the RTV-Harris detector is more pronounced. When the localization error is 4 pixels, the feature point repeatability rate has exceeded 50%. The RTV-Harris detector performs best for several reasons. First, the RTV method is able to attenuate texture and eliminate speckle noise, which prevents extracting invalid feature points to a certain extent. Second, the detection window of RTV-Harris gradually narrows, which improves the localization accuracy of feature points. These two factors contribute to the low false alarm rate and high repeatability rate.

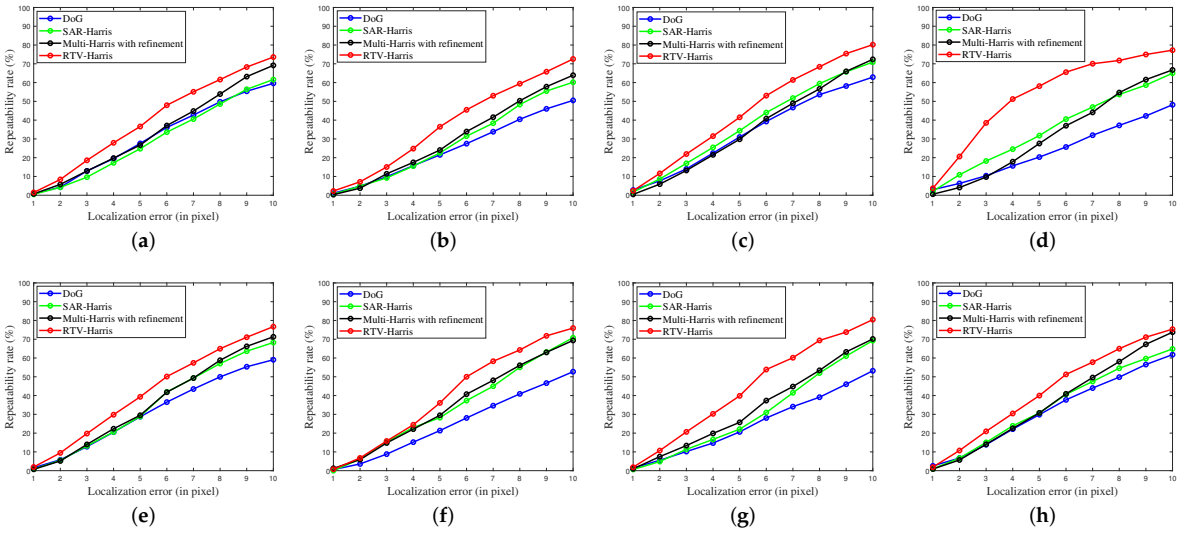


Figure 7. (**a**–**h**) are the feature points repeatability rate of the images pairs (**a**–**h**) in Figure 4, respectively. The independent variable of the line graph is the localization error.

3.5. Performance of EPCE Feature Descriptor

The performance of the EPCE feature descriptor was evaluated on the 11 large-scene optical and SAR image pairs by comparing it with descriptors constructed on the gradient of the Sobel edge and PC edge. For each group of experiments, the RTV-Harris detector with the same parameters was used to extract the feature points, and the matching method is the proposed coarse-to-fine method as described in Section 2.4. On average, the CMN of the EPCE descriptor is 32.55 ± 23.86 higher than that of the Sobel edge descriptor and 34.36 ± 22.81 higher than that of the PC edge descriptor. The resulting CMN metrics are shown in Figure 8. It shows that the EPCE descriptor provides a more precise characterization of feature points.

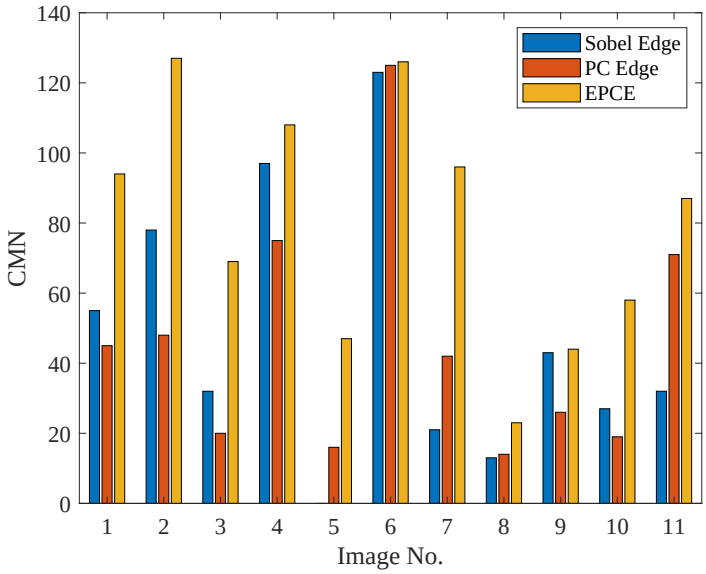


Figure 8. CMN of feature descriptors based on different edge detectors.

3.6. Overall Registration Performance of RTV-SIFT

In this subsection, we compare the effectiveness of the proposed RTV-SIFT method with the SAR-SIFT [16], Harris-PIIFD [27], and the other four state-of-the-art feature point-based methods, namely LNIFT [13], PC-SIFT and its enhanced version PCG-SIFT [29],

PSO-SIFT [15] and OS-SIFT [39]. To ensure the validity of the experimental results, we followed the parameter settings in the original method.

The comparison results are presented in Table 2. It can be inferred that the SAR-SIFT, LNIFT, PC-SIFT, and PCG-SIFT methods often fail to achieve image registration for optical and SAR images with significant nonlinear intensity differences. Although some methods perform better than others, they lack robustness in different scenarios. For example, OS-SIFT and PSO-SIFT fail to register in dense urban areas (image pair 7), while Harris-PIIFD and OS-SIFT fail to register in suburban areas with large rotation angles and high vegetation coverage (image pair 11). In addition, the existing methods detect fewer and more concentrated feature points, which leads to larger registration errors in the RMSE values. In contrast, the feature points detected by our RTV-SIFT are mainly distributed at the edges and thus have higher matching potential, resulting in higher CMN and CMR. Furthermore, since the detected feature points are more evenly distributed over the whole image, the registration error (RMSE) is smaller. Hence it is more robust in various situations. The last two rows of Table 2 also give the results of RTV-SIFT without/with the coarse-to-fine feature point matching strategy POED. The results indicate that RTV-SIFT achieves superior registration performance compared to existing algorithms only when coarse alignment is applied. The introduction of the fine-matching POED strategy leads to a significant improvement in the registration performance of RTV-SIFT. In addition, we also demonstrate the registration performance with N = 5 in Table 2. It can be seen that when N = 5, better results than the SOTA method can be achieved, and the average time consumption is only 70% of that with N = 8.

The qualitative comparison of the better performing Harris-PIIFD, OS-SIFT, PSO-SIFT, and our RTV-SIFT methods on the airport area, suburban area, and vegetation-covered area, respectively, are shown in Figure 9. It is visible that the correct matching points obtained from the OS-SIFT and PSO-SIFT algorithms are mainly concentrated on the roads with clear features. In contrast, our RTV-Harris detector produces a wider distribution of matching points. The concentrated distribution of matching points will increase the registration error in areas without matching points, while a uniform distribution of matching points can achieve more accurate alignment over the whole image range, as illustrated in Figure 10.

Finally, we also report the time spent on each step of RTV-SIFT to give a reference for subsequent studies, as shown in Table 3. Note that the results are the average time consumed for all the test image pairs.

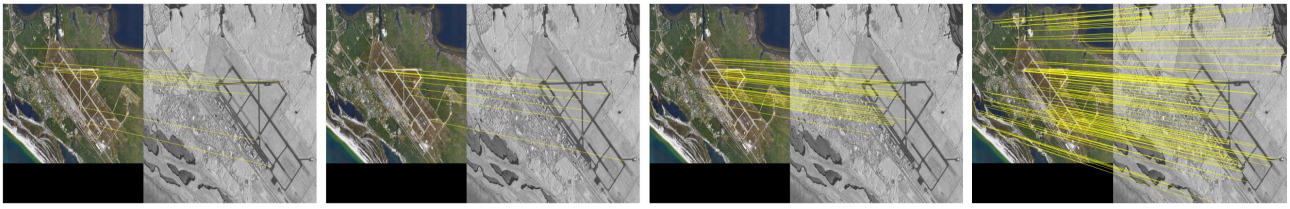
Table 2. Performance comparison of different registration methods.

Method	Criterion	Image Pairs 1	2	3	4	5	6	7	8	9	10	11
SAR-SIFT	CMN	1	0	1	5	0	2	0	0	0	0	24
	CMR(%)	20.00	0	33.33	71.43	0	18.18	0	0	0	0	88.89
	RMSE	10.68	130.21	322.64	2.36	71.54	3.50	486.83	477.25	165.60	386.00	1.72
	S_{cat}	-	-	-	0.021	-	-	-	-	-	-	0.175
	Time(s)	14.70	10.25	10.81	17.22	1.06	122.65	5.72	5.82	1.86	21.36	28.52
PC-SIFT	CMN	1	16	1	0	0	0	0	4	1	0	1
	CMR(%)	33.33	100	33.33	0	0	0	0	80.00	25.00	0	33.33
	RMSE	240.54	1.16	108.66	324.02	31.41	289.85	342.71	1.92	42.07	357.27	284.56
	S_{cat}	-	0.172	-	-	-	-	-	0.240	-	-	-
	Time(s)	10.15	3.97	6.13	8.65	0.61	8.64	5.36	4.84	0.77	12.69	14.70
PCG-SIFT	CMN	7	27	0	7	0	8	0	7	1	1	8
	CMR(%)	77.78	100	0	87.50	0	72.73	0	63.64	25.00	25.00	88.89
	RMSE	4.56	1.18	181.67	5.20	213.82	2.61	295.51	2.69	56.42	36.66	4.65
	S_{cat}	0.094	0.182	-	0.197	-	0.187	-	0.167	-	-	0.067
	Time(s)	9.28	4.11	6.03	7.90	0.6	8.15	5.25	4.86	0.79	11.89	17.81
Harris-PIIFD	CMN	8	23	13	9	5	8	3	8	8	12	0
	CMR(%)	80.00	**100**	**100**	64.29	83.33	57.14	27.27	66.67	80.00	68.42	0
	RMSE	2.41	**0.71**	1.39	3.82	1.94	3.77	5.33	4.62	2.77	3.26	625.37
	S_{cat}	0.146	0.206	0.208	0.368	0.337	0.272	0.298	0.321	0.345	0.266	-
	Time(s)	1.48	1.14	1.26	1.32	1.06	1.41	1.26	1.39	1.27	2.23	1.35
LNIFT	CMN	0	0	0	0	0	0	19	19	18	15	70
	CMR(%)	0	0	0	0	0	0	37.25	44.19	40.00	39.47	86.42
	RMSE	235.57	206.51	239.47	517.31	106.19	285.63	3.41	3.67	3.51	6.90	2.19
	S_{cat}	-	-	-	-	-	-	0.284	0.253	0.251	0.319	0.296
	Time(s)	44.27	32.82	36.26	50.77	20.04	45.08	34.58	50.98	24.00	49.93	40.49

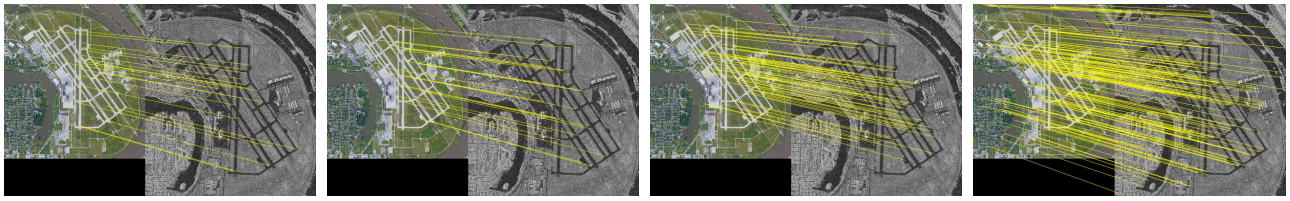
Table 2. *Cont.*

Method	Criterion	Image Pairs 1	2	3	4	5	6	7	8	9	10	11
OS-SIFT	CMN	11	25	5	15	3	15	1	4	5	11	1
	CMR(%)	**100**	**100**	83.33	88.24	75.00	88.24	20.00	80.00	62.50	73.33	33.33
	RMSE	1.37	1.08	2.05	1.99	28.33	1.89	5.44	4.01	3.80	2.67	126.29
	S_{cat}	0.109	0.201	0.276	0.265	0.284	0.243	-	0.253	0.366	0.182	-
	Time(s)	18.65	10.83	12.37	18.36	1.04	23.53	6.97	11.36	2.17	22.31	27.31
PSO-SIFT	CMN	33	63	4	10	9	27	0	2	17	6	14
	CMR(%)	**100**	96.92	26.67	71.43	90.00	80.95	0	28.57	**100**	66.67	46.67
	RMSE	1.26	1.38	5.09	5.13	1.67	2.76	61.36	6.71	**1.80**	2.55	5.20
	S_{cat}	0.087	0.174	0.109	0.168	0.364	0.174	-	-	0.228	0.053	0.246
	Time(s)	41.00	9.91	22.44	27.48	0.48	40.21	11.24	6.55	0.88	70.73	99.19
RTV-SIFT with POED (N = 5)	CMN	34	53	26	35	7	86	30	10	18	32	36
	CMR(%)	**100**	**100**	86.67	**100**	70	94.51	93.75	58.82	85.71	84.21	81.81
	RMSE	1.18	1.52	2.09	1.68	2.39	1.78	2.64	3.28	2.33	1.54	2.27
	S_{cat}	0.197	0.226	0.276	0.439	0.318	0.283	0.313	0.129	0.333	0.267	0.241
	Time(s)	23.96	10.86	14.52	21.72	1.79	38.34	22.84	22.29	2.29	22.32	26.41
RTV-SIFT without POED (N = 8)	CMN	46	51	23	31	8	78	10	12	21	17	55
	CMR(%)	97.87	**100**	**100**	**100**	**100**	89.66	76.92	**100**	91.3	**100**	**100**
	RMSE	1.04	1.11	1.02	0.85	**0.72**	1.95	2.86	**1.85**	2.03	1.75	**1.01**
	S_{cat}	0.161	0.195	0.230	0.291	0.355	0.257	0.247	0.165	0.277	0.266	0.206
	Time(s)	31.08	15.75	18.8	29.84	2.22	46.39	26.07	28.99	3.10	29.42	44.94
RTV-SIFT with POED (N = 8)	CMN	**94**	**127**	**69**	**108**	**47**	**126**	**96**	**23**	**44**	**58**	**87**
	CMR(%)	**100**	**100**	**100**	**100**	**100**	**100**	**100**	54.76	95.65	**100**	**100**
	RMSE	**0.64**	0.75	**0.80**	**0.68**	2.27	**0.75**	**0.69**	3.89	1.91	**0.77**	3.12
	S_{cat}	0.258	0.272	0.30	0.416	0.35	0.274	0.266	0.237	0.296	0.312	0.328
	Time(s)	32.43	17.29	20.35	31.59	2.41	49.45	28.81	28.11	3.10	30.59	48.62

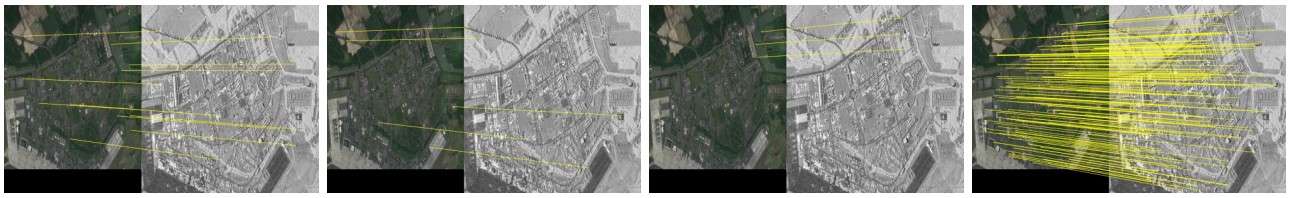
Note: The values of CMR are shown in percentage. Bolded fonts are the best performers in this metric.



(**a**)



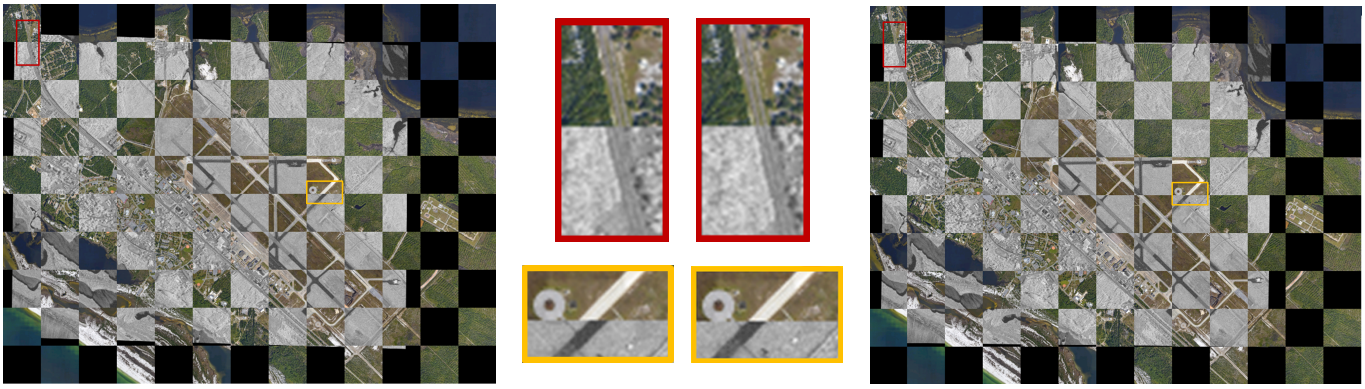
(**b**)



(**c**)

(**d**)

Figure 9. Qualitative comparison of different registration methods, from left to right, Harris-PIIFD, OS-SIFT, PSO-SIFT, and RTV-SIFT (ours). (**a**–**d**) are the results on test image pair 1, 2, 7, and 11, respectively.



(**a**) Result of PSO-SIFT (**b**) Zoomed-in view of patches (**c**) Result of RTV-SIFT

Figure 10. Qualitative comparison of the registration results of PSO-SIFT and RTV-SIFT (ours) on test image pair 1. The registration results are shown as checkerboard images. The left patches of the zoomed-in views in (**b**) are from the result of PSO-SIFT, and the right patches are from that of RTV-SIFT.

Table 3. Average consumption time on each step of RTV-SIFT.

RTV-Harris Space Construction	Keypoints Detection	EPCE Feature Description	Corse-to-Fine Match	Final Time
5.82 s	0.38 s	18.39 s	2.02 s	26.61 s

3.7. Validation under Different Conditions

Except for the interference from speckle noise, SAR images are unaffected by the weather conditions and time of day they are taken. However, optical images are susceptible to various factors such as illumination intensity and weather conditions. In this subsection, we employ simulated images under different imaging conditions to test the robustness of the proposed RTV-SIFT registration method. The simulated images were generated by processing real optical images with varying illumination, different levels of noise interference, and cloud occlusion, as depicted in Figure 11. Since the number and distribution of matching points significantly influence the registration effect, we utilize CMN and S_{cat} to evaluate the algorithm's performance.

3.7.1. Illumination Intensity

Variations in illumination intensity are the most common weather conditions that affect the contrast of optical images, thereby interfering with the registration results. We simulated optical images of different illumination intensities by attenuating and enhancing the intensity of the image to varying degrees. In order to vary the illumination intensity uniformly from weak to strong, we set 11 illumination levels as $[0.5, 0.6, 0.7, 0.8, 0.9, 1, \frac{1}{0.9}, \frac{1}{0.8}, \frac{1}{0.7}, \frac{1}{0.6}, \frac{1}{0.5}]$.

Figure 12a,d show the average CMN and average S_{cat} values of the 11 pairs of test images. The proposed RTV-SIFT performs better and is more stable than the other two competing algorithms in all cases at different illumination levels. This is attributed to the fact that the EPCE descriptor inherits the PC's insensitivity to light intensity and contrast [40].

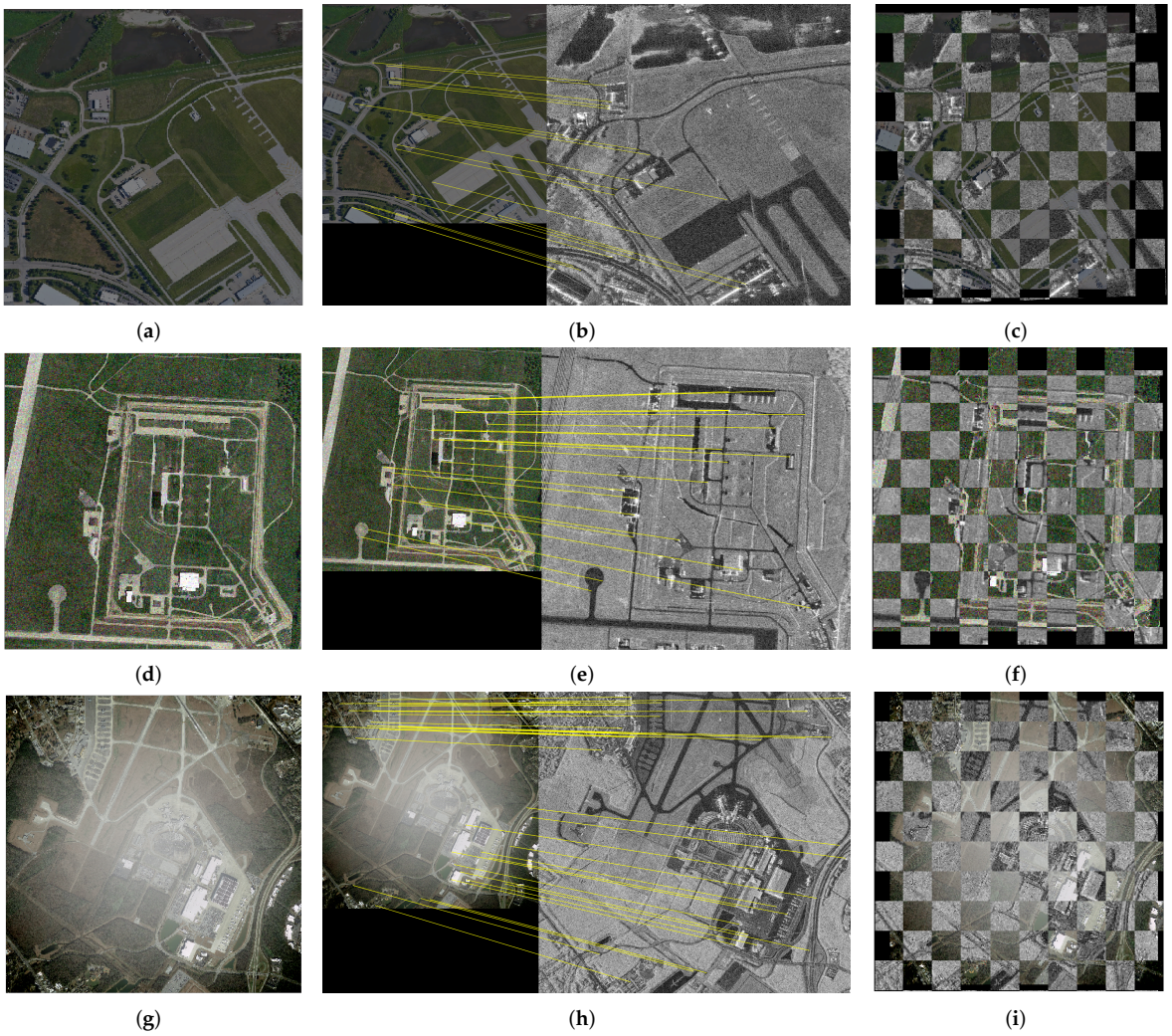


Figure 11. Example images of simulated optical images under different imaging condition and their registration results by RTV-SIFT method. (**a**,**d**,**g**) are the simulated optical images with illumination variation, noise, and cloud covering, respectively. (**b**,**e**,**h**) show the corresponding points in the image pairs. (**c**,**f**,**i**) show the checkerboard images of the registration results.

3.7.2. Noise Interference

Since the space optical cameras mainly use linear array CCDs (Charge-coupled Devices) as the sensor and the readout noise of CCD cameras follows Gaussian distribution [41], we simulated noise interference to optical images by adding different levels of Gaussian noise. The mean values of Gaussian noise were set to 0 and the variance was divided into 11 levels sampled at equal intervals from 0 to 0.1.

As shown in Figure 12b,e, the RTV-SIFT is able to maintain good performance all the time with different levels of noise disturbances, with almost no fluctuations in the average CMN and average S_{cat} values. This proves that the RTV-SIFT is robust to noise.

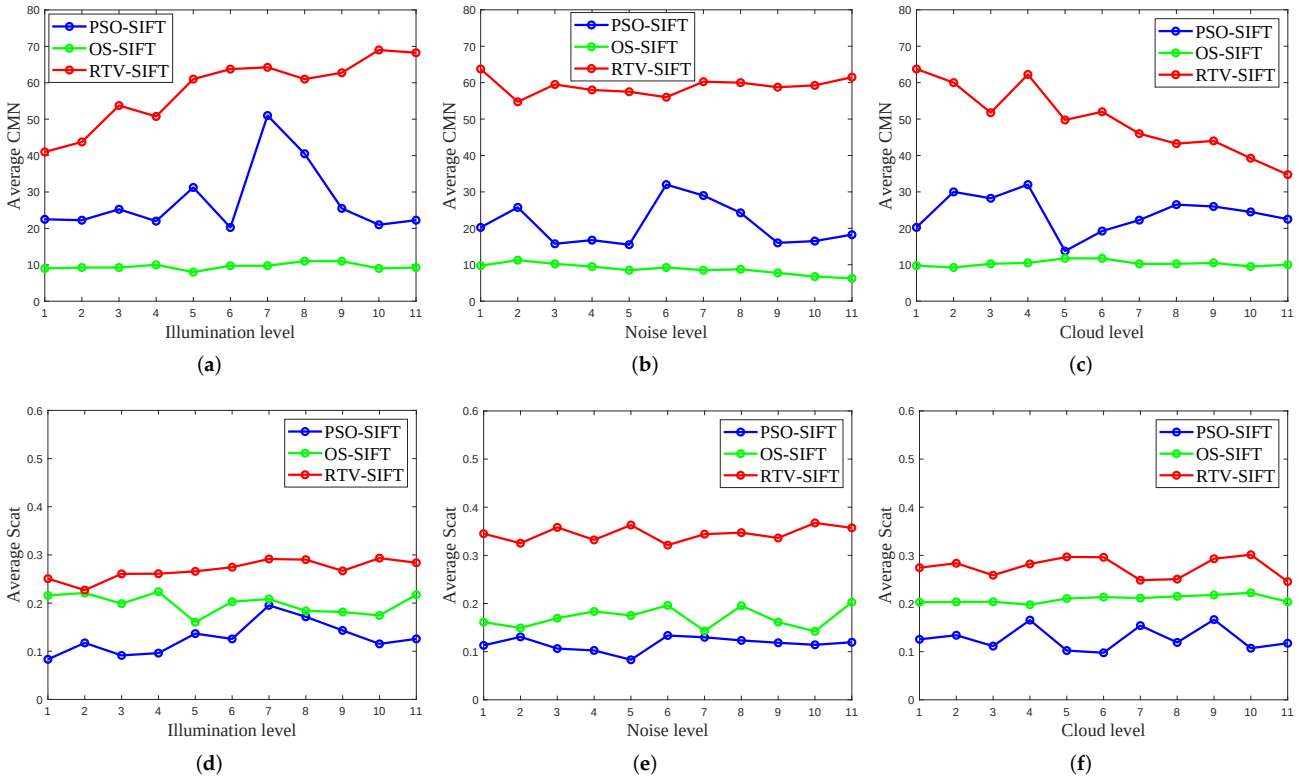


Figure 12. Average CMN (top row) and average S_{cat} (bottom row) of the PSO-SIFT, OS-SIFT, and RTV-SIFT (ours) on the test images under different simulated optical imaging conditions. (**a**,**d**) Results of simulated illumination variation. (**b**,**e**) Results of simulated noise interference. (**c**,**f**) Results of simulated cloud obscuration.

3.7.3. Cloud Covering

Optical remote sensing images can be negatively impacted by cloud occlusion during capture [42], resulting in degraded image quality and affecting feature extraction. A good registration algorithm should be robust to cloud occlusion to ensure optimal performance. To evaluate the impact of cloud occlusion on registration results, we used a mask with controlled transparency and extent to simulate clouds in optical images, with the cloud center set at the image center. 11 levels were set to represent varying cloud thicknesses and influence ranges.

The results in Figure 12c,f show that the average CMN and average S_{cat} of RTV-SIFT decrease with the increase of cloud thickness and occlusion area, but its performance is still far superior to the other two algorithms.

3.8. Summary of Experimental Results

Based on the above experimental results, we can draw the following conclusions.

1. Our proposed RTV-Harris feature point detector is robust to speckle noise and texture, so the extracted feature points are mainly distributed at the edges of the structure with a higher repeatability rate than the traditional DoG and multiscale-Harris approaches.
2. The EPCE feature descriptor can effectively overcome the nonlinear intensity differences between optical and SAR images, and is more accurate than the descriptors constructed on the Sobel and PC edges.
3. The POED based fine matching method can effectively increase the number of correct corresponding points and make their distribution more uniform, as shown in the last two rows of Table 2.
4. The RTV-SIFT method outperforms other algorithms in various scenes and imaging conditions, showcasing its superior robustness and adaptability.

4. Discussion

Despite significant radiometric differences between optical and SAR images, their structural information is consistent and stable. Our proposed method leverages this structural information to effectively address the two problems outlined in Section 1. Firstly, we use the RTV method to smooth speckle noise and texture regions while preserving edge accuracy. By detecting feature points on the structure image, we significantly increase the likelihood of detecting the homology point on both images. The experiments in Section 3.4 suggest this property. Next, we construct EPCE descriptors on the structure image to assign each feature point an accurate, unique descriptor. The structural edge descriptors of homonymous points are stable and similar, leading to a significant improvement in the correct matching rate of feature points as shown in Table 2.

In summary, the experiments conducted in this paper provide valuable insights into the registration of optical and SAR images and demonstrate the effectiveness of the proposed methods. The findings in this study have important implications for future research in the field of remote sensing image applications. however, since our method relies on the structural features of the images, it may not perform well in weakly structured regions, as depicted in Figure 13. These areas are localized with very small fields of view (256 m $\times$ 256 m), thus lacking significant structural features, leading to registration failure. Nevertheless, in practical engineering applications, we can still harness structural information to register large scene images, thereby achieving registration of local regions with weakly structural features.

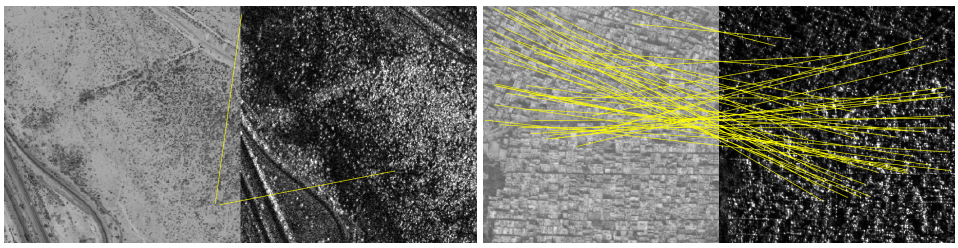


Figure 13. Registration effects in local weakly structured regions.

5. Conclusions

In conclusion, we have proposed a novel algorithm for registering optical and SAR images based on structure extraction and structural edge descriptors. Our approach utilizes the RTV-Harris detector to extract feature points located mainly on the structural edges, resulting in a high repeatability rate. The EPCE feature descriptors constructed on the RTV iteration space effectively overcome the intensity nonlinear difference between the optical and SAR images, obtaining accurate descriptors. Furthermore, the POED-based fine-matching method combines the position and principal direction information of feature points, resulting in more precise correspondences and improved registration accuracy. In comparison to several state-of-the-art methods, we found that the proposed RTV-SIFT method achieves superior registration results and is more robust to illumination variations, noise, and cloud occlusion. Our work provides a novel idea for optical and SAR image registration and demonstrates the potential of utilizing multi-modal image structural information for registration. We believe that our proposed method can be extended to other multimodal image registration fields, making significant contributions to the remote sensing domain.

Author Contributions: Conceptualization, funding acquisition and project administration, J.L.; Methodology and software, S.P., J.G. and L.H.; Validation, K.G. and Y.Z.; Data collection and processing, C.Z. and W.Z.; Writing—original draft, S.P.; Writing—review & editing, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China under Grants Nos U19B2030, 61976167, 62101416, the Natural Science Basic Research Program of Shaanxi under Grant No 2022JQ-708, and the Fundamental ResearchFunds for the Central Universities.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

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

References

1. Wan, L.; Xiang, Y.; You, H. An Object-Based Hierarchical Compound Classification Method for Change Detection in Heterogeneous Optical and SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9941–9959. [CrossRef]
2. Nie, M.; Ling, L.; Wei, X. A Novel Fusion and Target Detection Method of Airborne SAR Images and Optical Images. In Proceedings of the International Conference on Radar, Shanghai, China, 16–19 October 2006.
3. Poulain, V.; Inglada, J.; Spigai, M.; Tourneret, J.Y.; Marthon, P. High-resolution optical and SAR image fusion for building database updating. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 2900–2910. [CrossRef]
4. Jiang, X.; Ma, J.; Xiao, G.; Shao, Z.; Guo, X. A review of multimodal image matching: Methods and applications. *Inf. Fusion* **2021**, *73*, 22–71. [CrossRef]
5. Suri, S.; Reinartz, P. Mutual-Information-Based Registration of TerraSAR-X and Ikonos Imagery in Urban Areas. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 939–949. [CrossRef]
6. Lewis, J.P. Fast Normalized Cross-Correlation. *Circuits Syst. Signal Process.* **2009**, *28*, 819–843.
7. Liu, Y.; Wang, Q. Multi-sensor image registration based on local feature and its attributes set. In Proceedings of the IEEE 10th International Conference on Signal Processing Proceedings, Beijing, China, 24–28 October 2010; pp. 1053–1055.
8. 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]
9. Quan, D.; Wang, S.; Liang, X.; Wang, R.; Fang, S.; Hou, B.; Jiao, L. Deep generative matching network for optical and SAR image registration. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 6215–6218.
10. Merkle, N.; Auer, S.; Müller, R.; Reinartz, P. Exploring the potential of conditional adversarial networks for optical and SAR image matching. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1811–1820. [CrossRef]
11. Hughes, L.H.; 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]
12. Zampieri, A.; Charpiat, G.; Girard, N.; Tarabalka, Y. Multimodal image alignment through a multiscale chain of neural networks with application to remote sensing. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 657–673.
13. 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]
14. Ye, Y.; Zhu, B.; Tang, T.; Yang, C.; Xu, Q.; Zhang, G. A robust multimodal remote sensing image registration method and system using steerable filters with first-and second-order gradients. *ISPRS J. Photogramm. Remote Sens.* **2022**, *188*, 331–350. [CrossRef]
15. 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]
16. Dellinger, F.; Delon, J.; Gousseau, Y.; Michel, J.; Tupin, F. SAR-SIFT: A SIFT-like algorithm for SAR images. *IEEE Trans. Geosci. Remote Sens.* **2013**, *53*, 453–466. [CrossRef]
17. Low, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
18. Lindeberg, T. Feature Detection with Automatic Scale Selection. *Int. J. Comput. Vis.* **1998**, *30*, 79–116. [CrossRef]
19. Zhang, W. Combination of SIFT and Canny Edge Detection for Registration Between SAR and Optical Images. *IEEE Geosci. Remote Sens. Lett.* **2020**, *19*, 1–5. [CrossRef]
20. Fan, B.; Huo, C.; Pan, C.; Kong, Q. Registration of optical and SAR satellite images by exploring the spatial relationship of the improved SIFT. *IEEE Geosci. Remote Sens. Lett.* **2012**, *10*, 657–661. [CrossRef]
21. Xie, Z.; Liu, J.; Liu, C.; Zuo, Y.; Chen, X. Optical and SAR Image Registration Using Complexity Analysis and Binary Descriptor in Suburban Areas. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. . [CrossRef]
22. Li, J.; Hu, Q.; Ai, M. RIFT: Multi-modal image matching based on radiation-variation insensitive feature transform. *IEEE Trans. Image Process.* **2019**, *29*, 3296–3310. [CrossRef] [PubMed]
23. Forero, M.G.; Mambuscay, C.L.; Monroy, M.F.; Miranda, S.L.; Méndez, D.; Valencia, M.O.; Gomez Selvaraj, M. Comparative analysis of detectors and feature descriptors for multispectral image matching in rice crops. *Plants* **2021**, *10*, 1791. [CrossRef]
24. Sharma, S.K.; Jain, K.; Shukla, A.K. A Comparative Analysis of Feature Detectors and Descriptors for Image Stitching. *Appl. Sci.* **2023**, *13*, 6015. [CrossRef]
25. Mikolajczyk, K.; Schmid, C. Indexing based on scale invariant interest points. In Proceedings of the Proceedings Eighth IEEE International Conference on Computer Vision, ICCV 2001, Vancouver, BC, Canada, 7–14 July 2001; Volume 1, pp. 525–531.

26. Mikolajczyk, K.; Schmid, C. A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1615–1630. [CrossRef]
27. Chen, J.; Tian, J.; Lee, N.; Zheng, J.; Smith, R.T.; Laine, A.F. A Partial Intensity Invariant Feature Descriptor for Multimodal Retinal Image Registration. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 1707–1718. [CrossRef] [PubMed]
28. Fan, J.; Wu, Y.; Li, M.; Liang, W.; Cao, Y. SAR and optical image registration using nonlinear diffusion and phase congruency structural descriptor. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5368–5379. [CrossRef]
29. Shuai, J.; Jzang, U.; Wang, B.; Zhu, X.; Sun, X. Registration of SAR and Optical Images by Weighted Sift Based on Phase Congruency. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018.
30. Yu, Q.; Jiang, Y.; Zhao, W.; Sun, T. High-Precision Pixelwise SAR—Optical Image Registration via Flow Fusion Estimation Based on an Attention Mechanism. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 3958–3971. [CrossRef]
31. Kovesi, P. Image Features from Phase Congruency. *Videre J. Comput. Vis. Res.* **1999**, *1*, 1–26.
32. Kovesi, P. Phase congruency: A low-level image invariant. *Psychol. Res.* **2000**, *64*, 136–148. [CrossRef] [PubMed]
33. Kovesi, P. Phase congruency detects corners and edges. In *The Australian Pattern Recognition Society Conference: DICTA*; Csiro Publishing: Clayton, Australia, 2003; pp. 309–318.
34. Xu, L.; Yan, Q.; Xia, Y.; Jia, J. Structure extraction from texture via relative total variation. *ACM Trans. Graph. (TOG)* **2012**, *31*, 1–10. [CrossRef]
35. 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]
36. Schmid, C.; Mohr, R.; Bauckhage, C. Evaluation of interest point detectors. *Int. J. Comput. Vis.* **2000**, *37*, 151–172. [CrossRef]
37. Gonçalves, H.; Gonçalves, J.A.; Corte-Real, L. Measures for an objective evaluation of the geometric correction process quality. *IEEE Geosci. Remote Sens. Lett.* **2009**, *6*, 292–296. [CrossRef]
38. Xiang, Y.; Tao, R.; Wang, F.; You, H.; Han, B. Automatic Registration of Optical and SAR Images Via Improved Phase Congruency Model. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 5847–5861. [CrossRef]
39. 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]
40. 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]
41. Konnik, M.; Welsh, J. High-level numerical simulations of noise in CCD and CMOS photosensors: Review and tutorial. *arXiv* **2014**, arXiv:1412.4031.
42. Jaruwatanadilok, S.; Ishimaru, A.; Kuga, Y. Optical imaging through clouds and fog. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1834–1843. [CrossRef]