Article

A Robust 3D Density Descriptor Based on Histogram of Oriented Primary Edge Structure for SAR and Optical Image Co-Registration

Shuo Li ^{1,2,3} , Xiaolei Lv ^{1,2,3,*} , Jian Ren ⁴ and Jian Li ⁵

¹ Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, Chinese Academy of Sciences, Beijing 100190, China; lishuo193@mailsucas.ac.cn

² Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China

³ School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China

⁴ Lv Liangbei Company of Shanxi Communications Holding Group, Lvliang 033100, China; cbkzzx@sxql.com

⁵ Beijing Xingtian Information Technology Co., Ltd., Beijing 102200, China; jian.li@bj-xtd.com

* Correspondence: lvxl@aircas.ac.cn

Abstract: The co-registration between SAR and optical images is a challenging task because of the speckle noise of SAR and the nonlinear radiation distortions (NRD), particularly in the one-look situation. In this paper, we propose a novel density descriptor based on the histogram of oriented primary edge structure (HOPES) for the co-registration of SAR and optical images, aiming to describe the shape structure of patches more firm. In order to extract the primary edge structure, we develop the novel multi-scale sigmoid Gabor (MSG) detector and a primary edge fusion algorithm. Based on the HOPES, we propose the co-registration method. To obtain stable and uniform keypoints, the non-maximum suppressed SAR-Harris (NMS-SAR-Harris) and deviding grids methods are used. NMS-SSD fast template matching and fast sample consensus (FSC) algorithm are used to further complete and optimize matching. We use two one-look simulated SAR images to demonstrate that the signal-to-noise ratio (SNR) of MSG is more than 10 dB higher than other state-of-the-stage detectors; the binary edge maps and F-score show that MSG has more accurate positioning performance. Compared with the other state-of-the-stage co-registration methods, the image co-registration results obtained on seven pairs of test images show that, the correct match rate (CMR) and the root mean squared error (RMSE) improve by more than 25% and 15% on average, respectively. It is experimentally demonstrated that the HOPES is robust against speckle noise and NRD, which can effectively improve the matching success rate and accuracy.

Keywords: optical and SAR; image co-registration; primary edge detector; density descriptor

Citation: Li, S.; Lv, X.; Ren, J.; Li, J. A Robust 3D Density Descriptor Based on Histogram of Oriented Primary Edge Structure for SAR and Optical Image Co-Registration. *Remote Sens.* **2022**, *14*, 630. <https://doi.org/10.3390/rs14030630>

Academic Editor: Mi Wang

Received: 27 December 2021

Accepted: 27 January 2022

Published: 28 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 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

Synthetic aperture radar (SAR), as an active microwave imaging system, can obtain images regardless of time or cloud cover. However, the imaging characteristics make it difficult to conform to human vision and be interpreted like optical images. The complementary information of optical and SAR images plays a significant role in GCP extraction [1,2], image fusion [3], change detection [4], etc. The co-registration accuracy of SAR and optical images directly affects these applications. However, due to the speckle noise, the nonlinear radiometric distortion (NRD), the angular differences, and the radiometric differences between SAR and optical images, co-registration is challenging [5–7].

The previous image matching methods are divided into the area-based matching methods and feature-based matching methods. The area-based matching methods mainly include the normalized correlation methods (NCC) [8,9], mutual information methods (MI) [10–12], frequency domain-based methods [13], etc. They are registered by pixel

values and similarity measure. Because they are sensible to NRD and noise, they are more suitable for homomodal images rather than multi-sensor images. Feature-based matching methods aim at mining common features to form descriptors, mainly including the point features [14], line features [15], edge features [16], and contour features [17], etc. The feature-based matching methods are more robust and can reduce the effects of NRD and noise to a certain extent, so it is more suitable for the co-registration of SAR and optical.

In the traditional image matching field, the scale-invariant feature transform (SIFT) usually has high performance [18], and numerous improved algorithms have emerged to further enhance the performance [19–21]. However, due to the speckle noise and NRD, the traditional SIFT algorithm does not work well on SAR images. To address this problem, F Dellinger et al. proposed the SAR-SIFT, they redefined the gradient extraction part of SIFT using the ratio of exponentially weighted averages (ROEWA) to improve the robustness to speckle noise, which shows the high performance in SAR image co-registration [22]. For SAR and optical image co-registration, Xiang et al. proposed the OS-SIFT, which uses multi-scale ROEWA for SAR and multi-scale Sobel for optical in gradient extraction [23]. Yu et al. combined the spatial feature detection with local frequency-domain description and proposed an improved nonlinear SIFT-based co-registration framework [24]. All of these methods improved the robustness of the SIFT-like algorithm in SAR images.

In recent years, robust feature descriptors have been widely used in SAR and optical image co-registration. These methods incorporated several feature descriptions to improve the robustness against speckle noise and NRD, and significantly improved the accuracy and success rate of co-registration. Fan et al. proposed the uniform nonlinear diffusion-based Harris corner point extraction (UND-Harris) and phase consistent structure descriptors (PCSD) to match SAR and optical images [25]. Ye et al. proposed the histogram of orientation phase consistency (HOPC), which combines the advantages of histogram of oriented gradient (HOG) and phase consistency (PC) to achieve multimodal image matching [26]. They subsequently integrated the MMPC-lap feature detector with the local histogram of oriented phase congruency (LHOPC) to further increase performance [27]. Xiang et al. investigated the energy minimization method and the higher-order singular value decomposition method in the PC algorithm, improved the gradient extraction method of the traditional PC method, and further improved the robustness of the PC method in SAR and optical matching [28]. Xiong et al. proposed the rank-based local self-similarity (RLSS) to describe the local shape of an image and used RLSS descriptors of multiple local regions to construct dense-RLSS to complete the matching [29]. Li et al. constructed a maximum index map (MIM) with Log-Gabor to describe features and achieved rotation invariance by multiple MIM, and they refer to this as the radiation-invariant feature transform (RIFT) [30].

With the development of deep learning techniques, many researchers have tried to construct feature descriptions by using deep learning. He et al. proposed a remote sensing image matching technique based on siamese convolutional neural network to learn feature and similarity metrics [31]. Hughes et al. constructed SAR and optical image matching networks using semi-supervised learning, aiming to overcome dataset limitations [32]. Zhang et al. constructed a general workflow for multimodal remote sensing image co-registration based on a siamese fully convolutional network and adopted the strategy of maximizing the feature distance between positive and negative samples. [33]. Then, they used the Siamese convolutional network to learn pixelwise deep dense features to further improve the robustness of the network [34]. Dou et al. designed a generic matching network (GMN) based on the generic adversarial network (GaN) to increase the amount of training data and improve the matching performance [35]. Hughes et al. used GaN networks to generate very high resolution (VHR) SAR image blocks to achieve VHR SAR and optical matching [36]. Song et al. constructed MAP-Net to extract high-level semantic information by using spatial pyramid aggregated pooling (SPAP) and attention mechanism to complete cross-modal image matching [37].

However, although the deep learning methods are more likely to achieve performance leadership, they are limited by datasets and difficult to make breakthroughs in generaliza-

tion ability due to the factors such as resolution, shooting angle, and polarization. Therefore, traditional algorithms are still more widely used at this stage. The SIFT-like algorithms are slightly limited by the area features and inefficient at large images. The algorithms based on the robust feature descriptors usually achieve better results. However, in the case of one-look SAR images, the strong speckle noise causes an offset in the edge localization, which will lead to a degradation of co-registration performance. It is difficult to balance the localization ability of edges with the suppression ability of noise, so the co-registration performance is limited to full-resolution images.

In order to balance the localization ability of edges with the suppression ability of noise in the one-look SAR images and improve the co-registration performance, in this paper, we propose a novel 3D density feature descriptor based on primary structure information to achieve SAR and optical co-registration. The primary edge structure is defined as the edge response where the contrast changes significantly, which is usually manifested as the edge structure of the object that is significantly different from the surrounding ground objects. The primary edge can reflect the structure feature of nearby ground objects and has a high similarity between the optical and SAR images. Correspondingly, the weak edge has a low edge response value, which is usually generated by noise or inconspicuous object edges. The weak edge exhibits significant differences between SAR and optical images. It bases on a feature-based method and consists of three steps: keypoints detection, feature extraction, and feature matching. Compared with other algorithms, the main contributions of the proposed method are summarized as follows:

- The primary edge extraction operator multi-scale sigmoid Gabor (MSG) is proposed to improve the noise rejection ability and maintain the edge localization ability. The MSG is compared against the traditional operator ratio of exponentially weighted averages (ROEWA) [38], the state-of-the-art operators ratio based edge detector (RBED) [39], and unbiased difference ratio edge detector (UDR) [40]. In comparison to the other detectors above, MSG has the higher signal-to-noise ratio (SNR) and edge positioning performance.
- Based on primary edge structure information, we propose a novel 3D density feature descriptor, histogram of orientated primary edge structure (HOPES). It consists of more robust multi-angle structural information of SAR and optical images. Compared with the traditional method MI, the state-of-the-art methods CFOG and HOPC, HOPES can obtain more sharper similarity measure map.
- In the co-registration process, the non-maximum suppressed SAR-Harris (NMS-SAR-Harris) and deviding grids method are used to obtain stable and uniform keypoints. The HOPES descriptor is used to extract the features of keypoints. The NMS-SSD fast template match method is used to match feature. The fast sample consensus (FSC) algorithm [41] is used to remove outliers.

The remainder of this paper is organized as follows: In Section 2, we firstly propose the multi-scale sigmoid Gabor filter and the primary edge fusion algorithm; and then, we introduce the HOPES in detail; finally, we introduce the co-registration process, including keypoint extraction, and NMS-SSD fast matching. In Section 3, the performance of MSG, the accuracy and robustness of HOPES matching, and the effect of parameters on performance are evaluated and discussed experimentally. Finally, conclusions are provided in Section 4.

2. Methodology

2.1. Multi-Scale Sigmoid Gabor Filter (MSG)

Due to the noise and NRD, using edge structure features is a more robust co-registration method for SAR and optical images. However, it is a great challenge to extract edge features robustly and accurately. Noise and weak contrast areas cause weak edges, which are different from optical images and unsuitable for matching features. In contrast, the primary edges reflect the edge structure feature of the strong contrast region and have higher similarity with the optical image. In order to obtain primary edges and suppress the influence of speckle noise, a multi-scale sigmoid Gabor filter (MSG) is designed.

Mehrotra et al. demonstrated that the odd-symmetric part of the Gabor filter is an efficient and robust edge detection operator [42], and the two-dimensional Gabor filter at angle θ is defined as:

$$Go(x, y) = \exp\left[-\frac{x^2 + y^2}{2\sigma^2}\right] \sin[\omega(x \cos \theta + y \sin \theta)] \quad (1)$$

where ω is the frequency of the sine function and σ is the scale of the Gaussian function. Figure 1 gives the two-dimensional Gabor filters on $\theta = 0$ and $\theta = \pi/4$.

Figure 1. Gabor filter window at two angles: (a) $\theta = 0$. (b) $\theta = \pi/4$.

Two adjacent non-overlapping Gabor windows are given by the following equation:

$$\begin{aligned} Go_1^{\sigma\theta}(x, y) &= Go(x, y), & x \sin \theta - y \cos \theta &\geq 0 \\ Go_2^{\sigma\theta}(x, y) &= -Go(x, y), & x \sin \theta - y \cos \theta &< 0 \end{aligned} \quad (2)$$

The local means μ_1 and μ_2 of added both windows can be obtained by computing the convolution of the filter with the grayscale of original image:

$$\begin{aligned} \mu_1 &= I(x, y) \otimes Go_1^{\sigma\theta} \\ \mu_2 &= I(x, y) \otimes Go_2^{\sigma\theta} \end{aligned} \quad (3)$$

Then we can obtain the result of single-scale Gabor edge detector:

$$R^{\sigma\theta} = \min(\mu_1/\mu_2, \mu_2/\mu_1) \quad (4)$$

In Gabor filter defined by Equation (1), σ reflects the scale property of the filter. Build a scale space of n scales $S_n = \{\sigma_1, \dots, \sigma_n\}$, where the relationship between two adjacent scale parameters is $\sigma_{i+1}/\sigma_i = k$. Thus, the scale-edge strength map (S-ESM) and scale-edge direction map (S-EDM) can be obtained:

$$\begin{aligned} S - ESM &= \{ESM_1, \dots, ESM_n\} \\ S - EDM &= \{EDM_1, \dots, EDM_n\} \end{aligned} \quad (5)$$

where ESM_n and EDM_n are given by:

$$\begin{aligned} ESM_n &= 1 - \min_{\theta} R^{\sigma_n\theta} \\ EDM_n &= \arg \min_{\theta} R^{\sigma_n\theta} + \pi/2 \end{aligned} \quad (6)$$

The ESM at different scale is shown as Figure 2:

Figure 2. The ESM of different scale: (a) Simulated SAR image. (b) $n = 1$. (c) $n = 3$. (d) $n = 5$.

At a small scale, the Gabor filter performs better at localizing edges, but it is more sensitive to noise. At a wide scale, the Gabor filter has less edge localization capability but can suppress noise more effectively. Therefore a fusion method is considered to suppress noise and maintain edges. It is generally considered that the lower response of the Gabor filter is resulted by weak edges. The weak edges are usually cause by noise and unsharp edges, and they need to be suppressed. So, we proposed a sigmoid function to build the multi-scale sigmoid Gabor filter (MSG), which aims to reduce value at the lower response of the filter. A measure of filter response spread at single-scale can be generated by dividing the sum of responses by the highest response. Then we normalize it by the number of scales being used, and thus can obtain a fractional measure of spread that varies between 0 and 1. This spread is given by:

$$s(x) = \frac{1}{N} \left(\frac{\sum_n A_n(x)}{\varepsilon + A_{\max}(x)} \right) \quad (7)$$

where N is the total number of scale, $A_{\max}(x)$ is the maximum value of the filter response at that scale, and ε is used to avoid division by zero. Then the weights at this scale can be expressed as:

$$W(x) = \frac{1}{1 + e^{\gamma(c-s(x))}} \quad (8)$$

where c is the cutoff value of the filter; the edge response will be penalized to a certain extent below it. γ is the gain factor that controls the cutoff response; changing the value of γ can adjust the weight of primary and weak edges; increasing the value of γ enhances the suppression of weak edges and highlights primary edges. However, a value of γ that is excessively large tends to result in the loss of edge features, as shown in Figure 3. Therefore, the weights in the scale space are expressed as:

$$W_s = \{W_1, \dots, W_n\} \quad (9)$$

Figure 3. Effect of γ on strong and weak edges: (a) simulated SAR image. (b–d) $\gamma = \{3, 9, 15\}$.

The primary edge fusion algorithm in the scale space is described as follows:

1. If $ESM_{\sigma_i}(x, y) > Threshold$ holds for $i \in \{1, \dots, n\}$, then (x, y) is defined as the true edge point.
2. For the true edge points, weights in the scale space are:

$$W_s(x, y) = \{W_1(x, y), \dots, W_n(x, y)\} \quad (10)$$

While the weights at weak edges decrease significantly as the scale increases, the weights at primary edges remain relatively constant at different scales. Therefore we choose the minimum weight strategy to suppress noise and refine edges. Figure 4 shows the effect of different weight strategies, and it can be seen that the minimum weight strategy can obtain the minimum noise and the best edge localization performance.

3. The fused ESM is obtained from the following equation:

$$ESM_{\text{fusion}}(x, y) = \sum_{i=1}^n W_{\min}(x, y) \times ESM_i(x, y) \quad (11)$$

Figure 4. Impact of different strategies on W : (a) maximum weight strategy; (b) average weight strategy; (c) minimum weight strategy.

2.2. Histogram of Oriented Primary Edge Structure

Inspired by the histogram of oriented gradient (HOG), we develop a 3D density descriptor called the histogram of oriented primary edge structure (HOPES) using the primary edge structure produced via MSG. Since the images have been coarsely adjusted, the same point coordinates can be used to extract the template window and search window separately. Figure 5 gives the main flow of constructing HOPES 3D density descriptor.

Figure 5. Main flow of constructing HOPES 3D density descriptor.

Define orientation space $\theta = \{\theta_1 \dots, \theta_k\}$, where $\theta_k = k\pi/M$, M is the number of orientations, $k = 0, 1, \dots, M-1$ is the index of orientation. Firstly, construct the Gabor bilateral filter template in orientation-scale space from Equations (1) and (2), and then the orientation-scale edge strength map (OS-ESM) in orientation-scale space can be obtained as:

$$OS-ESM = \begin{Bmatrix} ESM_1^1 & ESM_2^1 & \dots & ESM_n^1 \\ ESM_1^2 & ESM_2^2 & \dots & ESM_n^2 \\ \dots & \dots & \dots & \dots \\ ESM_1^k & ESM_2^k & \dots & ESM_n^k \end{Bmatrix} \quad (12)$$

where ESM_n^k denotes ESM of the k -th orientation at the n -th scale:

$$ESM_n^k = 1 - R^{\sigma_n \theta_k} \quad (13)$$

Perform primary edge fusion algorithm on S-ESM at each orientation, and the primary edge feature map under the orientation space (O-PEFM) can be obtained as:

$$O - PEFM = \{ESM_{fusion}^1, \dots, ESM_{fusion}^k\} \quad (14)$$

where ESM_{fusion}^k denotes the primary edge obtained by the primary edge fusion algorithm at the k -th orientation:

$$ESM_{fusion}^k = \sum_1^n W_k \times ESM_n^k \quad (15)$$

Convolve the O-PEFM by a 3D Gaussian-like kernel g , then we can get the 3D density structure descriptor HOPES. This 3D Gaussian-like kernel consists of a two-dimensional Gaussian kernel in the X and Y directions and a one-dimensional Gaussian kernel in the Z direction. The convolution in the X and Y directions reduces noise, while convolution in the Z direction smooths the oriented gradient, hence reducing directional distortion produced by local geometric and intensity distortion. To demonstrate the role of Z-direction Gaussian convolution, we add multiplicative noise with a mean of 0 and a variance of 0.06 to a high-resolution optical image to simulate the noise distribution of the SAR image, and add distortion to simulate the directional distortion of the SAR image. Compare the gradient direction histograms of HOPES without Z-direction convolution, with $[1, 2, 1]^T$ convolution kernel [43], with $\sigma = 1, size = 5$, and $\sigma = 3, size = 11$ Gaussian kernels, (see Figure 6), and their statistical variances are in Table 1. As can be seen, the gradient direction histogram obtained by using a Gaussian kernel for Z-direction convolution is more similar to the original one, and it should be noted that using a larger Gaussian kernel and σ can achieve a better smoothing effect, but at the same time, it is more likely to cause overfitting phenomenon and lose its original gradient direction histogram feature. Therefore, in practical applications, we choose a Gaussian kernel with $\sigma = 1, size = 5$. Finally, L2 regularization is applied to the Z-direction to overcome the grayscale difference between SAR and optical images to obtain the final HOPES descriptor. HOPES is by constructed using all pixels, so it can be considered as a density descriptor.

Table 1. Influence of different Z-direction convolution strategies on the variance of gradient direction histogram.

Conv Kernel	No Conv	$[1, 2, 1]^T$	$Gauss_{\sigma=1, size=5}$	$Gauss_{\sigma=3, size=11}$
Variance	149.9107	139.9921	129.2999	39.5024

To generate the structure descriptor, the proposed HOPES descriptor computes a histogram of primary edge for M directions centered on each pixel. Visualize the HOPES descriptor, as shown in Figure 7a. Firstly, divide the obtained HOPES descriptor into $n \times n$ subregions, add up the gradient directions in the subregions, count the gradient values of all directions in all subregions, and normalize them. Take their size as the size of gradient vectors and the direction as the direction of the gradient vectors, then we can get the visualized HOPES descriptor. Figure 7b shows that the HOPES descriptor is able to obtain distinct gradient directional features at the edges, while the gradient direction is more balanced at the noise with no significant directionality.

Figure 6. Comparison of Z-direction convolution of HOPES descriptors: (a) original image; (b) distorted and noise contaminated images; (c) HOPES descriptors without the Z-direction convolution; (d) HOPES descriptors with $[1, 2, 1]^T$ Z-direction convolution; (e) HOPES descriptors with $\sigma = 1$, size = 5 Z-direction Gauss convolution. (f) HOPES descriptors with $\sigma = 3$, size = 11 Z-direction Gauss convolution.

Figure 7. HOPES descriptor visualization: (a) subregion of HOPES descriptor; (b) HOPES gradient direction of simulated SAR image.

2.3. Co-Registration Algorithm

The flowchart of the proposed algorithm is shown in Figure 8. The algorithm consists of three steps: keypoints detection, feature extraction, and feature matching. The step of feature extraction has been introduced above, so we will introduce step 1 keypoints extraction and step 3 feature matching next.

Figure 8. Flowchart of proposed match algorithm.

2.3.1. Keypoints Extraction

Firstly, the geography information is used to register the SAR and optical images coarsely. The initial registration is crucial to the entire registration process, and the incidence angle will affect registration results. We use two GF-3 images at the same region with different orbital directions and different incidence angles to prove that the increase of the incident angle will lead to a decrease in the registration success rate and accuracy. That is because the increase of incidence angle will cause the offset of the feature information and the increase of structural feature difference. In addition, an excessively large incidence angle may cause the obscuration of objects. So, the input SAR image is an orthoimage that has been coarsely adjusted by RD or RPC modal with digital elevation model (DEM). The purpose of using DEM is to maintain the stability of the primary structure. The coarse orthoimage eliminates most of the displacement, rotation, and distortion. Next, keypoints are extracted on SAR image by dividing grids or NMS-SAR-Harris algorithm. Dividing grids method aims to get evenly distributed points. In this method, we divide the $m \times n$ image uniformly on the row and column coordinates by x pixels, and then we can get the grid points. NMS-SAR-Harris uses non-maximum suppression based on SAR-Harris. It selects the point with the most significant response in a region as the keypoint, which effectively solves the problem of duplication of points and improves the matching efficiency, as shown in Figure 9:

Figure 9. Comparison of SAR-Harris and NMS-SAR-Harris for keypoint extraction: (a) keypoints obtained by SAR-Harris algorithm; (b) keypoints obtained by improved SAR-Harris algorithm. The latter improves the problem of points repetition in the region.

2.3.2. Feature Matching

The sum of squared difference (SSD) can be used as a similarity measure. Due to the complexity of 3D density descriptor, we transform the spatial convolution operation of SSD into the frequency domain by FFT for point multiplication calculation to reduce the computational complexity [43]. Non-maximum suppression is also introduced to obtain matching results with higher confidence. Suppose the HOPES descriptors of the reference template window and the search window are D_l and D_r , then the SSD of D_l and D_r are defined as:

$$S = \sum_{(x,y)} (D_l(x,y) - D_r(x,y))^2 \quad (16)$$

Expanding the above equation:

$$S = \sum_{(x,y)} D_l^2(x,y) + \sum_{(x,y)} D_r^2(x,y) - 2 \sum_{(x,y)} D_l(x,y) D_r(x,y) \quad (17)$$

The templates achieve optimal matching when the SSD metric is minimized. In the above equation, $\sum_{(x,y)} D_l^2(x,y)$ and $\sum_{(x,y)} D_r^2(x,y)$ are constant terms, so maximum $\sum_{(x,y)} D_l(x,y) D_r(x,y)$:

$$(x',y') = \arg \max_{(x',y')} \left(\sum_{(x,y)} D_l(x,y) D_r((x,y) - (x',y')) \right) \quad (18)$$

where (x',y') denotes the translational relationship between two windows. To speed up the computation, the convolution calculation $D_l(x,y) D_r((x,y) - (x',y'))$ is transferred to the frequency domain by the Fourier transform, then we have:

$$(x',y') = \arg \max_{(x',y')} \left(\sum_{(x,y)} F^{-1} \left\{ F\{D_l(x,y)\} \cdot F\{D_r((x,y) - (x',y'))\}^* \right\} \right) \quad (19)$$

where F and F^{-1} denote the Fourier transform and inverse transform, and $*$ denotes the complex conjugate. Using the above method, the translation relation (x',y') can be calculated. Inevitably, the matching results may have multi-peakedness, as in Figure 10a. In order to improve the matching accuracy and robustness, non-maximum suppression is used to select the maximum similarity point. We call it NMS-SSD. Firstly, the pixel values of similarity measure map are ordered decreasing, the first N extreme points are selected as seed points. The coordinates (x_i, y_i) are taken as the upper left point of window (x_i^{tl}, y_i^{tl}) , and the pixel values are recorded as matching scores. Given the search window size, and the lower right corner point of search window is (x_i^{rb}, y_i^{rb}) . The window area of all seed

points are compared with the window area of the highest matching score (main peak point), respectively, and if it is larger than the discriminant threshold N_t of the overlapping area ratio, as in Figure 10b, it will be removed from the seed points; if it is smaller than N_t , as in Figure 10c, it will be kept as a sub-peak point. If there are sub-peak points and the ratio of primary to secondary peak is greater than the threshold t , or only primary peak exists, then the primary peak will be used as matching point. The matching confidence is further improved by NMS-SSD.

Figure 10. Non-maximum suppressed SSD fast template matching: (a) if there are repetitive structural features in the region, the similarity measure map may have multiple peaks; (b) if the area ratio of overlapping regions is larger than discriminant threshold N_t , the seed point is rejected; (c) if the area ratio of the overlapping regions is smaller than discriminant threshold N_t , keep it as a sub-peak point.

3. Experimental Results and Discussion

In this section, we evaluate the edge extraction performance of the MSG operator using two simulated SAR images with added multiplicative noise. We also compare the similarity measure map of HOPES with the traditional method MI and two state-of-the-stage methods CFOG and HOPC. Then, seven pairs of SAR and optical images are used to test the HOPES and analyze the results. The co-registration performances are evaluated via objective and subjective methods. One is the evaluation criteria, and the other is to use a chessboard mosaic image and enlarged submaps. Finally, the influence of different parameters on the performance is experimentally evaluated and analyzed.

3.1. Comparison of Edge Extraction

In the proposed matching algorithm, the accurate acquisition of edge information is the key to successful matching. Therefore, in this section, we compare the proposed MSG with the existing popular SAR edge detection algorithms, including the traditional detector ROEWA [38], two state-of-the-stage detectors RBED [39], and UDR [40].

3.1.1. Datasets and Parameters Settings

Firstly, two sets of one-look simulated images are used for the experiment. To simulate the SAR image, we added speckle noise to the images, and make them to have a similar noise distribution to the true SAR image. Figure 11a, is a simulated one-look SAR image of size 1000×600 . Figure 11c is a simulated one-look SAR image of size 1500×1500 , in this image, the detection difficulty is obviously increased by the effects of transition zone size and edge contrast. The ground truth (GT) maps are shown as Figure 11b,d, which clearly indicate the edge, non-edge region and transition region. In the GT map, the black line represents the true location of the edge. The eight-neighborhood of the true edge is the transition region, which is the white area in the GT map. The black region in the GT map is the non-edge region.

Figure 11. (a,b) The first simulated SAR image and its ground truth (GT) map; (c,d) the second simulated SAR image and its ground truth (GT) map.

The detectors ROEWA $h_{ROEWA}(x, y)$, RBED $h_{RBED}(x, y)$, UDR $h_{UDR}(x, y)$ are noted as:

$$\begin{cases} h_{ROEWA}(x, y) = h_{ROEWA}^{\perp}(|x|)h_{Gaussian}^{\parallel}(y) \\ h_{ROEWA}^{\perp}(|x|) = \exp(-\zeta_{ROEWA}|x|) \\ h_{Gaussian}^{\parallel}(y) = \exp(-y^2/2\sigma_{ROEWA}^2) \\ \zeta_{ROEWA} > 0, \sigma_{ROEWA} > 1 \end{cases} \quad (20)$$

$$\begin{cases} h_{RBED}(x, y) = h_{RBED}^{\perp}(|x|)h_{RBED}^{\parallel}(y) \\ h_{RBED}^{\perp}(|x|) = |x|^{\alpha_{RBED}} \exp(-|x|^2/\beta_{RBED}^2) \\ h_{RBED}^{\parallel}(y) = \begin{cases} 1, & |y| \leq l_{RBED} \\ \exp\left(\frac{-(|y|-l_{RBED})^2}{2\sigma_{RBED}^2}\right), & |y| > l_{RBED} \end{cases} \\ \alpha_{RBED} > 0, \beta_{RBED} \neq 0, l_{RBED} \geq 0, \sigma_{RBED} > 1 \end{cases} \quad (21)$$

$$\begin{cases} h_{UDR}(x, y) = h_{UDR}^{\perp}(|x|)h_{UDR}^{\parallel}(y) \\ h_{UDR}^{\perp}(|x|) = |x|^{\alpha_{UDR}-1} \exp\left(\frac{|x|}{\beta}\right) \\ h_{UDR}^{\parallel}(y) = \begin{cases} 1, & |y| \leq l_{\parallel} \\ \exp\left(\frac{-(|y|-l_{\parallel})^2}{2\sigma_{\parallel}^2}\right), & |y| > l_{\parallel} \end{cases} \\ \alpha_{UDR} > 1, \beta_{UDR} > 0, l_{\parallel} \geq 0, \sigma_{\parallel} > 1 \end{cases} \quad (22)$$

In the next experiments, all possible combinations of parameters in the parameter space suggested by the original algorithm will be tried to obtain the best results. Furthermore, the parameter space of MSG is:

$$\begin{cases} \sigma_{MSG} = 1.4 + 0.2k_1, & k_1 = 0, 1, \dots, 5 \\ k_{MSG} = 1.2 + 0.1k_2, & k_2 = 0, 1, \dots, 8 \\ \gamma_{MSG} = 6 + k_3, & k_3 = 0, 1, \dots, 9 \end{cases} \quad (23)$$

3.1.2. Evaluation Criteria

Signal-to-noise ratio (SNR): We define the edge and transition regions in Figure 11b,d as signal S , and the other regions in Figure 11b,d as noise N . Use Equation (24) to calculate SNR:

$$SNR = 10\log_{10}(S/N) \quad (24)$$

F-score: In order to compare the edge localization performance between different algorithms, we use NSHT post-processed to obtain binary edges. The edge pixels are counted as true positive (TP) if the detector extracts them in the edge and transition regions, false positive (FP) if they are reported in the non-edge regions. The non edge pixels reported in the edge region are counted as false negative (FN), and the non edge pixels reported in the non edge region are counted as true negative (TN). Use Equation (25) to calculate the performance of the detection operator.

$$\begin{aligned} Precision &= \frac{TP}{TP + FP} \\ Recall &= \frac{TP}{TP + FN} \\ F-score &= 2 \times \frac{Precision \times Recall}{Precision + Recall} \end{aligned} \quad (25)$$

3.1.3. Results and Discussion

The edge strength maps obtained by ROEWA, RBED, UDR, and MSG are shown as Figure 12, their SNR are shown as Figure 13. Figure 14 gives the binary edge image after NSHT algorithm; and Table 2 gives F-scores of the four detectors.

Figure 12. Edge strength maps obtained by ROEWA, RBED, UDR, and MSG on the one-look simulated SAR images: (a,e) ROEWA; (b,f) RBED; (c,g) UDR; (d,h) MSG.

Figure 13. SNR of ROEWA, RBED, UDR and MSG on (a) the first one-look simulated SAR image and (b) the second one-look simulated SAR image.

Figure 14. Comparison of the binary edge maps obtained by (a,e) ROEWA; (b,f) RBED; (c,g) UDR; (d,h) MSG edge detectors.

Table 2. Performance comparison of detection operators.

Method	Data 1				Data 2			
	ROEWA	RBED	UDR	MSG	ROEWA	RBED	UDR	MSG
Precision	0.8938	0.9146	0.9555	0.9598	0.8093	0.8983	0.9394	0.9553
Recall	0.9039	0.9077	0.9521	0.9564	0.6893	0.7762	0.8655	0.8797
F-score	0.8988	0.9111	0.9538	0.9581	0.7445	0.8328	0.9010	0.9159

In the two sets of test data, ROEWA, RBED and UDR can extract edges, however, if we want to produce a sharp edge strength map, they will be inevitably affected by noise, and if we adjust parameters to avoid the influence of noise as much as possible, the edges will be thickened, which is not conducive to the subsequent registration task. The SNR of MSG is significantly better than other algorithms, it shows stronger edge response and noise suppression ability. A stronger SNR can be obtained by adjusting γ_{MSG} , however, it will result in the loss of some edge features due to the over-inhibition of low contrast edges. Experimental results of F-score show that MSG has better edge localization performance than other operators in one-look simulated SAR images. Among the four algorithms, MSG detects the most complete binary edge, as shown in Figure 14 and Table 2.

3.2. Comparison of the Feature Descriptors

We use traditional algorithm MI [10] and two state-of-the-art methods HOPC [26], CFOG [27] and HOPES to compare the differences of the four feature descriptors. Figure 15 shows similarity measure maps of CFOG, HOPC, MI and HOPES. We use the same template size and search radius for each group. HOPES can achieve correct matching under all four sets of data. It can get sharper and more obvious peaks at the matching point, while smoother at the non-matching area. Data 1 is a water pond, and it can be seen that due to the obvious temporal difference between SAR image and optical image, there is a certain difference between the template image and the search image, HOPES uses the primary edge feature of the template area to avoid the effect of temporal difference and thus achieves the co-registration. CFOG and HOPC have a higher matching success rate than MI because they use the structural information of the images, but there is a phenomenon of multi-peakedness. In comparison, HOPES has higher robustness.

Figure 15. Similarity measure maps of CFOG, HOPC, MI, and HOPES for the SAR and optical image pairs.

3.3. Comparison of SAR and Optical Co-Registration

In this subsection, we test seven sets of SAR and optical images with HOPC, CFOG, SAR-SIFT, OS-SIFT and our proposed algorithm. SAR-SIFT and OS-SIFT are the SIFT-like algorithms; HOPC and CFOG represent histogram of phase congruency and gradient, respectively. Subjective and objective criteria are used to evaluate the performance of the co-registration algorithms.

3.3.1. Datasets and Parameters Settings

In our experiment, to compare the co-registration performance between different algorithms, we select seven groups of SAR and optical images with different regions, resolutions, and time. As shown in Figure 16. Table 3 lists the information of each image pair.

Table 3. Information for the test images.

Pair	Source	Date	Size (Pixel)	Resolution	Region
A	GF3 Google Earth	09/2020 02/2013	2000 × 2000	10 m	Mountain
B	GF3 Google Earth	03/2021 02/2019	2500 × 2500	3 m	Urban
C	GF3 Google Earth	11/2020 02/2017	930 × 1020	5 m	Pond
D	GF3 Google Earth	12/2020 11/2018	1000 × 1000	1 m	River
E	GF3 Google Earth	11/2020 12/2017	1600 × 1600	5 m	Lake
F	GF3 Google Earth	11/2020 08/2017	1500 × 1500	5 m	Saltworks
G	GF3 Google Earth	12/2020 01/2010	4000 × 3000	1 m	Plateau

Figure 16. Experimental images. (a–g) are pair A–G in Table 3.

SAR images are from GF-3 satellite. Optical images are collected from Google Earth, mosaicked by different satellite images. Pair A is a mountainous area with an average elevation of about 1500 m. Apart from having generally similar trends of mountain range, the images vary widely and pose a challenge for co-registration. Pair B is the urban area with large radiometric differences. Surface features of Pair C and Pair F are the fish ponds and saltworks, which contain numerous repetitive texture features that can interfere with co-registration. Pair D is the tropical river area with large radiometric differences. Pair E contains a large lake area and the upper right part contains a urban area. Pair G is a plateau area with an average altitude of more than 4500m, with a resolution of 1m and large image differences, which is a big challenge for co-registration. All these images contain some temporal differences.

We compare our algorithm with CFOG, HOPC, SAR-SIFT and OS-SIFT. We apply the NMS-SAR-Harris and deviding grids methods to extract the keypoints in the HOPES matching algorithm to compare the differences of keypoint extraction methods, and apply NMS-SAR-Harris method to extract keypoints in CFOG and HOPC. We keep consistent with the relevant parameter settings in NMS-SAR-Harris of these three algorithm to ensure that the number of keypoints is the same. As the SIFT-like algorithms, SAR-SIFT and OS-SIFT follow the recommended settings of the original author. All algorithms use FSC algorithm for post-processing to eliminate outliers. The model is set to affine transformation with RMSE of 3. Parameters of HOPES are set to $\sigma_{MSG} = 1.4$, $k_{MSG} = 1.4$, $\gamma_{MSG} = 6$, scale parameter $N_{scale} = 3$, and orientation is set to $M = 8$.

3.3.2. Evaluation Criteria

As a subjective evaluation criteria, the checkboard mosaic image and enlarged sub-images are displayed to observe the effect and details of image co-registration. We also use the following criteria to analyze the performance objectively and quantitatively.

Root mean squared error (RMSE): We select 10~20 pairs of corresponding points manually to estimate the affine transformation matrix H . The RMSE is calculated as [44]:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N \left((x_i^{ref} - (x_i^{ses})')^2 + (y_i^{ref} - (y_i^{ses})')^2 \right)} \quad (26)$$

In Equation (26), N is the number of selected corresponding point pairs, (x_i^{ref}, y_i^{ref}) represents position of the i th keypoint in reference image, (x_i^{ses}, y_i^{ses}) represents position of the i th matched point in image to be registered. (x_i^{ses}, y_i^{ses}) is transformed by the affine matrix H to obtain its corresponding coordinates in reference image $((x_i^{ses})', (y_i^{ses})')$. A smaller RMSE indicates a higher accuracy of co-registration algorithm.

Number of correct matches (NCM): It counts the number of correct matches after FSC.

Correct match rate (CMR): It counts the success matching rate. It should be noted that SAR-SIFT and OS-SIFT are different from HOPES, CFOG, and HOPC algorithms, the SIFT-like algorithms adopt global search strategy and their initial keypoints are different with others. So we do not count CMR on them. CMR are defined as follows:

$$CMR = \frac{2 \times NCM}{N_1 + N_2} \quad (27)$$

In Equation (27), N_1 and N_2 are the number of keypoints extracted from two images, in our algorithm, $N_1 = N_2$.

3.3.3. Results and Discussion

The co-registration results are shown in Figure 17, and the statistics of the six co-registration strategies are shown in Table 4.

Table 4. NCM, MCR and RMSE values of different matching strategies.

Pair	Performance	HOPES (Harris)	HOPES (Grid)	CFOG	HOPC	SAR-SIFT	OS-SIFT
A	NCM	283	764	281	185	-	-
	CMR	80.86%	82.15%	80.29%	52.86%	-	-
	RMSE	1.3319	1.3474	2.1204	1.4159	-	-
B	NCM	798	862	843	541	-	-
	CMR	53.96%	58.16%	57.00%	36.58%	-	-
	RMSE	1.2873	1.2843	1.7119	1.1951	-	-
C	NCM	257	166	161	268	15	71
	CMR	92.91%	88.46%	60.07%	43.28%	-	-
	RMSE	0.9641	0.9546	1.7879	1.0875	3.7227	2.6154
D	NCM	194	184	182	182	-	31
	CMR	98.98%	93.88%	92.86%	92.86%	-	-
	RMSE	1.1712	1.1237	2.0324	1.1323	-	11.3869
E	NCM	472	236	456	428	18	217
	CMR	87.90%	81.66%	84.92%	79.70%	-	-
	RMSE	0.8637	0.8798	1.6992	0.9177	2.4089	3.0335
F	NCM	726	232	639	560	13	221
	CMR	96.41%	90.63%	84.86%	74.37%	-	-
	RMSE	0.9309	0.9468	1.0940	0.9637	4.5631	3.1541
G	NCM	370	1291	304	269	-	-
	CMR	79.23%	78.48%	65.10%	57.60%	-	-
	RMSE	1.2591	1.2479	1.8305	1.3279	-	-

Figure 17. Co-registration results of Pair A–G.

Our HOPES achieves the best CMR in all seven data sets and the best RMSE in six data sets. There is little difference between NMS-SAR-Harris and deviding grids strategy. Two SIFT-like algorithms, SAR-SIFT and OS-SIFT, achieved the worst results. As a co-registration algorithm designed for SAR images, SAR-SIFT is difficult to complete the co-registration between optical and SAR in the most cases. OS-SIFT improves the gradient extraction in SAR-SIFT, making it more competent for optical and SAR co-registration. It can complete the co-registration in the textured scenes such as lakes and ponds, but the registration accuracy is not high. At the same time, both algorithms are time-consuming

in co-registration of images with larger size and richer texture because of global search. To visually compare the co-registration results, Figures 18–24 gives the mosaic board map after HOPES matching and a local zoom of the board map obtained by the successful co-registration algorithm. Next, we analyze and discuss the co-registration results for each pair of images.

In Pair A, due to the terrain and squint of SAR, the imaging results are very different from the optical image. At the same time, in areas with large topographic changes, the orthophoto used DEM has smear, which will interfere with feature extraction. SAR-SIFT and OS-SIFT failed to register. CFOG (Figure 18c) and HOPC (Figure 18d) have a certain offset, and the offset of CFOG is larger. HOPES (Figure 18b) uses primary edges to construct primary structural features, it has the highest accuracy and the smallest offset, which is consistent with the RMSE results in Table 4. In Pair B, SAR-SIFT and OS-SIFT fail to register, the offset of CFOG (Figure 19c) is the largest, and there is little difference between HOPC (Figure 19d) and HOPES (Figure 19b). HOPC achieves the best RMSE in Table 4. In Pair C, due to the rich texture details, all algorithms complete the co-registration. HOPES (Figure 20b) has the best CMR and RMSE, CFOG (Figure 20c) and OS-SIFT (Figure 20f) have offset at two marks, HOPC (Figure 20d) and SAR-SIFT (Figure 20e) have offset at mark ①. In Pair D, HOPES (Figure 21b) has the best CMR and RMSE, and CFOG (Figure 21c) has a small range of offset at three marks. SAR-SIFT fails to register successfully. Although OS-SIFT (Figure 21e) completes the co-registration, the accuracy is the worst. The RMSE is 11.3869, and there are obvious offsets at three marks. In Pair E, HOPES (Figure 22b) has the best CMR and RMSE, with no significant offset of two marks. However, CFOG (Figure 22c) and HOPC (Figure 22d) have offset at mark ②, and SAR-SIFT (Figure 22e) and OS-SIFT (Figure 22f) have offset at two marks. In Pair F, HOPES (Figure 23b) has the best CMR and RMSE, CFOG (Figure 23c) has an insignificant offset at mark ①, HOPC (Figure 23d) has an insignificant offset at mark ②, and SAR-SIFT (Figure 23e) and OS-SIFT (Figure 23f) have significant offsets at two marks. In Pair G, similar to Pair A, SAR and optical images of the plateau area are very different, SAR-SIFT and OS-SIFT fail to register. The offset of HOPES (Figure 24b) is the smallest, and it has the best CMR and RMSE. CFOG (Figure 24c) and HOPC (Figure 24d) both have offset.

Figure 18. Checkboard mosaic image and enlarged sub-images of Pair A: (a) checkboard mosaic image of HOPES; (b–d) enlarged sub-images of HOPES, CFOG, HOPC.

Figure 19. Checkboard mosaic image and enlarged sub-images of Pair B: (a) checkboard mosaic image of HOPES; (b–d) enlarged sub-images of HOPES, CFOG, HOPC.

Figure 20. Checkboard mosaic image and enlarged sub-images of Pair C: (a) checkboard mosaic image of HOPES; (b–f) enlarged sub-images of HOPES, CFOG, HOPC, SAR-SIFT, OS-SIFT.

Figure 21. Checkboard mosaic image and enlarged sub-images of Pair D: (a) checkboard mosaic image of HOPES; (b–e) enlarged sub-images of HOPES, CFOG, HOPC, OS-SIFT.

Figure 22. Checkboard mosaic image and enlarged sub-images of Pair E: (a) checkboard mosaic image of HOPES; (b–f) enlarged sub-images of HOPES, CFOG, HOPC, SAR-SIFT, OS-SIFT.

Figure 23. Checkboard mosaic image and enlarged sub-images of Pair F: (a) checkboard mosaic image of HOPES; (b–f) enlarged sub-images of HOPES, CFOG, HOPC, SAR-SIFT, OS-SIFT.

Figure 24. Checkboard mosaic image and enlarged sub-images of Pair G: (a) checkboard mosaic image of HOPES; (b–d) enlarged sub-images of HOPES, CFOG, HOPC.

To compare the efficiency of the above algorithms, we test them with Pair C, Pair E and Pair F and count the running time of each algorithm, as Table 5. The experiment is conducted on a laptop with an Intel Core i7-10875H processor and 32 GB memory.

Table 5. The comparison of algorithms efficiency.

Pair	Keypoints Number	HOPES	CFOG	HOPC	SAR-SIFT	OS-SIFT
C	268	85 s	54 s	62 s	67 s	53 s
E	537	152 s	94 s	117 s	266 s	235 s
F	753	237 s	136 s	163 s	264 s	249 s
Time per point		0.31 s	0.19 s	0.22 s	-	-

Among the three data sets, the CFOG algorithm has the best running time, which because CFOG computes only single-scale gradient information with relatively low operational complexity, however, it has the lowest co-registration accuracy among the first three algorithms. HOPC has a higher computational complexity due to the phase congruency algorithm. While HOPES adopts a multi-scale fusion algorithm, which increases the computational complexity, but also brings higher co-registration accuracy and success rate. However, the SIFT-like algorithms adopt a global search strategy and have high dimension descriptors, so their computational complexity are the largest. It should be noted that the number of feature points extracted by SIFT-like algorithms is not the same as the first three algorithms, so it is not strictly comparable. We count the average running time for each point of the first three algorithms, there is an approximate linear relationship between the efficiency of the algorithm and the number of keypoints. Extracting the deeper feature information is the key to improving the co-registration accuracy and success rate, but it also increases the computational complexity, which is a tradeoff.

To summarize, the SIFT-like algorithms such as SAR-SIFT and OS-SIFT are more sensitive to nonlinear radiation differences. They can successfully register the images with

obvious boundaries and texture (such as lakes, rivers, etc.), but they are not suitable for significant nonlinear radiation differences caused by radiation distortion. HOPC uses the phase congruency model with illumination and contrast invariance to construct the histogram of phase congruency. As an improved algorithm of HOPC, CFOG uses gradient information to construct the histogram of gradient direction. Experiments show that they are more robust to nonlinear radiation differences than SIFT-like algorithms, but their structural feature extraction methods are not optimized for SAR images. In the practical applications, especially the one-look SAR images, the speckle noise of SAR greatly affects the success rate and accuracy of co-registration. Compared with the above four algorithms, our HOPES adopts a multi-scale primary edge fusion algorithm, constructs more robust and higher SNR primary edge structure features, and retains the edge positioning ability. Therefore, HOPES has better RMSE and NCM, and is more robust to different scenes.

3.4. Comparison of Parameters Settings

In order to compare the effects of σ_{MSG} , γ_{MSG} and N_{Scale} on the HOPES co-registration performance, we choose Pair C for testing. First, the scale factor is set to $k_{MSG} = 1.4$, the number of scales is set to $N_{Scale} = 4$, the orientation is set to $M = 8$, $\gamma_{MSG} = 1, 3, 5, \dots, 15$, $\sigma_{MSG} = 1.4, 1.6, 1.8, 2.0$, the statistics results of RMSE and NCM are shown as Figure 25a,b. When σ_{MSG} and γ_{MSG} are small, both the RMSE and NCM are poor due to the influence of SAR speckle noise. As γ_{MSG} increases, the primary edge fusion function plays a role at this time, which provides a higher edge localization ability and suppresses the speckle noise, so that the RMSE gradually decreases and NCM gradually increases. However, with γ_{MSG} increasing, some edges will be excessively suppressed, resulting in the loss of structural features, and it makes the RMSE increase and NCM decrease. There is a tradeoff in the choice of σ_{MSG} . With a too small choice of σ_{MSG} , the influence of noise cannot be satisfactorily reduced. In contrast, A smaller number of scales results in a larger noise impact, and a larger number of scales will produce a coarser edge structure, both of which are detrimental to co-registration.

Figure 25. Effects of different parameters on co-registration performance. (a) The statistics results of RMSE with different σ_{MSG} . (b) The statistics results of NCM with different σ_{MSG} . (c) The statistics results of RMSE with different γ_{MSG} . (d) The statistics results of NCM with different γ_{MSG} .

4. Conclusions

In this study, we propose a primary structure extraction algorithm to extract the primary edges of SAR images; based on this, we develop a SAR and optical co-registration algorithm called HOPES to overcome the difficulties caused by strong speckle noise and complex NRD.

We design a primary structure extraction algorithm, including Multi-scale Sigmoid Gabor filter, primary edge fusion algorithm and minimum weight strategy, to suppress speckle noise and obtain the primary edge structure information in SAR image, especially the one-look SAR images. The test results of simulated SAR images show that the primary structure extraction algorithm can obtain higher SNR edge strength map and edge positioning accuracy. The proposed co-registration method is composed of NMS-SAR-Harris and deviding grids for keypoints extraction, HOPES structure feature descriptor, NMS-SSD

fast template matching, and FSC outlier removal. NMS-SAR-Harris and deviding grids methods can obtain keypoints with obvious features and uniform keypoints, respectively, HOPES is a histogram of primary structure based on MSG. It is a 3D density structure feature that can reflect the matching region's structure feature. NMS-SSD fast template matching and FSC algorithm further improve the confidence of matching results. Image co-registration experiments show that the proposed co-registration method is robust to speckle noise and NRD between SAR image and optical image; it improves the success rate and accuracy of matching effectively.

Author Contributions: Conceptualization, S.L. and X.L.; methodology, S.L.; software, S.L.; validation, S.L. and X.L.; formal analysis, S.L.; investigation, S.L., X.L., J.R. and J.L.; resources, S.L., X.L. and J.R.; data curation, S.L., X.L. and J.R.; writing—original draft preparation, S.L.; writing—review and editing, S.L., X.L. and J.L.; visualization, S.L.; supervision, X.L.; project administration, X.L.; funding acquisition, X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the LuTan-1 L-Band Spaceborne Bistatic SAR Data Processing Program, grant number E0H2080702.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data will be available upon request to the corresponding author.

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

References

- Long, H.; Fu, K.; Han, C. An automatic method on detecting image control points from SAR imagery based on Optical Image Patches. In Proceedings of the 2009 Joint Urban Remote Sensing Event, Shanghai, China, 20–22 May 2009; pp. 1–4.
- Bürgmann, T.; Koppe, W.; Schmitt, M. Matching of TerraSAR-X derived ground control points to optical image patches using deep learning. *ISPRS J. Photogramm. Remote Sens.* **2019**, *158*, 241–248. [\[CrossRef\]](#)
- Kulkarni, S.C.; Rege, P.P. Pixel level fusion techniques for SAR and optical images: A review. *Inf. Fusion* **2020**, *59*, 13–29. [\[CrossRef\]](#)
- Wan, L.; Zhang, T.; You, H. Multi-sensor remote sensing image change detection based on sorted histograms. *Int. J. Remote Sens.* **2018**, *39*, 3753–3775. [\[CrossRef\]](#)
- Kai, L.; Xueqing, Z. Review of research on registration of sar and optical remote sensing image based on feature. In Proceedings of the 2018 IEEE 3rd International Conference on Signal and Image Processing (ICSIP), Shenzhen, China, 13–15 July 2018; pp. 111–115.
- 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. [\[CrossRef\]](#)
- Ma, J.; Jiang, X.; Fan, A.; Jiang, J.; Yan, J. Image matching from handcrafted to deep features: A survey. *Int. J. Comput. Vis.* **2021**, *129*, 23–79. [\[CrossRef\]](#)
- Zhao, F.; Huang, Q.; Gao, W. Image matching by normalized cross-correlation. In Proceedings of the 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, Toulouse, France, 14–19 May 2006; Volume 2, p. II.
- Shi, W.; Su, F.; Wang, R.; Fan, J. A visual circle based image registration algorithm for optical and SAR imagery. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 2109–2112.
- Maes, F.; Collignon, A.; Vandermeulen, D.; Marchal, G.; Suetens, P. Multimodality image registration by maximization of mutual information. *IEEE Trans. Med Imaging* **1997**, *16*, 187–198. [\[CrossRef\]](#)
- Shu, L.; Tan, T. SAR and SPOT image registration based on mutual information with contrast measure. In Proceedings of the 2007 IEEE International Conference on Image Processing, San Antonio, TX, USA, 16–19 September 2007; Volume 5, p. V-429.
- Suri, S.; Reinartz, P. Mutual-information-based registration of TerraSAR-X and Ikonos imagery in urban areas. *IEEE Trans. Geosci. Remote Sens.* **2009**, *48*, 939–949. [\[CrossRef\]](#)
- Kuglin, C.D. The phase correlation image alignment method. *Proc. Int. Conf. Cybern. Soc.* **1975**, 163–165.
- Yu, L.; Zhang, D.; Holden, E.J. A fast and fully automatic registration approach based on point features for multi-source remote-sensing images. *Comput. Geosci.* **2008**, *34*, 838–848. [\[CrossRef\]](#)
- Shi, X.; Jiang, J. Automatic registration method for optical remote sensing images with large background variations using line segments. *Remote Sens.* **2016**, *8*, 426. [\[CrossRef\]](#)
- Kim, Y.S.; Lee, J.H.; Ra, J.B. Multi-sensor image registration based on intensity and edge orientation information. *Pattern Recognit.* **2008**, *41*, 3356–3365. [\[CrossRef\]](#)
- Li, H.; Manjunath, B.; Mitra, S.K. A contour-based approach to multisensor image registration. *IEEE Trans. Image Process.* **1995**, *4*, 320–334. [\[CrossRef\]](#)

18. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [\[CrossRef\]](#)
19. Ke, Y.; Sukthankar, R. PCA-SIFT: A more distinctive representation for local image descriptors. In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 27 June–2 July 2004; Volume 2, p. II.
20. Bay, H.; Tuytelaars, T.; Van Gool, L. Surf: Speeded up robust features. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 404–417.
21. Morel, J.M.; Yu, G. ASIFT: A new framework for fully affine invariant image comparison. *SIAM J. Imaging Sci.* **2009**, *2*, 438–469. [\[CrossRef\]](#)
22. Dellinger, F.; Delon, J.; Gousseau, Y.; Michel, J.; Tupin, F. SAR-SIFT: A SIFT-like algorithm for SAR images. *IEEE Trans. Geosci. Remote Sens.* **2014**, *53*, 453–466. [\[CrossRef\]](#)
23. 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. [\[CrossRef\]](#)
24. Yu, Q.; Ni, D.; Jiang, Y.; Yan, Y.; An, J.; Sun, T. Universal SAR and optical image registration via a novel SIFT framework based on nonlinear diffusion and a polar spatial-frequency descriptor. *ISPRS J. Photogramm. Remote Sens.* **2021**, *171*, 1–17. [\[CrossRef\]](#)
25. 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. [\[CrossRef\]](#)
26. Ye, Y.; Shen, L. Hopc: A novel similarity metric based on geometric structural properties for multi-modal remote sensing image matching. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *3*, 9. [\[CrossRef\]](#)
27. Ye, Y.; Shan, J.; Hao, S.; Bruzzone, L.; Qin, Y. A local phase based invariant feature for remote sensing image matching. *ISPRS J. Photogramm. Remote Sens.* **2018**, *142*, 205–221. [\[CrossRef\]](#)
28. Xiang, Y.; Tao, R.; Wan, L.; Wang, F.; You, H. OS-PC: Combining feature representation and 3D phase correlation for subpixel optical and SAR image registration. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 6451–6466. [\[CrossRef\]](#)
29. Xiong, X.; Xu, Q.; Jin, G.; Zhang, H.; Gao, X. Rank-based local self-similarity descriptor for optical-to-SAR image matching. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 1742–1746. [\[CrossRef\]](#)
30. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
31. He, H.; Chen, M.; Chen, T.; Li, D. Matching of remote sensing images with complex background variations via Siamese convolutional neural network. *Remote Sens.* **2018**, *10*, 355. [\[CrossRef\]](#)
32. Hughes, L.; Schmitt, M. A semi-supervised approach to sar-optical image matching. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *4*, 1–8. [\[CrossRef\]](#)
33. Zhang, H.; Ni, W.; Yan, W.; Xiang, D.; Wu, J.; Yang, X.; 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. [\[CrossRef\]](#)
34. Zhang, H.; Lei, L.; Ni, W.; Tang, T.; Wu, J.; Xiang, D.; Kuang, G. Optical and SAR Image Matching Using Pixelwise Deep Dense Features. *IEEE Geosci. Remote Sens. Lett.* **2020**, *19*, 6000705. [\[CrossRef\]](#)
35. 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.
36. Hughes, L.H.; Schmitt, M.; Zhu, X.X. Mining hard negative samples for SAR-optical image matching using generative adversarial networks. *Remote Sens.* **2018**, *10*, 1552. [\[CrossRef\]](#)
37. Cui, S.; Ma, A.; Zhang, L.; Xu, M.; Zhong, Y. MAP-Net: SAR and Optical Image Matching via Image-Based Convolutional Network With Attention Mechanism and Spatial Pyramid Aggregated Pooling. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1000513. [\[CrossRef\]](#)
38. Fjortoft, R.; Lopes, A.; Marthon, P.; Cubero-Castan, E. An optimal multiedge detector for SAR image segmentation. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 793–802. [\[CrossRef\]](#)
39. Wei, Q.R.; Wang, Y.K.; Xie, P.Y. Sar Edge Detector with High Localization Accuracy. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 3744–3747.
40. Wei, Q.; Feng, D.; Jia, W. UDR: An Approximate Unbiased Difference-Ratio Edge Detector for SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 6688–6705. [\[CrossRef\]](#)
41. 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.* **2014**, *12*, 43–47. [\[CrossRef\]](#)
42. Mehrotra, R.; Namuduri, K.R.; Ranganathan, N. Gabor filter-based edge detection. *Pattern Recognit.* **1992**, *25*, 1479–1494. [\[CrossRef\]](#)
43. 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. [\[CrossRef\]](#)
44. Wang, L.; Sun, M.; Liu, J.; Cao, L.; Ma, G. A Robust Algorithm Based on Phase Congruency for Optical and SAR Image Registration in Suburban Areas. *Remote Sens.* **2020**, *12*, 3339. [\[CrossRef\]](#)