



Article

LPHOG: A Line Feature and Point Feature Combined Rotation Invariant Method for Heterologous Image Registration

Jianmeng He ^{1,2}, Xin Jiang ^{1,*} , Zhicheng Hao ¹, Ming Zhu ¹, Wen Gao ³ and Shi Liu ^{1,2}

¹ Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China; hejianmeng21@mails.ucas.ac.cn (J.H.); haozhicheng@ciomp.ac.cn (Z.H.); zhuming@ciomp.ac.cn (M.Z.); liushi211@mails.ucas.ac.cn (S.L.)

² University of Chinese Academy of Sciences, Beijing 100049, China

³ BYD Auto Industry Company Ltd., Shenzhen 518118, China; gaowen@ciomp.ac.cn

* Correspondence: jiangxin@ciomp.ac.cn; Tel.: +86-139-4485-1852

Abstract: Remote sensing image registration has been a very important research topic, especially the registration of heterologous images. In the research of the past few years, numerous registration algorithms for heterogenic images have been developed, especially feature-based matching algorithms, such as point feature-based or line feature-based matching methods. However, there are few matching algorithms that combine line and point features. Therefore, this study proposes a matching algorithm that combines line features and point features while achieving good rotation invariance. It comprises LSD detection of line features, keypoint extraction, and HOG-like feature descriptor construction. The matching performance is compared with state-of-the-art matching algorithms on three heterogeneous image datasets (optical–SAR dataset, optical–infrared dataset, and optical–optical dataset), verifying our method’s rotational invariance by rotating images in each dataset. Finally, the experimental results show that our algorithm outperforms the state-of-the-art algorithms in terms of matching performance while possessing very good rotation invariance.

Keywords: heterologous image matching; line features; point features; LSD algorithm; HOG; FSC algorithm



Citation: He, J.; Jiang, X.; Hao, Z.; Zhu, M.; Gao, W.; Liu, S. LPHOG: A Line Feature and Point Feature Combined Rotation Invariant Method for Heterologous Image Registration. *Remote Sens.* **2023**, *15*, 4548. <https://doi.org/10.3390/rs15184548>

Academic Editor: Dusan Gleich

Received: 15 August 2023

Revised: 5 September 2023

Accepted: 13 September 2023

Published: 15 September 2023



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

1. Introduction

Remote sensing images, with their advantages of convenience and intuition, are becoming one of the effective means for people to observe, describe and analyze the characteristics of the Earth’s surface. At the same time, the quality of remote sensing images has been improving due to the maturity of sensor imaging technology, and the types of images are becoming more and more diversified, gradually developing in the direction of multi-modal, multi-spectral, multi-resolution and multi-temporal features [1]. Different types of images contain different information, and the use of multi-source remote sensing images together to realize information complementarity has become one of the hot topics in recent years. The prerequisite for the simultaneous use of heterogeneous images is the registration of heterogeneous images.

Over the years, many scholars in different countries have published review papers by summarizing the existing registration techniques. Zitová et al., in a review in 2003, outlined the basic steps of image registration and classified the registration techniques in detail [2]. There are five basic steps in image registration: feature detection, feature matching, mapping function design, image transformation and resampling [2]. The image registration scenarios can be classified according to the method of image acquisition: different angle image registration, different time phase image registration, different modal image registration, etc. As for the registration techniques, they are subdivided into grayscale-based registration methods, feature-based registration methods, and transformation domain-based registration methods. Among them, feature-based registration methods are in the mainstream. The

grayscale-based image registration algorithms count the grayscale features of multimodal images, calculate the similarity of grayscale features, and complete the image registration work. Grayscale-based matching algorithms suffer from numerous limitations, especially for heterogeneous images with large differences in grayscale values. Feature-based heterogeneous image registration relies on the point features [3], line features [4] and plane features [5] of the image, calculates the similarity of the features of the two images and completes the feature matching [6–8]. Li et al. summarized the difficulties and modalities of feature-based visible image and SAR remote sensing image registration [9]. Due to the imaging characteristics of synthetic aperture radar, SAR images often have speckle noise, which affects the feature extraction in the registration process. Meanwhile, due to different radiation characteristics, the same object will demonstrate different gray values in visible and SAR images. In addition, due to the characteristics of side-view imaging of synthetic aperture radar, SAR images will be superimposed on the mask, shrinkage and other characteristics, which further increases the difficulty of registration. Li Kai summarized the classical operators and algorithms commonly used in the process of optical and SAR image registration [9]. Based on point features: Moravec [10], Harris [11], SUSAN [12], SIFT [13], SURF [14]. Based on line features: ROA [15], registration based on chain code and line features, method combining line features and histogram, method based on adaptive algorithm and line features. Based on plane features: MRF [16], level set based methods [17], multi-scale registration.

Moreover, there are many methods based on deep learning, such as Mu-net [18], PCNet [19], RFNet [20] and Fourier-Net [21]. In order to adapt various types of multimodal images, Mu-Net uses the structural similarity to design a loss function that allows Mu-net to achieve comprehensive and accurate registration [18]. With the help of phase congruency, PCNet enhances the similarity of multimodal images, thus improving the performance of registration [19]. RFNet combines image registration and image fusion and improves the performance of fine registration via the feedback of image fusion [20]. In order to conserve resources and improve speed, Fourier-Net uses a parameter-free model-driven decoder to replace expansive paths in a U-net style network [21].

In recent years, the state-of-the-art heterogeneous image matching algorithms based on point features have included PSO-SIFT [22], OS-SIFT [23], RIFT [24] and LNIFT [25]. PSO-SIFT overcomes the problem of intensity differences in heterogeneous remote sensing images by introducing a new gradient definition, and then finely corrects the matching performance through positional differences, scale differences, and dominant orientation differences between pairs of feature points in combination with the results of the initial matching. PSO-SIFT abandons the strategy of SIFT to use the gradient obtained by subtracting the gradients of the two neighboring pixels in the Gaussian scale space of the image, instead using the Sobel operator to compute the gradient of keypoints in the Gaussian scale space of the image, which in turn optimizes the computation of the dominant orientation of the feature points. Meanwhile, PSO-SIFT adopts a circular neighborhood with radius 12σ and 17 bins to determine the log-polar sectors in the dominant orientation of the keypoints to construct the GLOH-like [26] feature descriptor, instead of using the original 4×4 square sector SIFT descriptor. In the matching process, PSO-SIFT firstly obtains the initial matching results with the nearest neighbor ratio algorithm, then optimizes the distance calculation method for feature point pairs by combining the positional difference, scale difference, and dominant orientation difference of the initial matched keypoints, and finally rejects the wrong matched point pairs with the FSC algorithm [27]. The experimental results of PSO-SIFT show that PSO-SIFT outperforms SURF [14], SAR-SIFT [28] and MS-SIFT [29] in multi-spectral and multi-sensor remote sensing image matching.

The OS-SIFT algorithm divides optical and SAR image matching into three steps. The first step is to detect the feature points in the Harris scale space of optical and SAR images, respectively, the second step is to determine the orientation of the feature points and construct the feature descriptors, and the third step is keypoint matching. The idea of Harris scale space is derived from DOG scale space. In DOG space scale space, keypoints are

extracted by finding local maxima. In Harris scale space, the corner points are detected with the multiscale Harris function. The multiscale Harris function can be derived by replacing the first derivative of DOG scale space with the multiscale gradient computations [23]. In order to overcome the performance impact of the significant difference between optical and SAR images on the repetition rate of keypoint detection, the OS-SIFT algorithm adopts different gradient calculation methods for optical and SAR images when constructing the Harris scale space. The Sobel operator is used to calculate the image gradient for optical images, and the ROWEA [30] operator is used to calculate the image gradient for SAR images. After detecting the keypoints in the Harris scale space, the position of the keypoints is also finely corrected by the least squares method. Similarly to PSO-SIFT, OS-SIFT also uses a circular neighborhood with a radius of 12σ (σ is the parameter of the first scale, $\sigma = 2$ in OS-SIFT) and 17 bins to determine the dominant orientation of the keypoints in the log-polar sectors to construct the GLOH-like keypoint feature descriptors, instead of using the original 4×4 square sector SIFT descriptor. Finally, OS-SIFT matches point pairs by the nearest neighbor ratio and eliminates false matching point pairs by FSC [27]. The OS-SIFT experimental results show that it outperforms SIFT-M [31] and PSO-SIFT algorithms in terms of matching accuracy.

In order to solve the problem of intensity and gradient sensitivity to nonlinear radiation differences in the process of feature detection and descriptor construction, RIFT proposes a feature detection method based on phase consistency and a maximum index map to construct the descriptor. The experimental results of RIFT show that in terms of matching performance, it is better than the algorithms of SAR-SIFT [28], LSS [32], PIIFD [33] and others.

LNIFT proposes an algorithm to improve the matching performance by reducing the nonlinear radiometric differences of heterogeneous images. LNIFT firstly employs a mean filter to filter the multimodal images to obtain normalized image pairs to reduce the modal differences between the images. Then, feature points are detected on the normalized images according to the improved ORB [34], and HOG [35] feature descriptors are constructed on the normalized images to enhance the rotational invariance of matching. The experimental results show that LNIFT achieves better matching results than SIFT, PSO-SIFT, OS-SIFT and RIFT on multiple multimodal image datasets.

Heterogeneous image matching algorithms based on line features in recent years mainly combined some control points of line features and line features. Meng et al. used Gaussian gamma-type double window (GCS) [36] and LSD [37] linear feature detection, and then extracted the control points as the to-be-matched points to achieve the matching between optical and SAR images [38]. Sui et al. used different line feature extraction methods for optical images and SAR images [39]. For optical images, line features are extracted directly using LSD detector, while for SAR images, some pretreatment will be performed first. The Lee filter [40] is used to reduce the effect of scattering noise on a SAR image, then the edges of the SAR image are detected by using Gaussian gamma-shaped (GGS) bi-windows. The line features of the SAR image are obtained by Hough transform [41]. In order to improve the matching accuracy, the line features are extracted on the low-resolution image. Then, the transform relationship between the intersections of the line segments is calculated, which guides the conjugate line segment selection for fine matching. Finally, the images are matched according to the spatial consistency.

Although there have been many matching algorithms for heterogenous images in the past few years, including those based on point features and line features, these methods still have many limitations. These limitations include the following:

- Feature detection is too confined to point features or line features, which cannot well combine the advantages of point features and line features. This leads to limitations in feature detection.
- The step of extracting keypoints in the line feature-based matching algorithm is too complicated, which prevents the advantages of line features from being fully utilized.

- Constructing the dominant orientation of the keypoint still relies too much on the intensity and gradient of the local image patch around the keypoint, which will lead to uncontrollable differences in the dominant orientation between two points in the reference image and registration image. Because there are nonlinear differences between the intensity and gradient of heterogeneous images, it will also reduce the registration performance and the rotation invariance of the matching.

In this paper, we address the above limitations by proposing a rotation-invariant matching method based on the combination of line features and point features for heterogeneous images. This proposed method mainly consists of the following two approaches:

First, we use the LSD algorithm to extract the line features of the heterogeneous image, and then extract the points on the straight line segment as keypoints. When extracting the feature points on the line features, we first compare the gradient magnitude of multiple points in the vertical direction, perpendicular to the feature line segment at every position, and select the point with the largest gradient magnitude as the real feature point.

Second, in order to improve the rotation invariance, we no longer determine the dominant orientation of the feature point based on the intensity or gradient of the local image patch while constructing feature descriptor. We directly assign the tilt angle of the straight line segment to the orientation of the keypoints according to which line they are extracted from. At the same time, we rotate the image according to the tilt angle and center of the given straight line segment, and construct HOG-like feature descriptors on the obtained rotated image.

The rest of the paper is organized as follows: Section 2 provides a detailed description of all steps in the methodology. Section 3 describes the datasets used for the experiments. In Section 4, the matching performance and rotation invariance of each dataset are evaluated and discussed qualitatively and quantitatively in turn. In Section 5, future research directions are to be discussed. Finally, the conclusions are presented in Section 6.

2. Methodology

In this section, we describe the proposed LPHOG method in detail. Figure 1 displays the framework of our LPHOG.

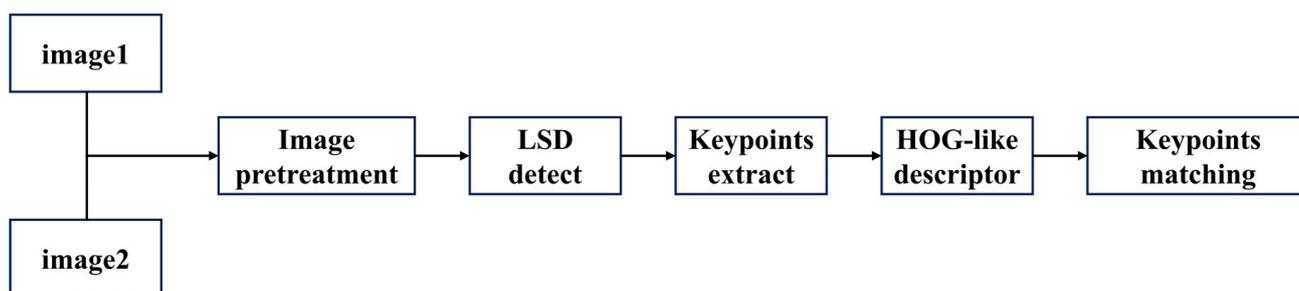


Figure 1. The framework of our LPHOG method.

2.1. Image Pretreatment

In this section on image preprocessing, we used different pretreatment methods for the images based on the characteristics of the three different datasets (optical–SAR dataset, optical–infrared dataset, optical–optical dataset), but all of these preprocessing methods are very simple.

2.1.1. Image Pretreatment of Optical-SAR Dataset

For the optical and SAR image datasets, we only preprocessed the SAR images. Figure 2 illustrates the pretreatment process for the SAR image. In order to enhance

the brightness of the SAR images, we applied logarithmic enhancement. The logarithmic enhancement function definition is,

$$I_{enhance_sar}(x, y) = \log(I_{sar}(x, y) + 1) \quad (1)$$

where $I_{enhance_sar}$ and I_{sar} represent the enhanced SAR image and original SAR image, respectively; (x, y) represents a 2D image coordinate.



Figure 2. The pretreatment process for SAR images.

However, the logarithmic function also enhances the speckle noise on the SAR image. In order to reduce the effect of speckle noise, we applied Gaussian filtering to smooth the SAR image on the enhanced SAR image $I_{enhance_sar}(x, y)$. The edges of the smoothed SAR image are not obvious, which will reduce the performance of LSD line feature detection. Therefore, we used bilateral filtering for the SAR image. We filtered the enhanced image after passing it through the logarithmic enhancement using a Gaussian filter with a standard deviation of $2\sqrt{2}$ in both the horizontal and vertical directions and a Gaussian kernel size of 17. We then filtered the Gaussian filtered image using a bilateral filter where the diameter of each pixel neighborhood used during filtering was set to 9. The filtering sigma was 100 in color space and 100 in coordinate space.

Figure 3 shows the original SAR image and the SAR image after pretreatment, respectively. Figure 4 shows the LSD detection results for the optical image, the original SAR image and the SAR image after pretreatment, respectively. Compared with the original SAR image, it can be seen that the preprocessed SAR image becomes smooth in the original region that is seriously impacted by speckle noise, which reduces the chance of detecting pseudo straight line features. At the same time, as with the line features of the optical image, the line features become more elongated.

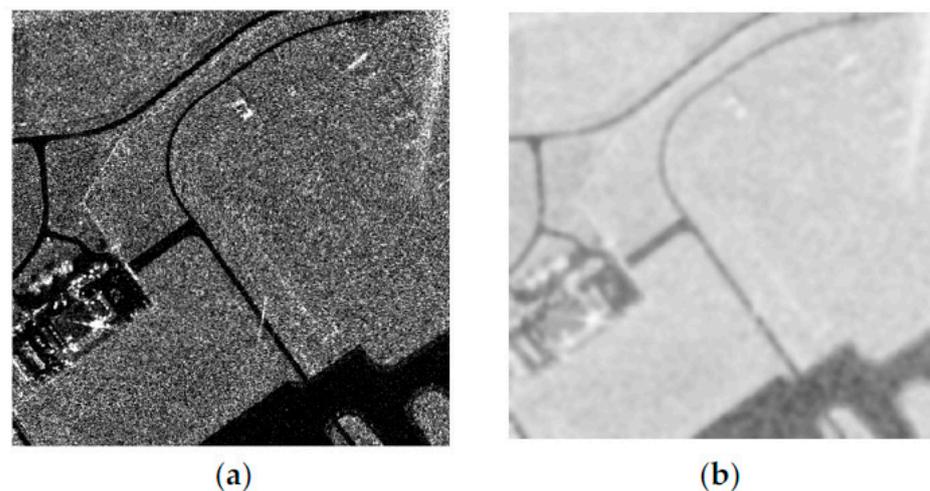


Figure 3. The change in a SAR image. (a) The original SAR image; (b) the SAR image after pretreatment.

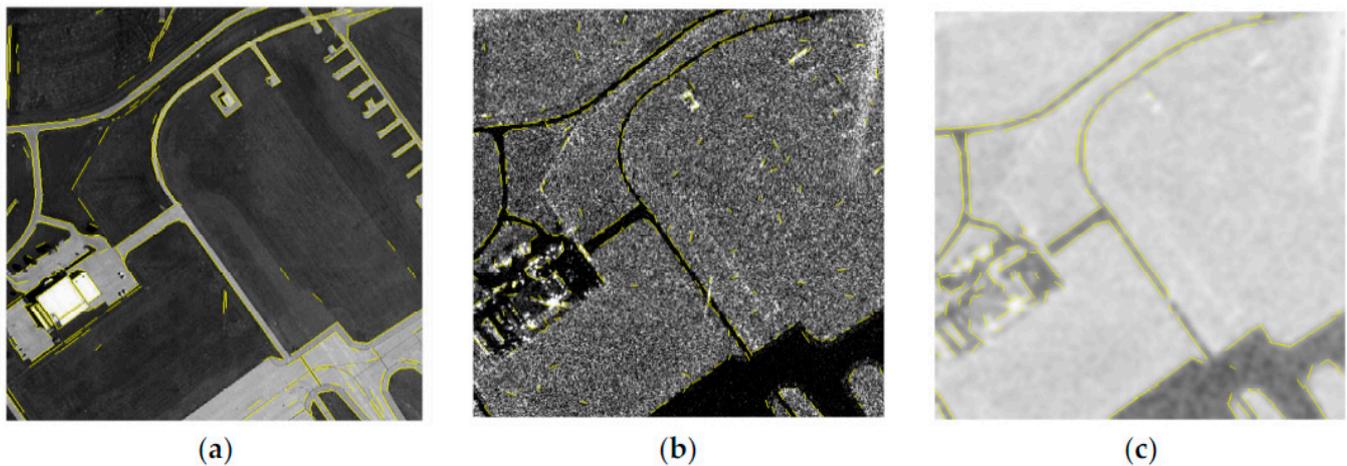


Figure 4. The results of LSD detection. (a) The results for the optical image; (b) the results for the original SAR image; (c) the results for the SAR image after pretreatment. The yellow lines represent line features.

2.1.2. Image Pretreatment of Optical-Infrared Dataset

For the optical and infrared image datasets, we did not preprocess the images but performed LSD straight line detection directly. Figure 5 shows the LSD detection results for the optical–infrared dataset.

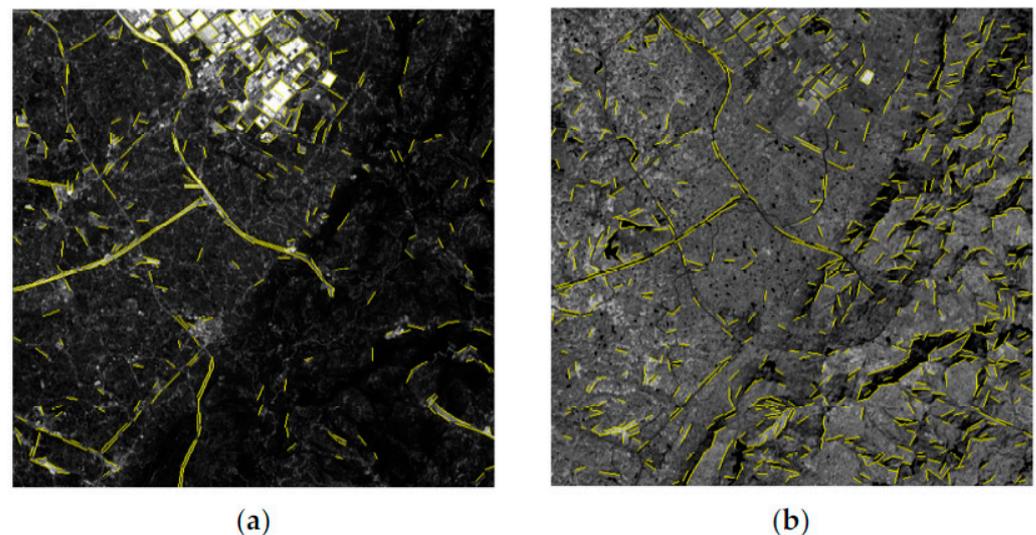


Figure 5. The results of LSD detection. (a) The results for the optical image; (b) the results for the infrared image. The yellow lines represent line features.

2.1.3. Image Pretreatment of Optical-Optical Dataset

For the optical–optical image dataset, we filtered the original images using bilateral filtering for each pair of optical images. Figure 6 displays the pretreatment process for optical images. We filtered each optical image using a bilateral filter where the diameter of each pixel neighborhood used during filtering was set to 5, filtering sigma to 150 in color space and 150 in coordinate space. Figure 7 shows the LSD detection results for a pair of images in the optical–optical dataset before and after pretreatment, respectively. From the red rectangle area, it can be seen that relative to the original image, the pretreated optical image reveals a greater number of line features. Since the imaging mechanism for the optical images was the same, while the two images correspond to the same region, more line features can imply more matching point pairs, which can improve the matching performance.



Figure 6. The pretreatment process for optical images in the optical–optical dataset.

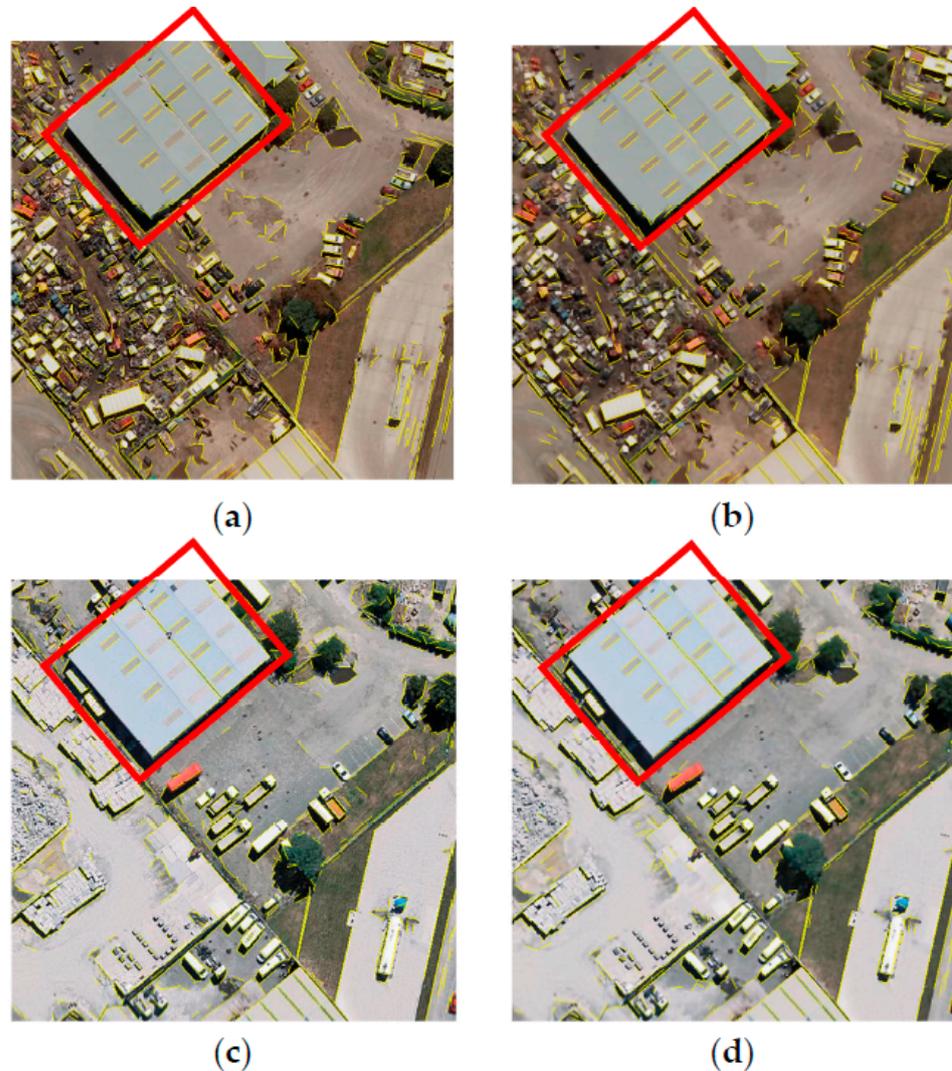


Figure 7. The results of LSD detection. (a) The results for the original reference image; (b) the results for the reference image after pretreatment; (c) the results for the original registration image; (d) the results for the registration image after pretreatment.

2.2. Review of LSD Algorithm and LSD Detection

LSD line detection uses a rectangle to approximate a line segment. LSD linear detection consists of the following main steps [37,42,43]:

- Perform Gaussian down-sampling of the input image at scale S . Then, calculate the gradient magnitude and direction of each pixel of the down-sampled image. LSD calculates the gradient of a given point (x, y) from four points below the point, mathematically defined as

$$g_x(x, y) = \frac{i(x+1, y) + i(x+1, y+1) - i(x, y) - i(x, y+1)}{2} \quad (2)$$

$$g_y(x, y) = \frac{i(x, y + 1) + i(x + 1, y + 1) - i(x, y) - i(x + 1, y)}{2} \quad (3)$$

where $i(x, y)$ is the image grayscale value at point (x, y) .

The gradient magnitude is

$$G(x, y) = \sqrt{g_x^2(x, y) + g_y^2(x, y)} \quad (4)$$

The gradient direction θ is computed as

$$\theta = \arctan\left(-\frac{g_x(x, y)}{g_y(x, y)}\right) \quad (5)$$

- According to the gradient magnitude, all of the points are pseudo-ordered from smallest to largest, and the "NOT USED" identifier is used to initialize the identification of all of the points [37].
- Set the gradient division threshold, and indicate the points with gradient magnitude less than threshold as "USED" [37].
- Adopt the point with the largest gradient amplitude as the seed point, and identify this point as "USED" [37].
- Using the seed point as the starting point, search for "NOT USED" points with gradient direction θ , and then set the searched points to "USED" [37].
- Based on all of the points obtained in the previous step, generate a rectangle R containing all of the points [37].
- Determine whether the density of "USED" points in this rectangle R satisfies the threshold D . If not, divide the rectangle R into multiple rectangular boxes until the threshold D is satisfied [37].
- Calculate the number of false alarm probability points NFA ; change the rectangle R until $NFA \leq \epsilon$, to obtain the linear features. NFA is defined as [37,42]

$$NFA = (NM)^{5/2} \gamma \cdot B(n, k, p) \quad (6)$$

where N and M are the number of columns and rows of the image (after scaling), γ is a total of different values for p tried, n is the total number of pixels in the rectangle, and k is the number of selected "USED" points in step 5. p represents a given precision, and $B(n, k, p)$ is the binomial tail. $B(n, k, p)$ is defined as [42]

$$B(n, k, p) = \sum_{j=k}^n \binom{n}{j} p^j (1-p)^{n-j} \quad (7)$$

Finally, LSD line segments are characterized by a rectangle determined by its center points, angle, length, and width. An LSD line segment is shown in Figure 8 [33,38].

The center point of line segment is (c_x, c_y) .

$$c_x = \frac{\sum_{i \in Region} G(i) \cdot x(i)}{\sum_{i \in Region} G(i)} \quad (8)$$

$$c_y = \frac{\sum_{i \in Region} G(i) \cdot y(i)}{\sum_{i \in Region} G(i)} \quad (9)$$

where $G(i)$ is the gradient magnitude of point i , and the index i runs over the points in the region.

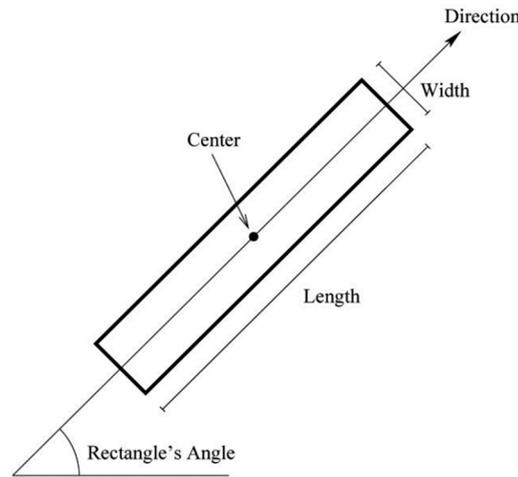


Figure 8. LSD line segment.

The main line segment's angle (rectangle's angle) is set to the angle of the eigenvector associated with the smallest eigenvalue of the matrix M [37]:

$$M = \begin{pmatrix} m^{xx} & m^{xy} \\ m^{xy} & m^{yy} \end{pmatrix} \quad (10)$$

with

$$m^{xx} = \frac{\sum_{i \in \text{Region}} G(i) \cdot (x(i) - c_x)^2}{\sum_{i \in \text{Region}} G(i)} \quad (11)$$

$$m^{yy} = \frac{\sum_{i \in \text{Region}} G(i) \cdot (y(i) - c_y)^2}{\sum_{i \in \text{Region}} G(i)} \quad (12)$$

$$m^{xy} = \frac{\sum_{i \in \text{Region}} G(i) \cdot (x(i) - c_x)(y(i) - c_y)}{\sum_{i \in \text{Region}} G(i)} \quad (13)$$

where $G(i)$ is the gradient magnitude of point i ; $x(i)$ and $y(i)$ represent the x coordinate of point i and y coordinate of point i , respectively.

In the LSD line feature detection for all of the images, we only set the parameter of scale $S = 1.2$, and the rest of the parameters are chosen adaptively by the LSD algorithm.

2.3. Keypoint Extraction

Figure 9 shows the two states before and after image rotation. Assuming that l_1 is a straight line detected by LSD, the center point of this straight line segment is (c_x, c_y) , and the tilt angle of l_1 is θ . We rotate the image using bilinear interpolation to obtain the rotated image and the corresponding horizontally oriented straight line segment l_1' by taking the center point (c_x, c_y) as the center of rotation and the tilt angle θ as the angle of rotation. Figure 10 displays the states of a real image pair and one of line features. The rotation matrix T is simply denoted as

$$\text{Rotated image} = T((c_x, c_y), \theta) \cdot \text{Original image} \quad (14)$$

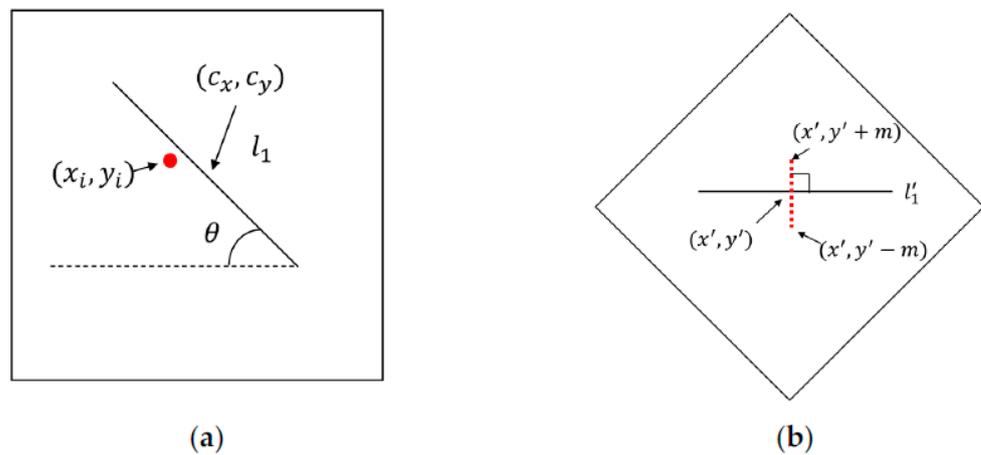


Figure 9. The states of image. (a) The image before rotation; (b) the image after rotation.

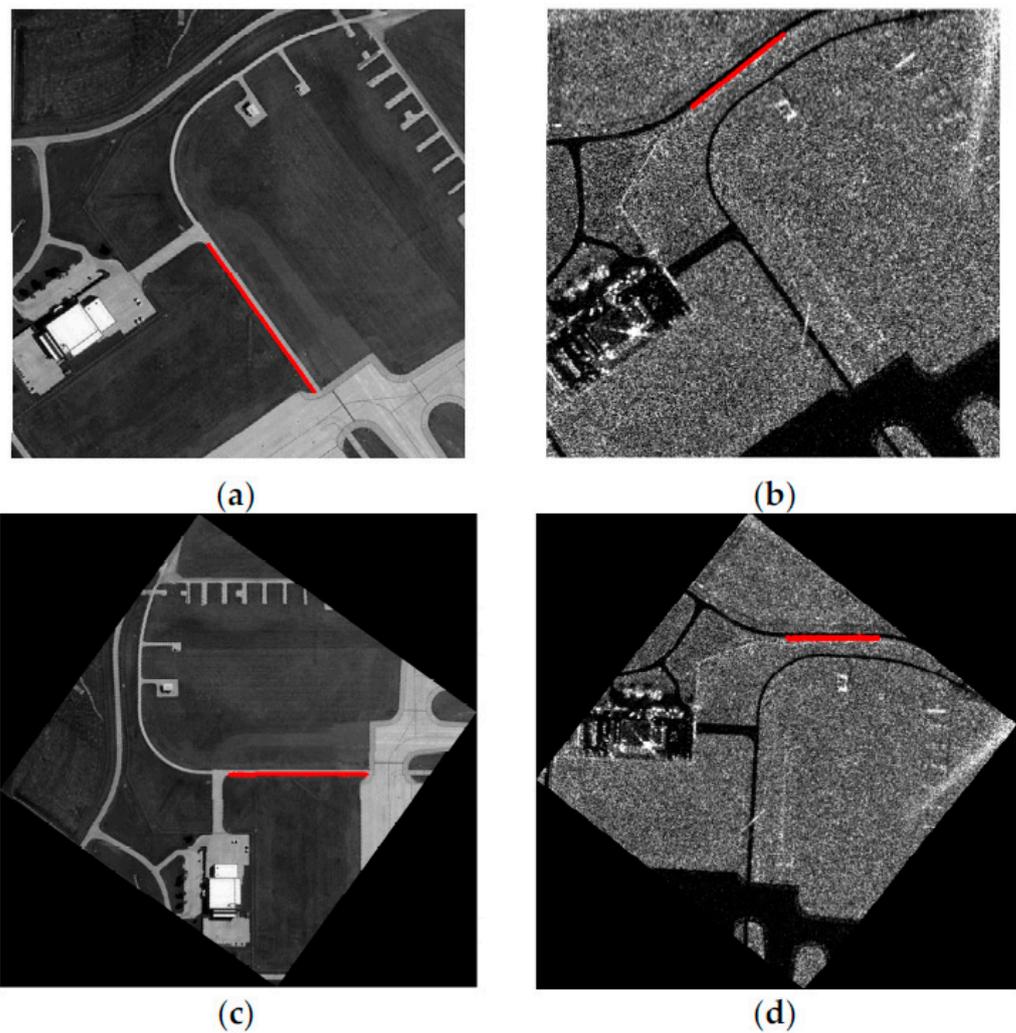


Figure 10. The states of real image. (a) The optical image and one of the line features before rotation; (b) the SAR image and one of the line features before rotation; (c) the optical image and one of the line features after rotation; (d) the SAR image and one of the line features after rotation. The red lines represent line features. We made the original image and rotated image the same size for easier illustration.

As shown in Figure 9, on each side of point (x', y') along the horizontal line l_1' , we select m points on the vertical direction to form a $2 \times m + 1$ point set $\{(x', y'), (x', y' - 1), \dots, (x', y' - m), (x', y' + 1), \dots, (x', y' + m)\}$. Compare the gradient magnitude of these $2 \times m + 1$ points on the rotated image by the Sobel operator, and take the point with the largest gradient as the candidate point $(x', y' \pm k)$.

$$G(x', y' \pm k) = \text{Max}\{G(x', y'), G(x', y' - 1), \dots, G(x', y' - m), G(x', y' + 1), \dots, G(x', y' + m)\} \tag{15}$$

where G represents the gradient magnitude. Calculate by Sobel operator and $k \leq m$.

The point (x_i, y_i) on the original image corresponding to point $(x', y' \pm k)$ is obtained by inverse rotation matrix T^{-1} .

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} = T^{-1} \cdot \begin{pmatrix} x' \\ y' \pm k \end{pmatrix} \tag{16}$$

Assuming that the length of any feature line is s_j , a total of M straight line segments are detected, and then a total of N_F keypoints are extracted.

$$N_F = \sum_{j=0}^M s_j \tag{17}$$

2.4. HOG-Like Descriptor

Figure 11 shows two states of a local image patch before and after rotation. As shown in Figure 11, for a feature point (x_i, y_i) , we rotate the image by taking the center of the straight line segment where the feature point is located in the step of “Keypoints extract” as the rotation center and the tilt angle θ of the straight line segment as the rotation angle. Then, we select a $J \times J$ local image patch centered on the point $(x', y' \pm k)$ to build a HOG-like descriptor for feature point (x_i, y_i) . Figure 12 illustrates the basic shape of the HOG-like feature descriptor. We divide this local image patch into 8×8 grids, where each grid represents 16×16 pixels. In our HOG-like descriptors, we only compute a distribution histogram with 4 bins, and the orientations belong to $[0^\circ, 360^\circ)$. Hence, the length of our HOG-like descriptor is $8 \times 8 \times 4 = 256$. There are three differences between our HOG-like descriptor and the HOG [35]: first, our descriptor is built on the rotated images; second, the grids in our descriptor have no overlaps between each other; third, we only use a 4-bin histogram to encode $[0^\circ, 360^\circ)$ orientations.

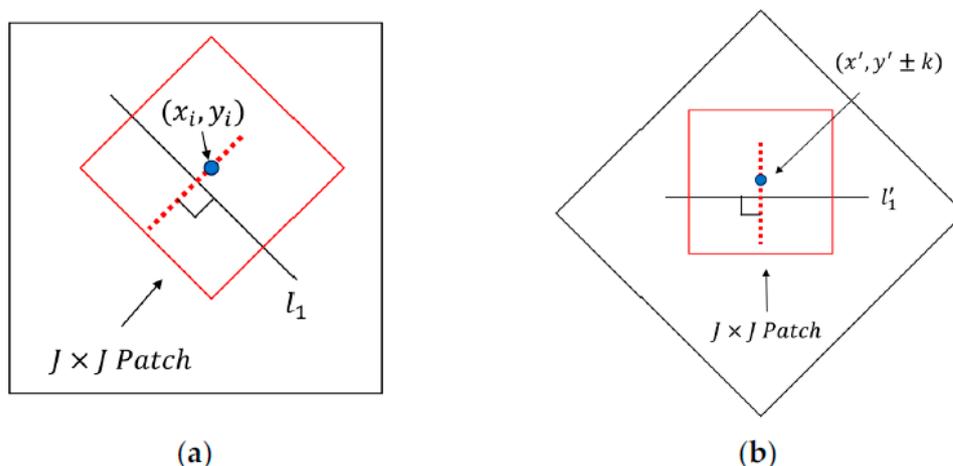


Figure 11. Two states of a local image patch before and after rotation. (a) The local image patch before rotation; (b) the local image patch after rotation.

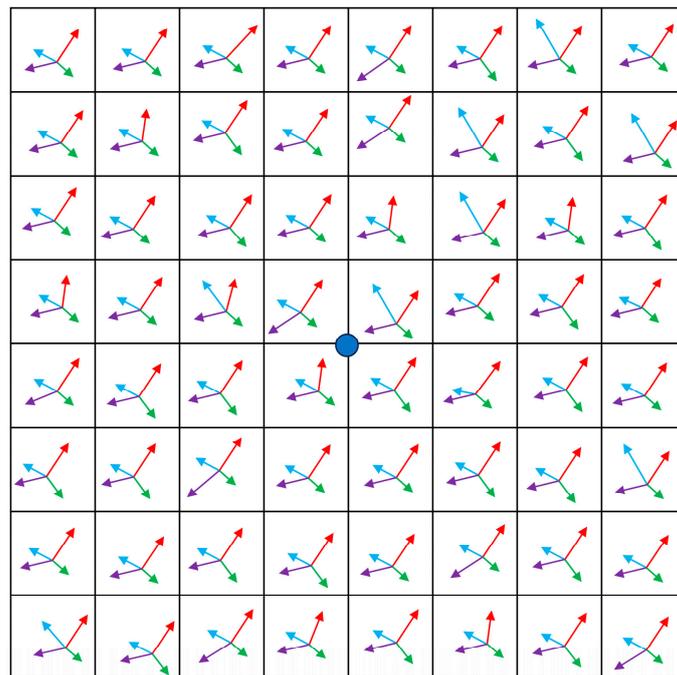


Figure 12. Our HOG-like descriptor, where blue dot represents the keypoint. The local image patch is divided into 8×8 grids. Each grid is encoded into a 4-bin histogram of oriented gradients, where the orientations are normalized into $[0^\circ, 360^\circ)$.

2.5. Keypoint Matching

The feature descriptors are matched by brute force searching, and then, false matching point pairs are eliminated by the FSC algorithm.

3. Datasets

We selected three real heterogeneous image datasets, 500 pairs of images selected for each dataset forming a total of 1500 pairs of images, including an optical–SAR dataset, optical–infrared dataset, and optical–optical dataset. At the same time, we rotate the registration image of each image pair. The rotation angles are from 0° to 360° with an interval of 15° . Thus, a total of 25 registration images of each image pair are obtained (the rotation angles are $[0^\circ, 15^\circ, 30^\circ, \dots, 315^\circ, 330^\circ, 345^\circ, 360^\circ)$) [24]. Overall, with all image pairs rotated in this way, a total of $25 \times 1500 = 37,500$ pairs of images are obtained. The sources and characteristics of each dataset are described in detail below.

3.1. Dataset 1

We selected 500 pairs of images of size 512×512 from OS-SATASET [44] to construct our Dataset 1. OS-SATASET is made by Key Laboratory of Technology in Geo-spatial Information Processing and Application System, Aerospace Information Research Institute, Chinese Academy of Sciences. OS-SATASET contains about 2600 pairs of 512×512 non-overlapping optical and SAR image pairs. The optical images are from Google Earth, the SAR images are from Gaofen-3, and the acquisition method for SAR images is in cluster mode. The resolution of both optical and SAR images is 1 m. Figure 13 shows sample images of six pairs from Dataset 1. We then rotated the SAR images at $[0^\circ, 360^\circ]$ with 15° intervals to obtain 25 optical–SAR image datasets with different rotation angles. Figure 14 shows six pairs of sample images with a 45° rotation angle. Figure 15 shows six pairs of sample images with a 240° rotation angle. (We do not show the sample images for the other rotation angles one by one due to space limitations. Since the background of the rotated SAR images is white, we add a black border to each rotated SAR image for easy viewing. The rotated SAR images are not down-sampled, and we reduced the rotated SAR images to the same size as the optical images just to make them easier to display.)

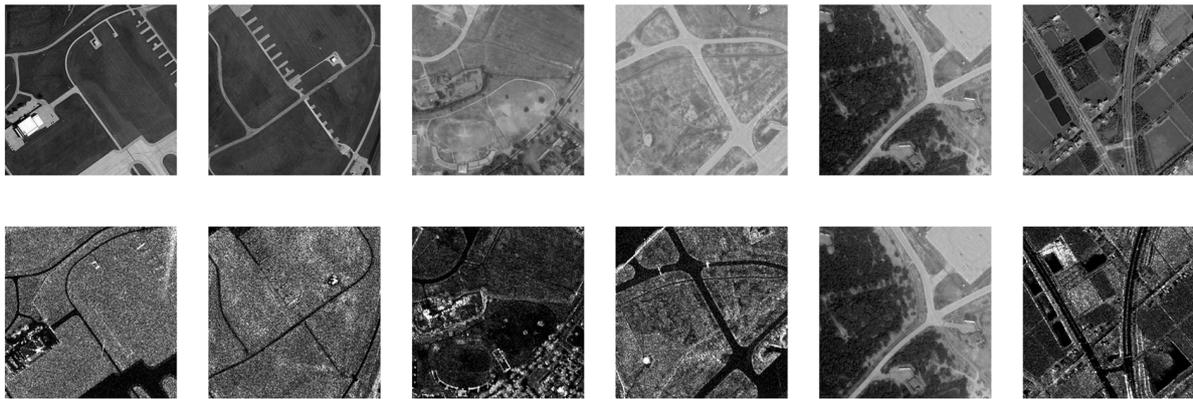


Figure 13. Sample image pairs from Dataset 1 (500 pairs) without rotation.

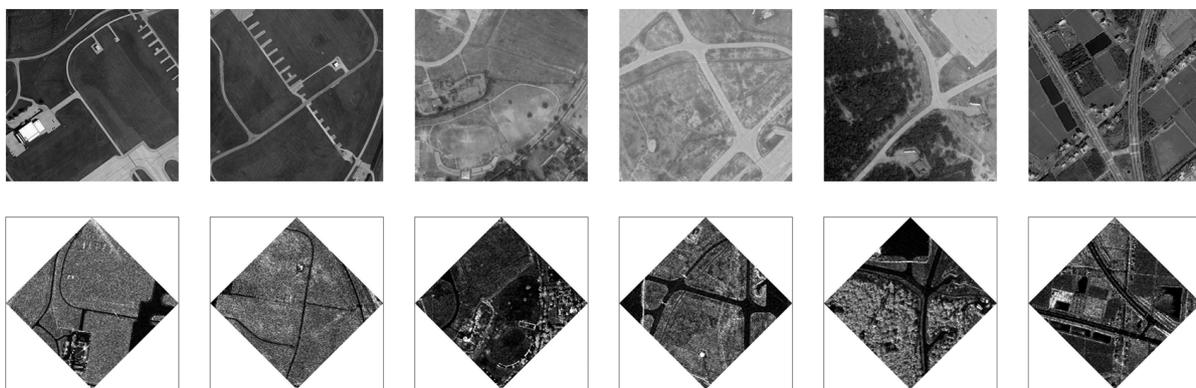


Figure 14. Sample image pairs from Dataset 1 (500 pairs) after 45° rotation.

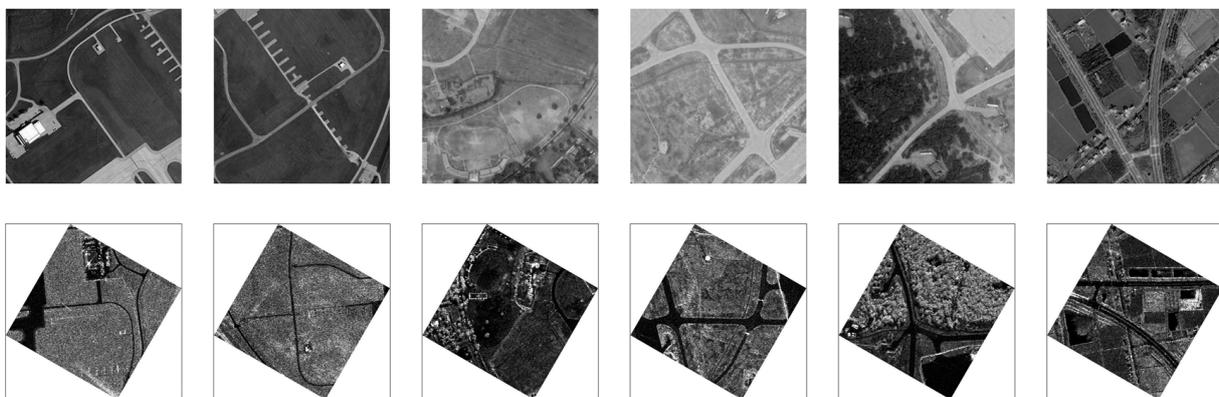


Figure 15. Sample image pairs from Dataset 1 (500 pairs) after 240° rotation.

3.2. Dataset 2

Ye et al. obtained a dataset of 5000 pairs of 512×512 optical–infrared optical image pairs with a resolution of 30 m by cropping multiple Landsat-8 satellite 7600×7800 images [18]. We selected 500 image pairs from these 5000 pairs to form Dataset 2. Figure 16 shows sample images of six pairs from Dataset 2. We then rotated the infrared images at $[0^\circ, 360^\circ]$ with 15° intervals to obtain 25 optical–infrared image datasets with different rotation angles. Figure 17 shows six pairs of sample images with a 45° rotation angle. Figure 18 shows six pairs of sample images with a 240° rotation angle. (We do not show the sample images for the other rotation angles one by one due to space limitations. Since the background of the rotated infrared images is white, we add a black border to each rotated infrared image for easy viewing. The rotated infrared images are not down-sampled, and

we reduced the rotated infrared images to the same size as the optical images just to make them easier to display.)

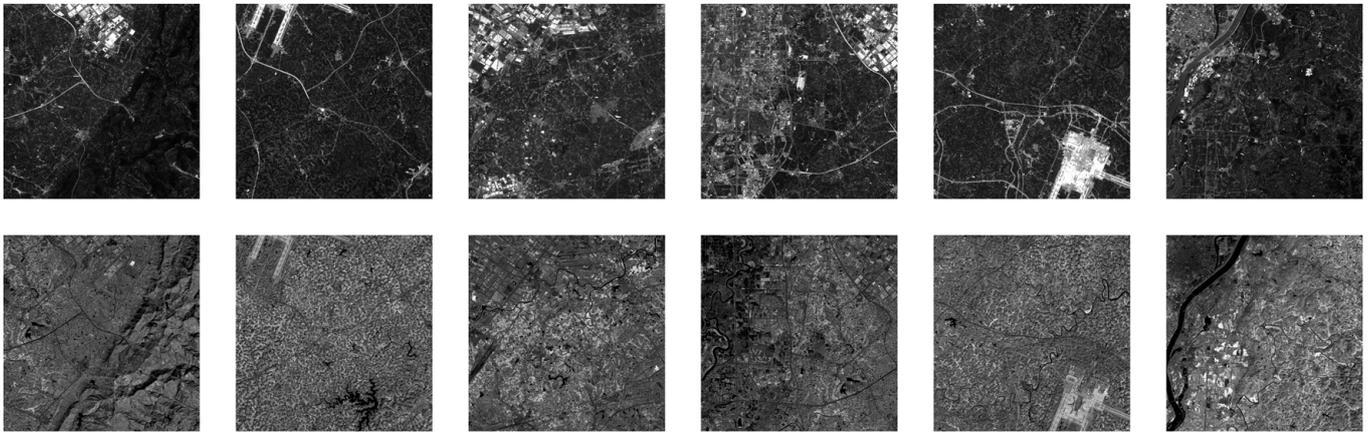


Figure 16. Sample image pairs from Dataset 2 (500 pairs) without rotation.

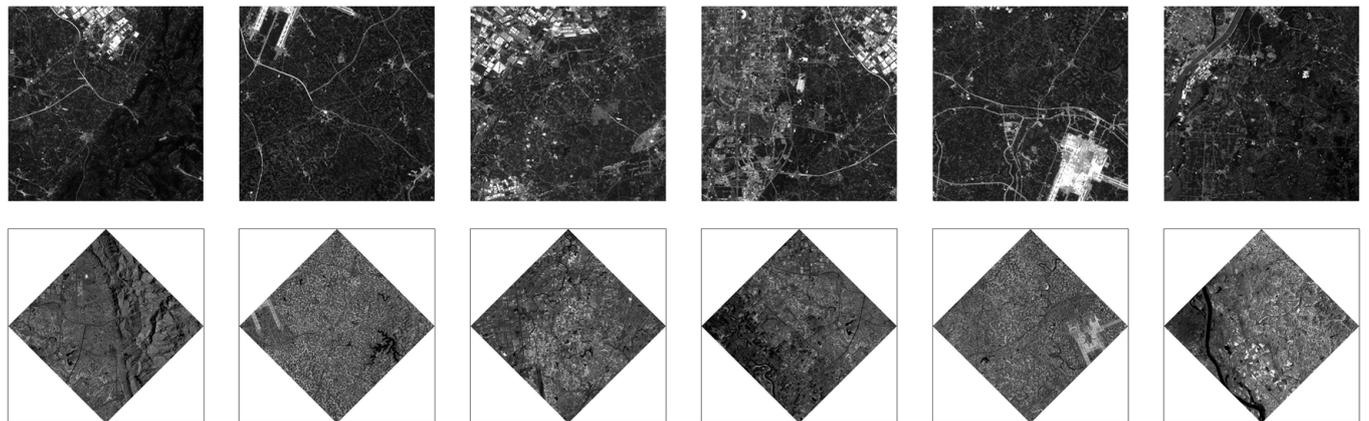


Figure 17. Sample image pairs from Dataset 2 (500 pairs) after 45° rotation.

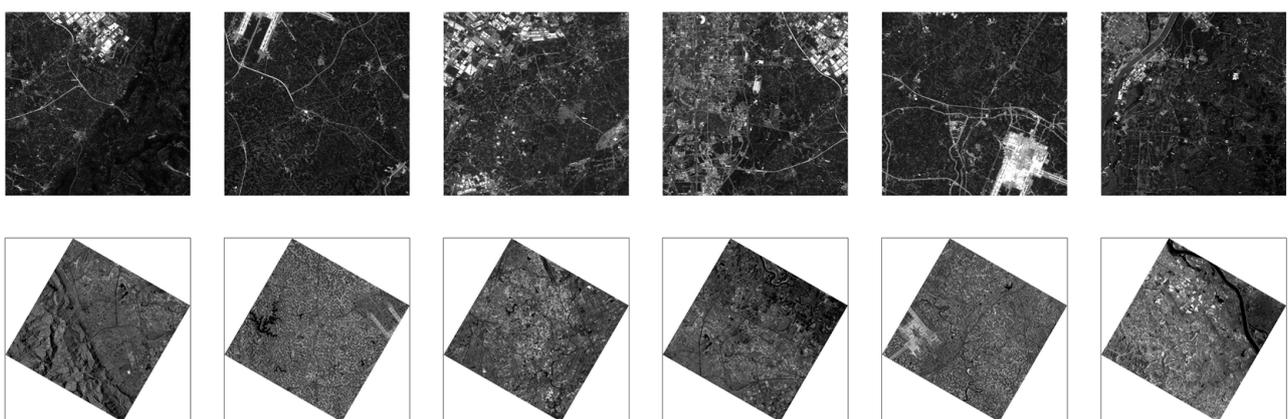


Figure 18. Sample image pairs from Dataset 2 (500 pairs) after 240° rotation.

3.3. Dataset 3

The optical–optical dataset is derived from the WHU building dataset [45]. The WHU building dataset contains two 32,507 × 15,354 aerial images representing buildings and scenery over a 20.5 square kilometer area of Christchurch, New Zealand in 2011 and 2016, respectively. The images' resolutions are both 0.075 m. Ye et al. constructed 5000 pairs of 512 × 512 optical–optical image pairs by cropping and multiple random

affine transformations [18]. From these 5000 pairs of images, we selected 500 pairs with differences in texture, buildings and color to construct our third original dataset to validate the matching effect of our algorithm in the same modality and at different times. Figure 19 shows sample images of six pairs from of Dataset 3. We then rotated the registration images at $[0^\circ, 360^\circ]$ with 15° intervals to obtain 25 optical–optical image datasets with different rotation angles. Figure 20 shows six pairs of sample images with a 45° rotation angle. Figure 21 shows six pairs of sample images with a 240° rotation angle. (We do not show the sample images for the other rotation angles one by one due to space limitations. Since the background of the rotated images is white, we add a black border to each rotated image for easy viewing. The rotated images are not down-sampled, and we reduced the rotated registration images to the same size as the reference images just to make them easier to display.)



Figure 19. Sample image pairs from Dataset 3 (500 pairs) without rotation.



Figure 20. Sample image pairs from Dataset 3 (500 pairs) after 45° rotation.



Figure 21. Sample image pairs from Dataset 3 (500 pairs) after 240° rotation.

4. Results

In this section, we comprehensively evaluate the performance of our method on all of the datasets listed above. Different from traditional methods that use several or dozens of

image pairs for tests, we used 37,500 image pairs for comparisons. The proposed method was compared with three baseline or state-of-the-art methods including PSO-SIFT, RIFT, and LNIFT. For fair comparisons, we used the official implementations of each method provided by the authors. At the same time, the thresholds of keypoint extraction for RIFT, LNIFT, and our LPHOG method were equally 5000 on each image, and we set a contrast threshold for PSO-SIFT to be very small (0.0001) to extract as many as feature points as possible. All experiments were conducted on a PC with AMD Ryzen 7 5800 H CPU at 3.2 GHz and 16 GB of RAM.

We qualitatively evaluated the matching performance by the correct matching images and mosaic grid maps of the sample images, as well as quantitatively evaluating the rotation invariance of the method by the number of correct matches (NCM) of the sample images at every rotation angle. Finally, we statistically assessed the matching performance of the method by the average NCM, $P_{NCM \geq 10}$ (percentage of image pairs with at least 10 correctly matched point pairs), $P_{NCM \geq 100}$ (percentage of image pairs with at least 100 correctly matched point pairs), and the average RMSE to quantitatively evaluate the matching performance of the algorithm, where the higher the NCM, $P_{NCM \geq 10}$ and $P_{NCM \geq 100}$, the better, while the lower the RMSE, the better. If the number of correct matches (NCM) of an image pair is not smaller than 10, this image pair is regarded as correctly matched, since an NCM value that is too small will make the robust estimation technique fail [25]. If one image pair is not correctly matched (i.e., $NCM < 10$), its RMSE is set to be 20 pixels [25].

The definition of $P_{NCM \geq 10}$

$$P_{NCM \geq 10} = \frac{\text{Numbers of image pairs with } NCM \geq 10}{500} \quad (18)$$

The definition of $P_{NCM \geq 100}$

$$P_{NCM \geq 100} = \frac{\text{Numbers of image pairs with } NCM \geq 100}{500} \quad (19)$$

The RMSE is computed as [46]

$$RMSE = \sqrt{\frac{1}{N_{cor}} \sum_{i=1}^{N_{cor}} \left((x_i^o - (x_i^s)')^2 + (y_i^o - (y_i^s)')^2 \right)} \quad (20)$$

where is the number of correctly matched keypoints after the fast sample consensus (FSC), and $((x_i^s)', (y_i^s)')$ denotes the transformed coordinates of (x_i^o, y_i^o) by the estimated transformation matrix H, N_{cor} denotes the number of correct matches.

4.1. Parameter Settings

There is only one parameter left (m in Equation (15)) that affects the performance of our method. We set $m = 2$ in Dataset 1 and Dataset 2 and $m = 4$ in Dataset 3.

4.2. Evaluation of Dataset 1

4.2.1. Qualitative Evaluation of Dataset 1

We selected three pairs of images from the optical–SAR dataset for qualitative comparison. Figures 22–24 show the registration results for Dataset 1 before rotation, after rotation of 45° and after rotation of 240° , respectively. Figure 25 shows the checkerboard mosaic images of our LPHOG method without rotation, after 45° rotation, and after 240° rotation, respectively. As shown in Figure 22, PSO-SIFT could only match the third pair of images, but could not correctly match the first and second pairs of images. After 45° rotation and 240° rotation, PSO-SIFT could not match these three pairs of images correctly, indicating that the overall matching performance of PSO-SIFT was poor and there was almost no rotation invariance. Without rotation, RIFT could match these three pairs of images well, but the matching performance of the first pair and the second pair decreased significantly

after 45° rotation, and the matching performance after 240° rotation improved somewhat compared with the results of 45° rotation, which meant that the matching performance of RIFT was good, but the rotation invariance was not robust. LNIFT could correctly match the first and second pairs of images without rotation, but failed on the third pair of images, and could correctly match only the second pair of images after 45° rotation and 240° rotation, which indicated that the overall matching performance of LNIFT was stronger than that of PSO-SIFT, but weaker than that of RIFT, and the robustness of rotation invariance was not weak. On the whole, the number of matching point pairs obtained with our LPHOG method was significantly higher than that of the above three methods in all three situations, and the number of matching point pairs did not significantly decrease or even increase when the rotation angle changed, which indicates that the matching performance of our method was significantly better than that of PSO-SIFT, RIFT, and LNIFT. At the same time, our method had good rotation invariance. Further, from the checkerboard mosaic images shown in Figure 25, all of our image matching accuracies were high.

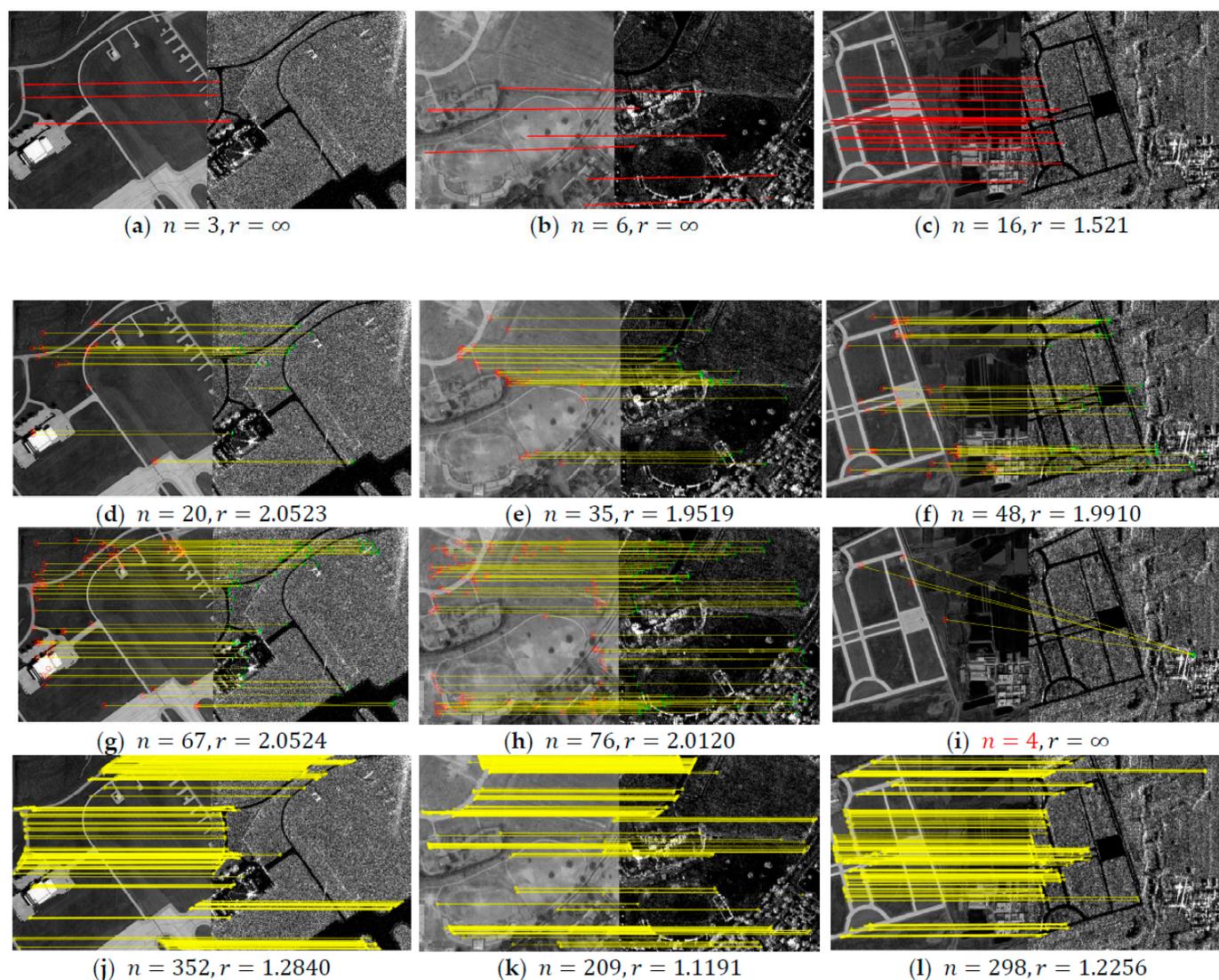


Figure 22. Registration results for Dataset 1 without rotation. (a–c) Registration results for PSO-SIFT; (d–f) registration results for RIFT; (g–i) registration results for LNIFT; (j–l) Registration of our LPHOG. n represents the number of correct match points; r represents RMSE. $r = \infty$ indicates failure to match this image pair.

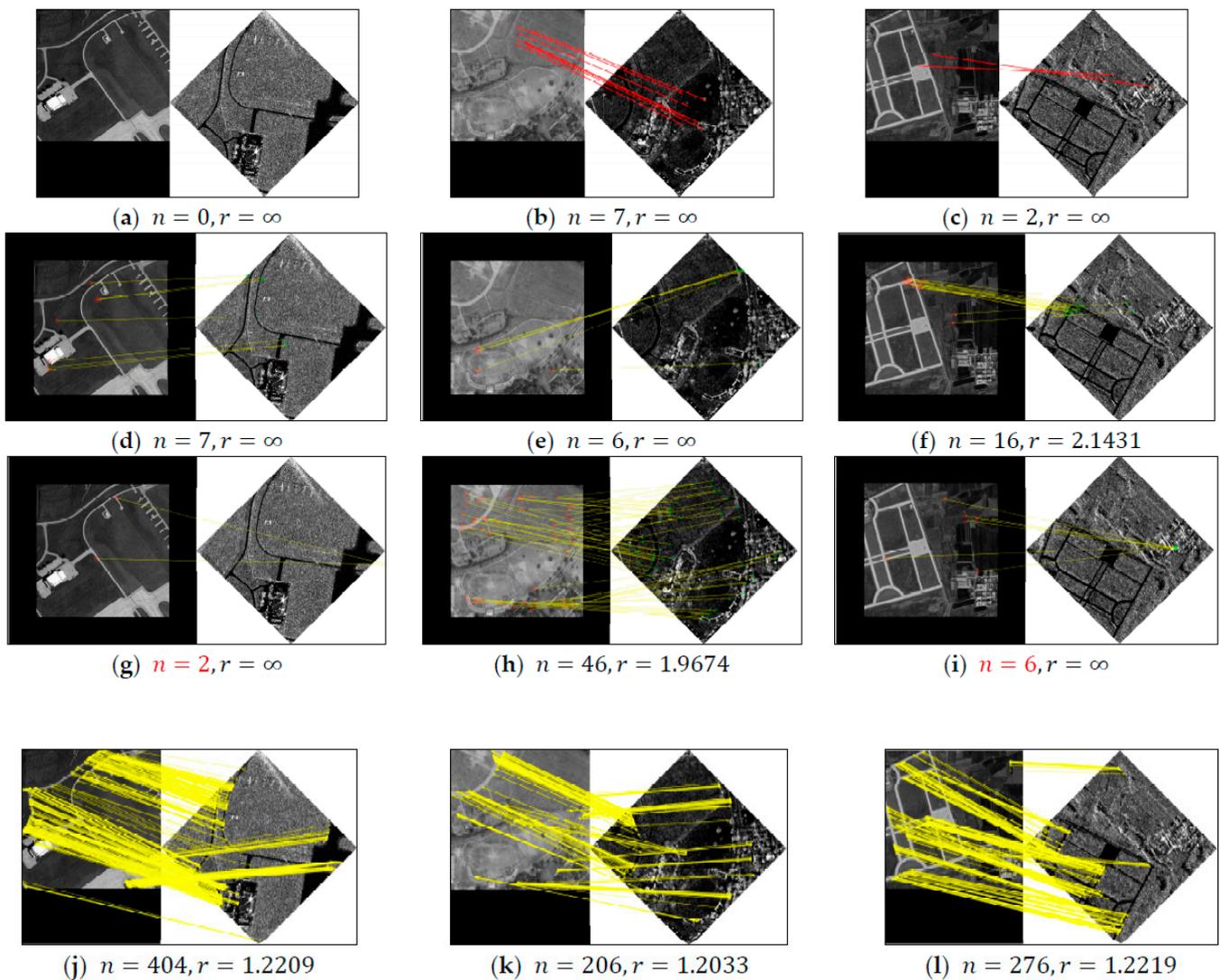


Figure 23. Registration results for Dataset 1 with 45° rotation. (a–c) Registration results for PSO-SIFT; (d–f) registration results for RIFT; (g–i) registration results for LNIFT; (j–l) registration of our LPHOG. n represents the number of correct match points; r represents RMSE. $r = \infty$ indicates failure to match this image pair.

4.2.2. Quantitative Evaluation of Dataset 1

Figure 26 shows the variation in NCM with rotation angles for the first pair of sample images, the second pair of sample images and the third pair of sample images in Dataset 1, respectively. We rotated the registration image of each sample image pair. The rotation angles were from 0° to 360° with an interval of 15° . Thus, a total of 25 registration images of each image pair were obtained (the rotation angles were $[0^\circ, 15^\circ, 30^\circ, \dots, 315^\circ, 330^\circ, 345^\circ, 360^\circ]$) [24]. As a whole, the NCM of our LPHOG method was at least three times higher than that of the other three algorithms (PSO-SIFT, RIFT, LNIFT) in every rotation angle, and the highest NCM was up to 20 times higher. From the first pair, the number of matching pairs for PSO-SIFT stayed close to 0. LNIFT could only achieve more than 50 correct match points at 0° , and the numbers of matching pairs for other rotation angles were close to those obtained with PSO-SIFT. The matching points for RIFT fluctuated, and the fluctuation was relatively small, staying above and below a small number of 20 matching pairs. The number of correctly matched pairs of points in our method stayed above 300 for most of the rotation angles, and only dropped below 250 for 195° . For the second pair, our method maintained over 190 match point pairs for the majority of angles, only dropping to near

160 for the 60° rotation angle. PSO-SIFT, like the first pair, maintained the number of match point pairs near 0. LNIFT worked better than RIFT for the majority of rotation angles at $[0^\circ, 45^\circ]$ and $[270^\circ, 360^\circ]$; but, for rotation angles at $[105^\circ, 240^\circ]$, RIFT was stronger than LNIFT. From the third pair, our method consistently maintained more than 250 correctly matched pairs of points, while the number of correctly matched pairs of points for the other three best-performing RIFTs was consistently less than 1/5 of ours.

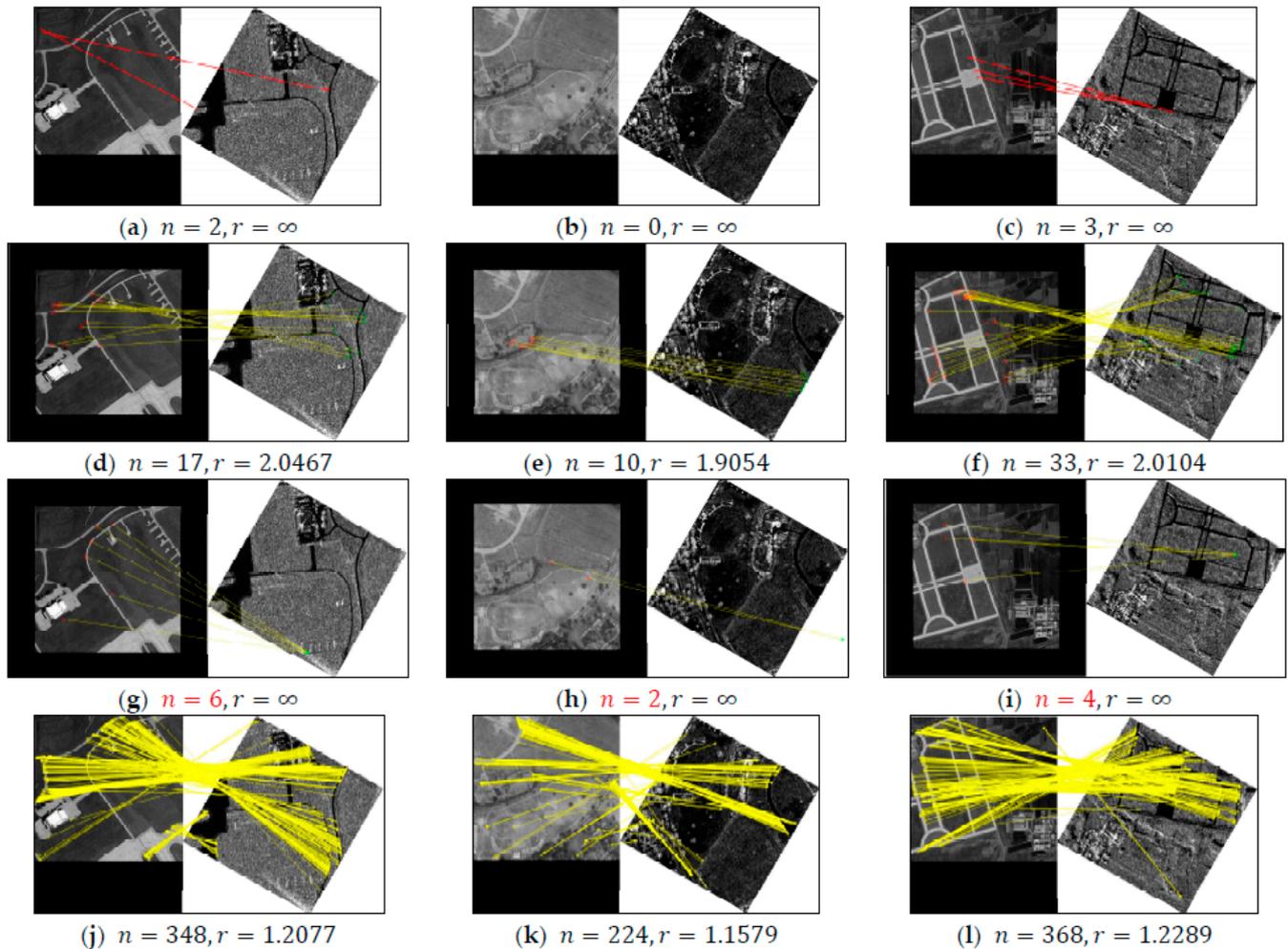


Figure 24. Registration results for Dataset 1 with 240° rotation. (a–c) Registration results for PSO-SIFT; (d–f) registration results for RIFT; (g–i) registration results for LNIFT; (j–l) registration results for our LPHOG method. n represents the number of correct match points; r represents RMSE. $r = \infty$ indicates failure to match this image pair.

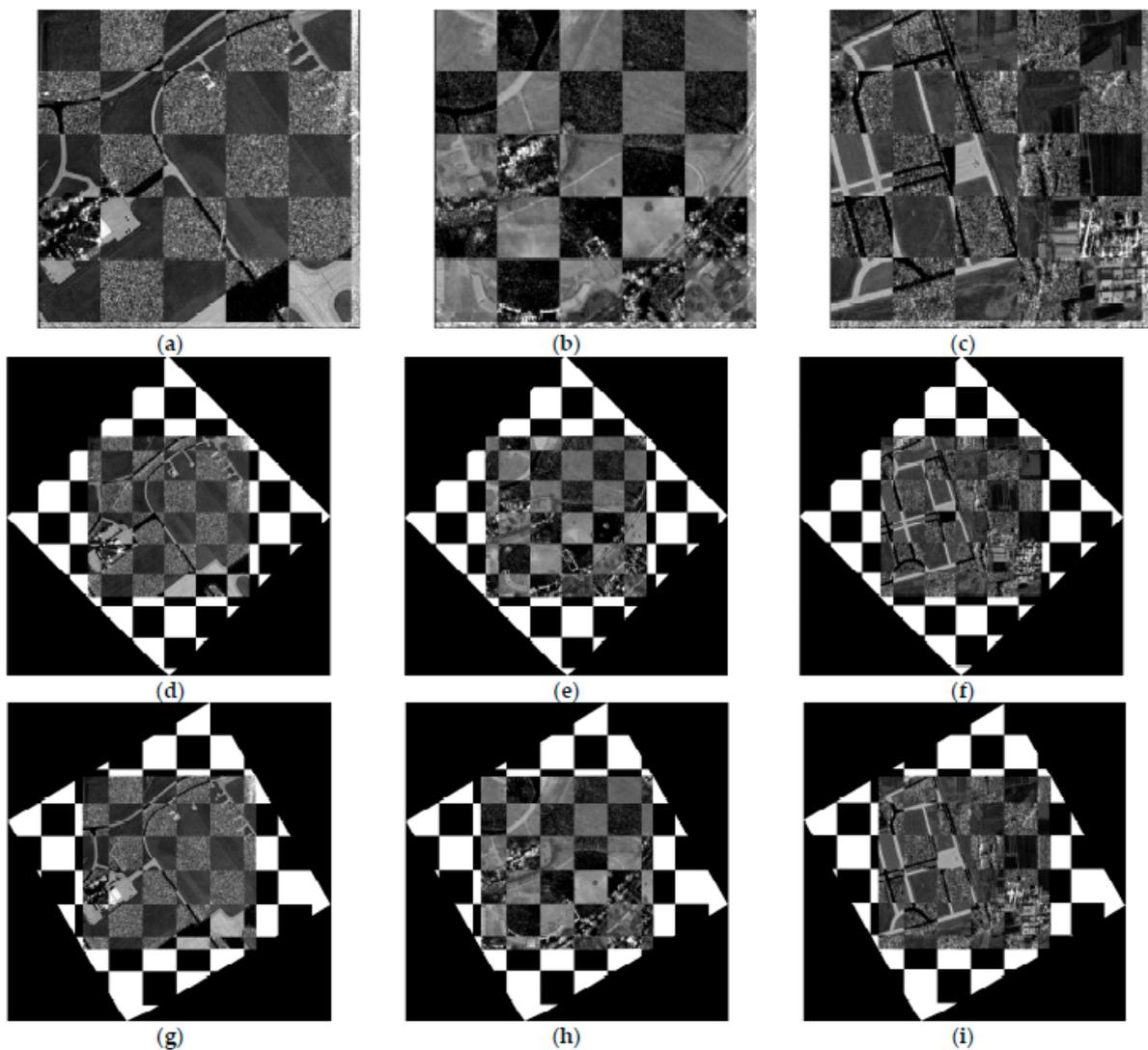


Figure 25. Checkerboard mosaic images of our LPHOG. (a–c) Without rotation; (d–f) with 45° rotation; (g–i) with 240° rotation.

Figure 27 shows the changes in $P_{NCM \geq 10}$, $P_{NCM \geq 100}$, average NCM, and average RMSE for the whole Dataset 1, respectively. The $P_{NCM \geq 10}$ of our method was always equal to 1, and $P_{NCM \geq 100}$ was basically maintained above 0.75, which indicated that our method could correctly match all of the images in Dataset 1 under every rotation angle, and most of the images were matched well under all rotation angles. The $P_{NCM \geq 10}$ of RIFT fluctuated between 0.5 and 0.8, and the $P_{NCM \geq 100}$ was maintained near 0 in most of the rotation angles, which indicated that RIFT could correctly match most of the images in most of the rotation angles, but the overall matching performance was poor. When the rotation angle was located in $[0^\circ, 90^\circ]$, the $P_{NCM \geq 10}$ of LNIFT could be maintained above 0.6, but the decline was very obvious in $[90^\circ, 180^\circ]$ (the change of P_1 in $[180^\circ, 360^\circ]$ and $[0^\circ, 180^\circ]$ was approximately symmetric). The $P_{NCM \geq 100}$ of LNIFT with the change in the rotation angles was similar to a parabola, and it dropped to 0 between $[135^\circ, 225^\circ]$, which indicated that the rotation invariance of LNIFT was poor. The average RMSE and $P_{NCM \geq 10}$ correlation was more obvious, with the average RMSE of our method staying near 1.12, and the other methods' RMSE being affected by $P_{NCM \geq 10}$ and located at a higher level. The average number of matched pairs of points for PSO-SIFT was maintained near 5; RIFT floated around 15–20; and the average number of correctly matched pairs of points for LNIFT showed a similar parabolic shape between $[0^\circ, 360^\circ]$, which reached the lowest level

between $[165^\circ, 195^\circ]$. Our LPHOG method stayed around 135, indicating that the overall matching performance of our method was very good and robust to rotation invariance.

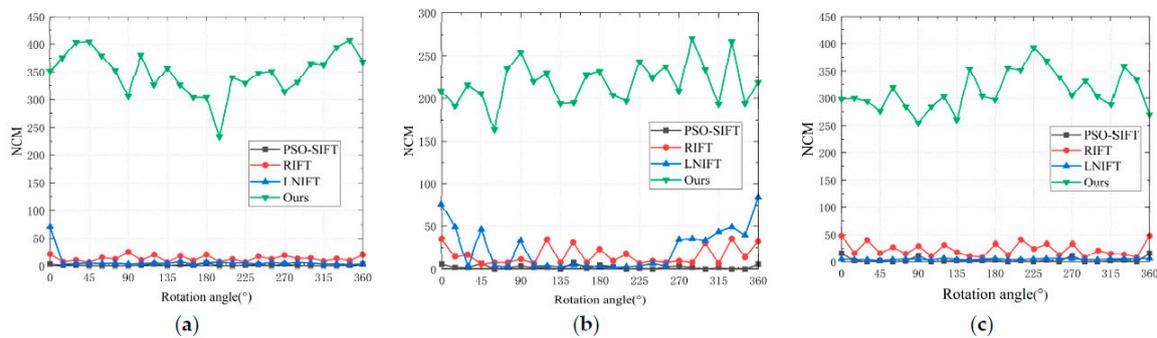


Figure 26. Changes in NCM for sample images (corresponding to those in Figure 22, respectively) from Dataset 1. (a) Pair 1; (b) Pair 2; (c) Pair 3.

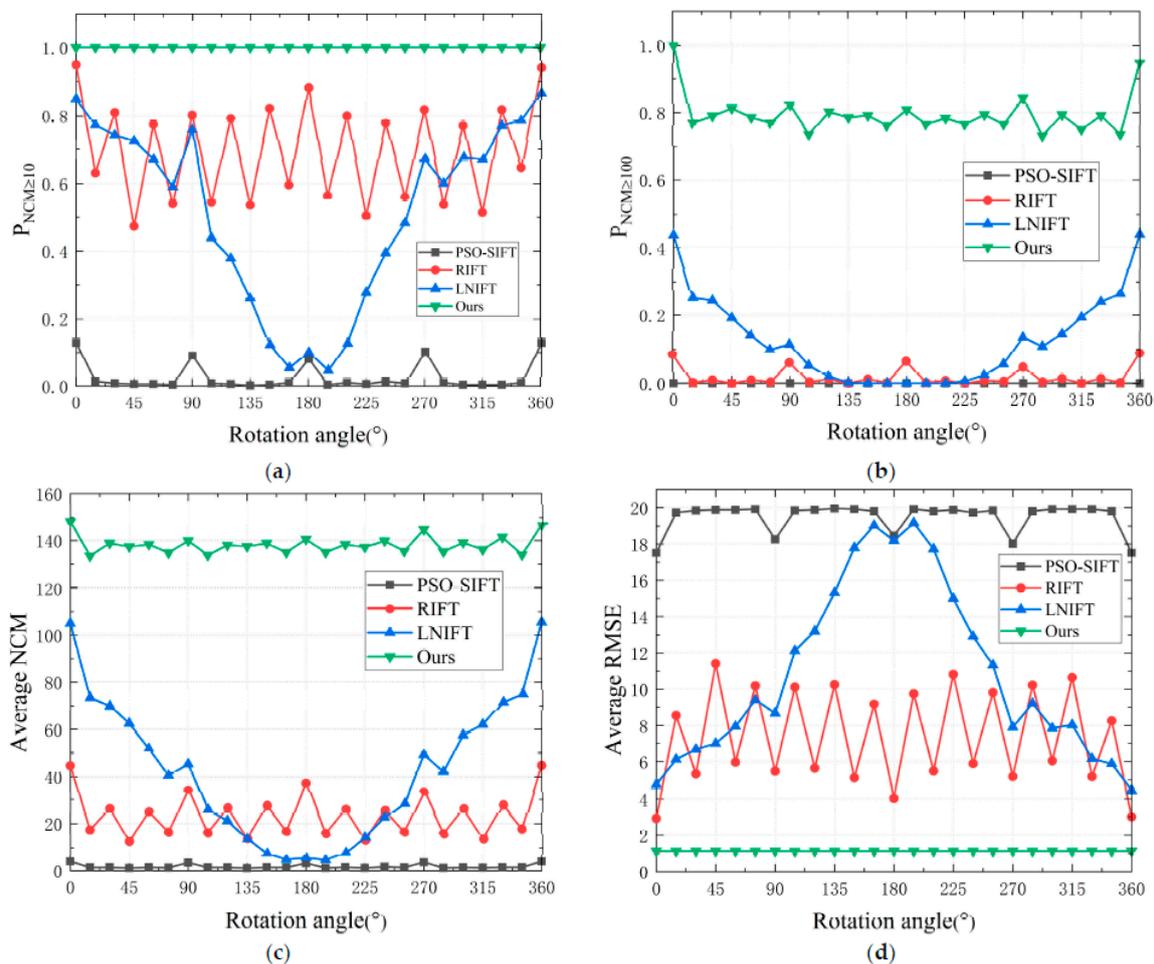


Figure 27. Changes in $P_{NCM \geq 10}$, $P_{NCM \geq 100}$, average NCM, and average RMSE for Dataset 1. (a) $P_{NCM \geq 10}$; (b) $P_{NCM \geq 100}$; (c) average NCM; (d) average RMSE.

The comparisons show that our method has wide applicability to the registration of many optical images and SAR images and adapts to many rotation angles. The matching performance of our method is very good, the number of matched pairs of points is always high, and the RMSE is always low. The rotation invariance of our method is highly robust.

4.3. Evaluation of Dataset 2

4.3.1. Qualitative Evaluation of Dataset 2

We selected three pairs of sample images from Dataset 2 to demonstrate the matching performance for optical images and infrared images. Figures 28–30 show the matching performance for sample images without rotation, after 45° rotation and after 240° rotation, respectively. Figure 31 shows the checkerboard mosaic images for our LPHOG method before rotation, after 45° and after 240° of rotation, respectively. Without rotation, PSO-SIFT could only correctly match the second pair of images; LNIFT could only correctly match the first and second pair; RIFT and our method could both correctly match these three pairs of images. After rotating 45°, PSO-SIFT failed on all three pairs of images; RIFT only correctly matched the third pair of images; LNIFT correctly matched the second pair of images, and our method correctly matched all three pairs of images. After rotating 240°, PSO-SIFT and LNIFT failed on all three pairs of images; RIFT correctly matched the first and second pairs of images; and our method correctly matched all three pairs of images. As seen from the checkerboard mosaic images in Figure 31, our method matched all three pairs of images with high accuracy in all three situations. The comparisons show that our method outperformed the other three methods in terms of matching performance, especially in terms of rotation invariance.

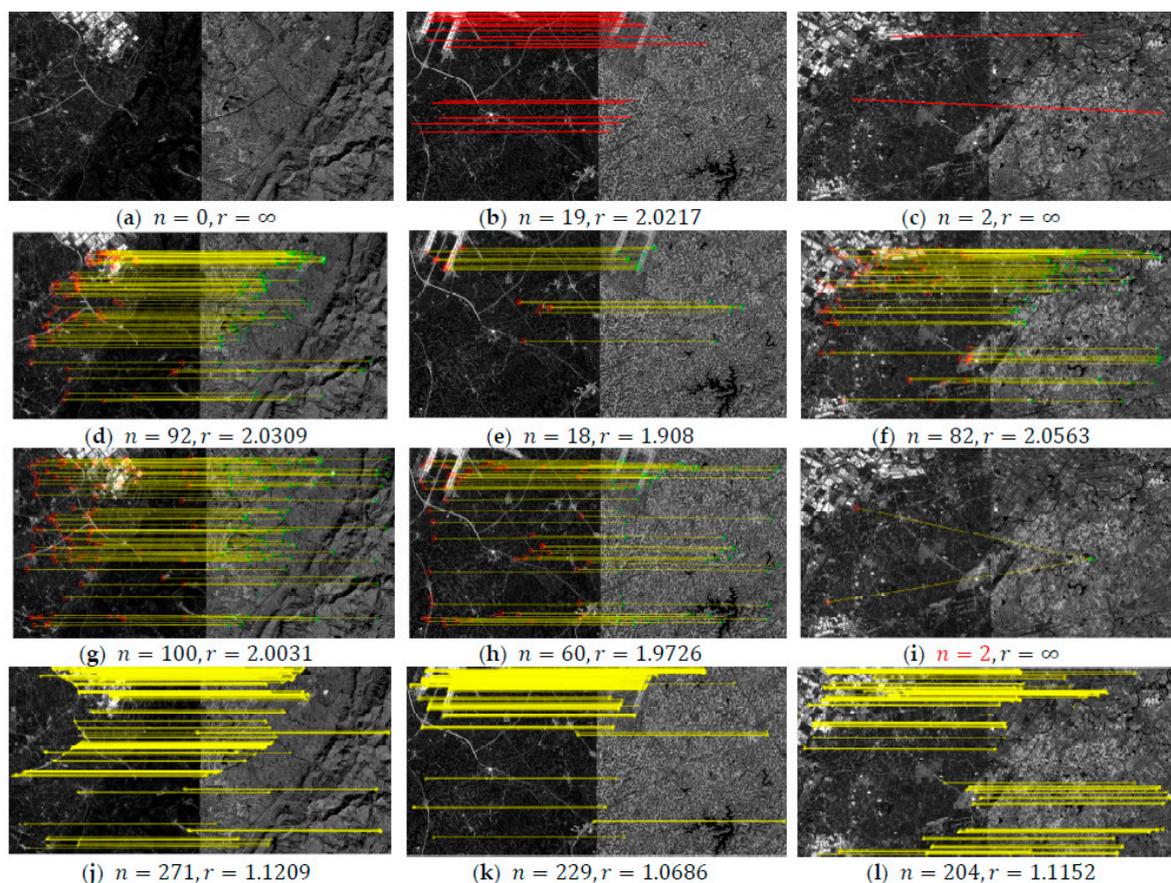


Figure 28. Registration results for Dataset 2 without rotation. (a–c) Registration results for PSO-SIFT; (d–f) registration results for RIFT; (g–i) registration results for LNIFT; (j–l) registration results for our LPHOG method. n represents the number of correct match points; r represents RMSE. $r = \infty$ indicates failure to match this image pair.

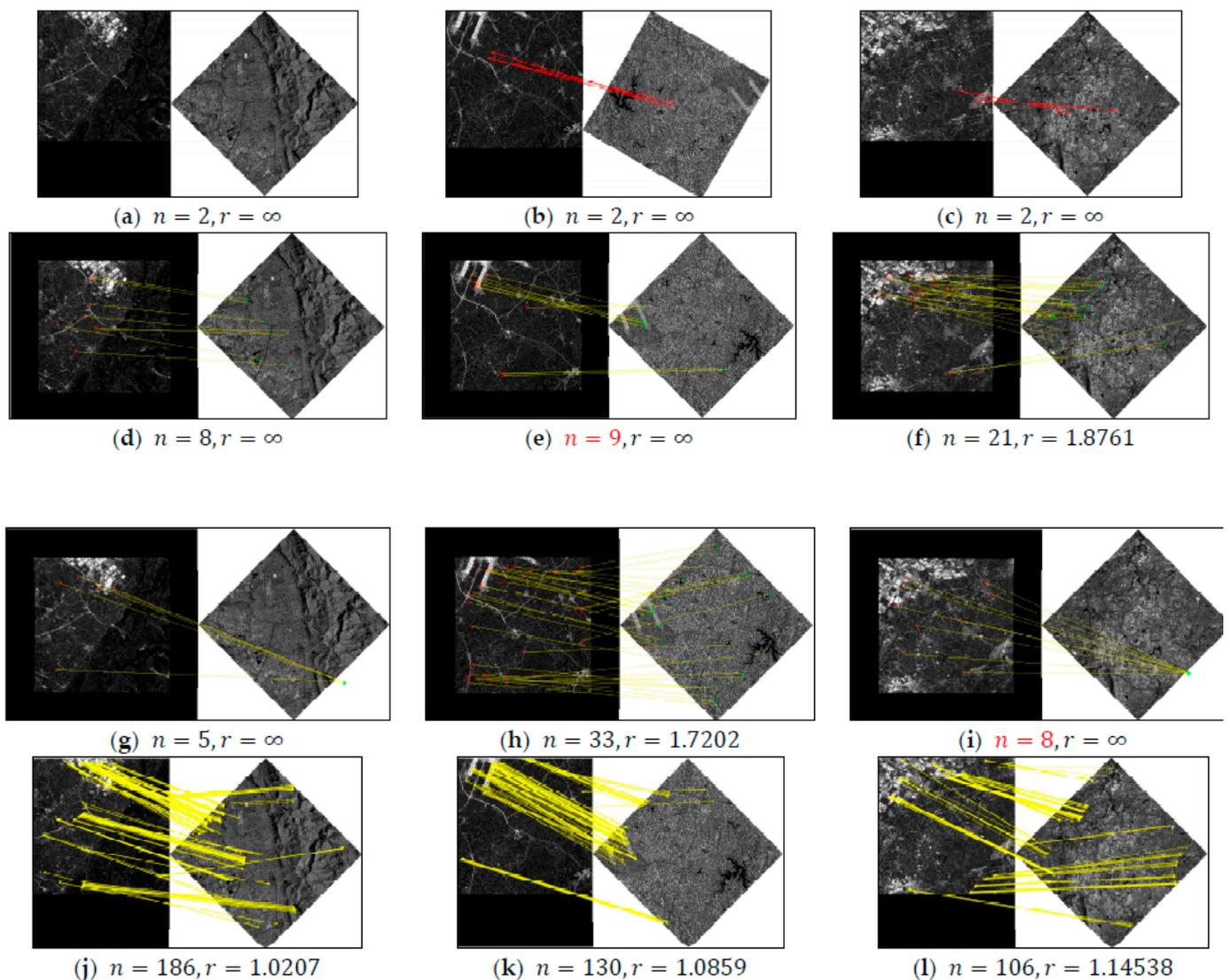


Figure 29. Registration results for Dataset 2 with 45° rotation. (a–c) Registration results for PSO-SIFT; (d–f) registration results for RIFT; (g–i) registration results for LNIFT; (j–l) registration results for our LPHOG method. n represents the number of correct match points; r represents RMSE. $r = \infty$ indicates failure to match this image pair.

4.3.2. Quantitative Evaluation of Dataset 2

Figure 32 shows the variation in NCM with rotation angles for the first pair of sample images, the second pair of sample images and the third pair of sample images, respectively. We rotated the registration image of each sample image pair. The rotation angles were from 0° to 360° with an interval of 15° . Thus, a total of 25 registration images for each image pair were obtained (the rotation angles were $[0^\circ, 15^\circ, 30^\circ, \dots, 315^\circ, 330^\circ, 345^\circ, 360^\circ]$) [24]. As a whole, our LPHOG method had more correctly matched point pairs than the other three methods for every rotation angle. Our method maintained the number of correctly matched point pairs above 100 for the first pair, above 50 for the second pair, and above 70 for the third pair at every rotation angle. The worst of the other methods, PSO-SIFT, correctly matched pairs of points close to 0 at any rotation angle for the first and third pairs; for the second pair, it was close to 0 for all angles except 0° and 360° . For the first and third pairs, the number of matched points for RIFT fluctuated between 20–75, and in a few rotation angles, it was up to 1/2 of ours. For the second pair, the NCM of RIFT was in the range of 5–25, and in a few rotation angles, it reached 1/10 of ours. For all of the sample images, the NCM of LNIFT at $[180^\circ, 360^\circ]$ varied basically symmetrically with

$[0^\circ, 180^\circ]$. The number of correctly matched pairs of points for LNIFT at $[0^\circ, 45^\circ]$ in some cases maintained a high level, but was still far below that of our method and approached 0 in most cases between $[45^\circ, 180^\circ]$.

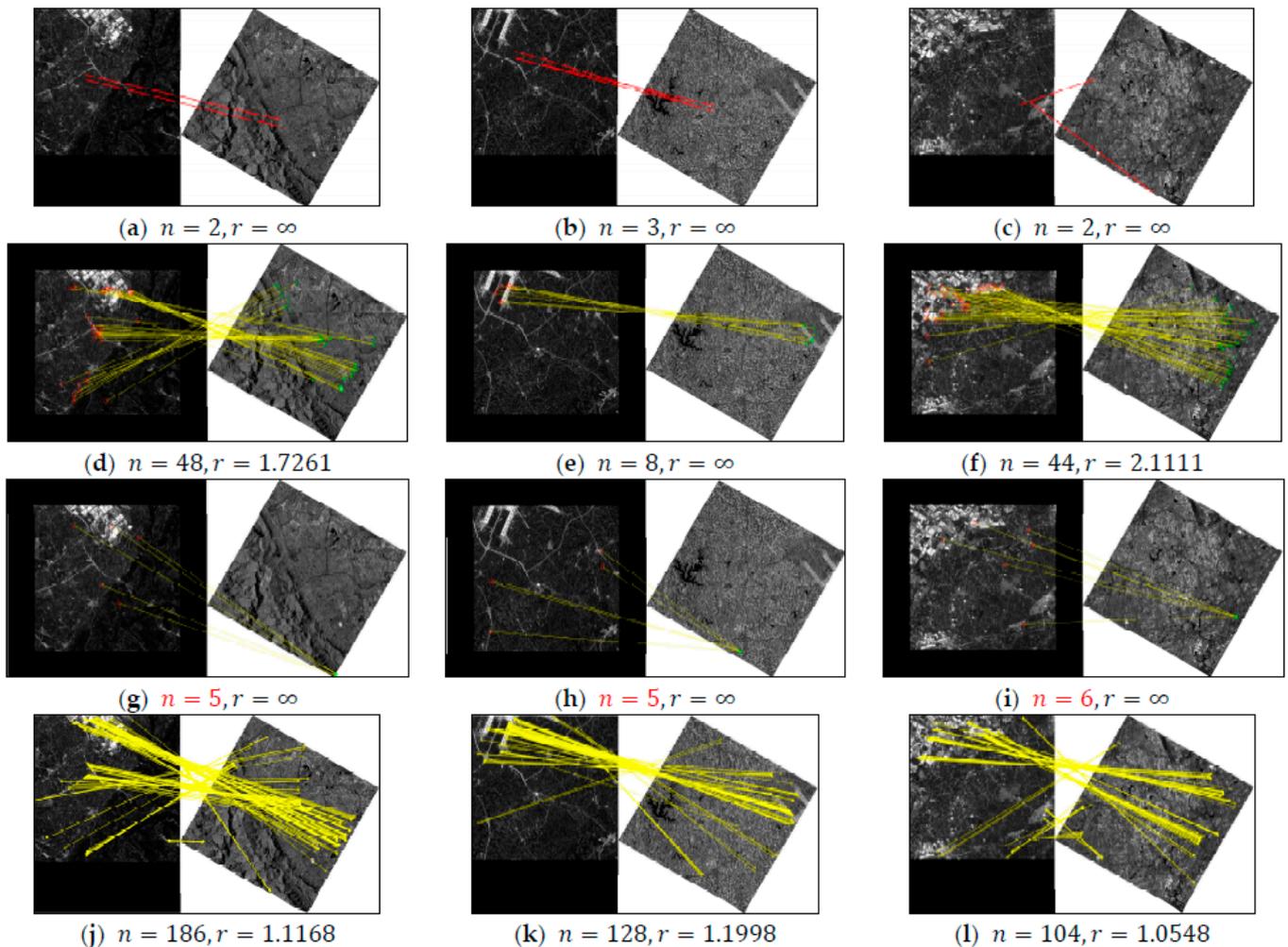


Figure 30. Registration results for Dataset 2 with 240° rotation. (a–c) Registration results for PSO-SIFT; (d–f) registration results for RIFT; (g–i) registration results for LNIFT; (j–l) registration results for our LPHOG method. n represents the number of correct match points; r represents RMSE. $r = \infty$ indicates failure to match this image pair.

Figure 33 shows the changes in $P_{NCM \geq 10}$, $P_{NCM \geq 100}$, average NCM, and average RMSE for the whole Dataset 2, respectively. As a whole, our $P_{NCM \geq 10}$, $P_{NCM \geq 100}$ and average NCM were larger than those of PSO-SIFT, RIFT and LNIFT at every rotation angle, and the average RMSE was smaller than those of PSO-SIFT, RIFT and LNIFT at every rotation angle, which indicated that our algorithm's matching performance, rotation invariance and matching accuracy were better than those of these three algorithms. Among them, the $P_{NCM \geq 100}$ and the average number of matched points with our method had more significant jumps at 0° , 90° , 180° , 270° and 360° , which should be due to the fact that the line feature detection was not affected by the image borders in these rotation angle cases. How to eliminate the influence of image borders on line feature detection at other rotation angles will be one of our key research topics in the future.

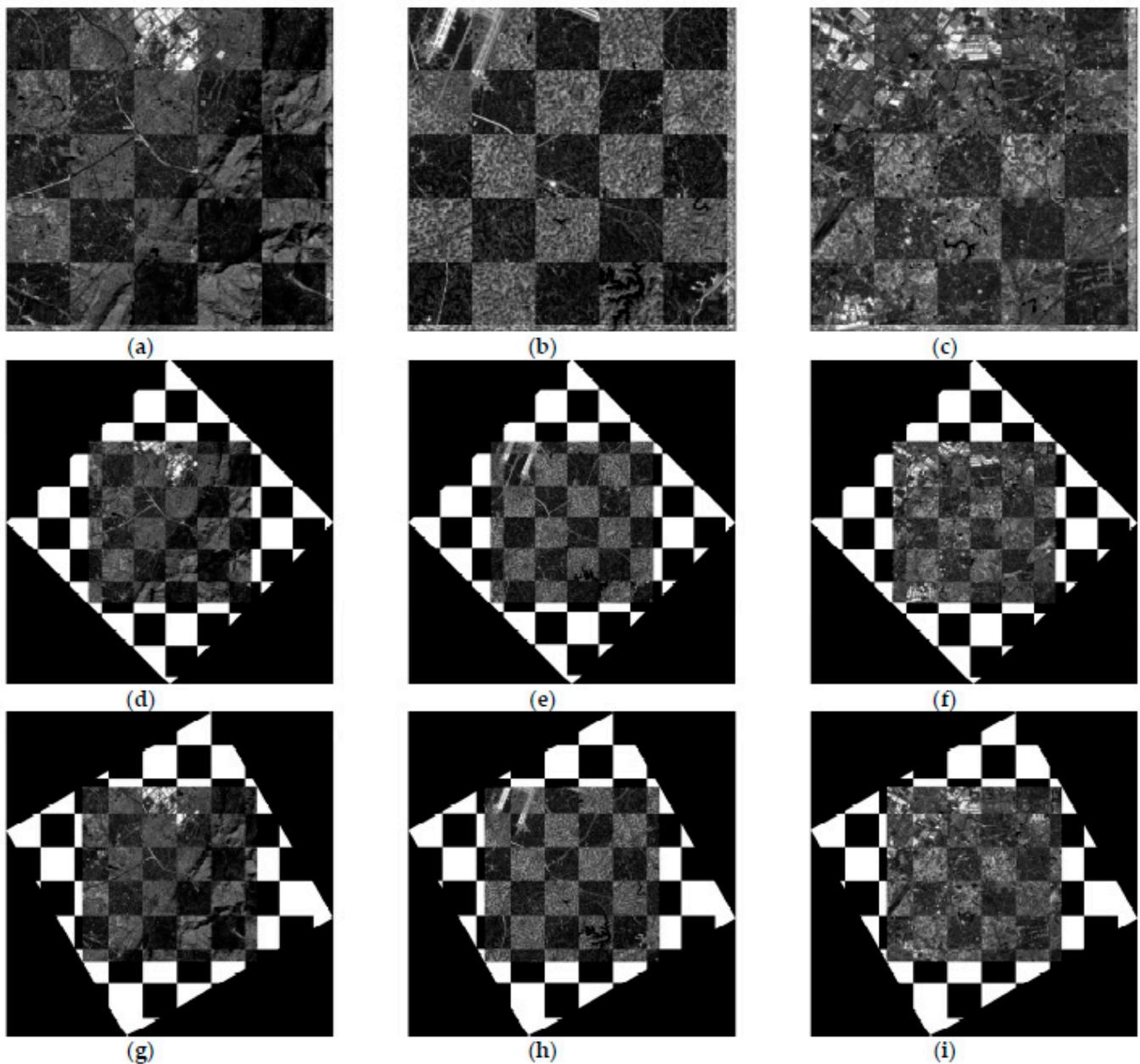


Figure 31. The checkerboard mosaic images of our LPHOG method for sample images in Dataset 2. (a–c) Without rotation; (d–f) after 45° rotation; (g–i) after 240° rotation.

The comparisons show that our LPHOG method has wide applicability to the registration of many optical images and infrared images, and adapts to many rotation angles. The number of matched pairs of points is always high, and the RMSE is always low. The rotation invariance of our method is highly robust for optical and infrared image registration.

4.4. Evaluation of Dataset 3

4.4.1. Qualitative Evaluation of Dataset 3

We selected three pairs of sample images from Dataset 3 to compare the matching performance of optical and optical matching. Figures 34–36 show the matching performance for sample images without rotation, after 45° rotation and after 240° rotation, respectively. Figure 37 shows the checkerboard mosaic images of our method before rotation and after 45° and 240° rotation, respectively. Without rotation, PSO-SIFT and RIFT could correctly match three pairs of images; LNIFT could only match the second pair of images; our

LPHOG method could correctly match these three pairs of images, and the number of matched points was much higher than that of the other three methods. As shown in Figure 35, after 45° of rotation, PSO-SIFT and LNIFT failed to match these three pairs of images; both RIFT and our method could correctly match these three pairs of images. As shown in Figure 36, after 240° rotation, LNIFT failed to match the three pairs of images; PSO-SIFT matched the third pair of images correctly; RIFT matched the three pairs of images correctly, but the number of matched pairs was lower; our method matched the three pairs of images correctly and the numbers of matched pairs were high. As shown in Figure 37, our method’s matching accuracy was very high. (Some regions are misaligned because the two images correspond to different times, and the corresponding object sizes and shapes do not exactly coincide.)

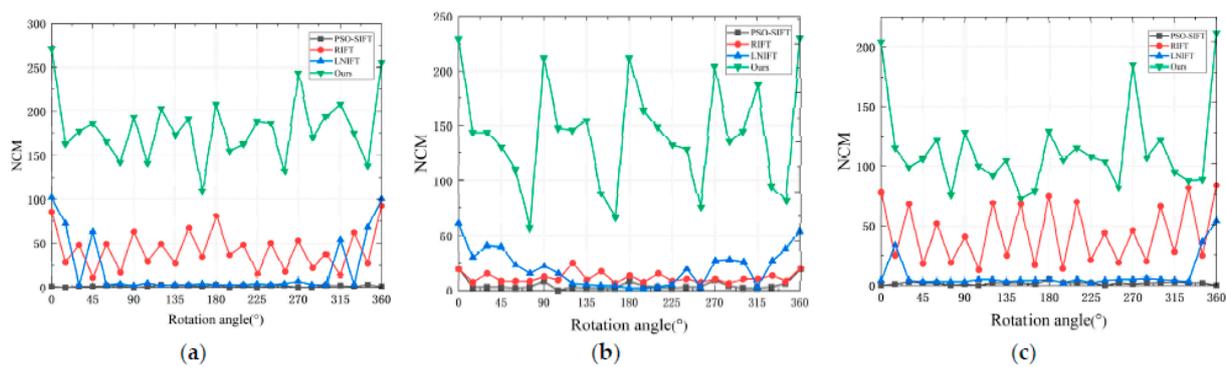


Figure 32. Changes in NCM for the sample images (corresponding to those in Figure 28, respectively) in Dataset 2. (a) Pair 1; (b) Pair 2; (c) Pair 3.

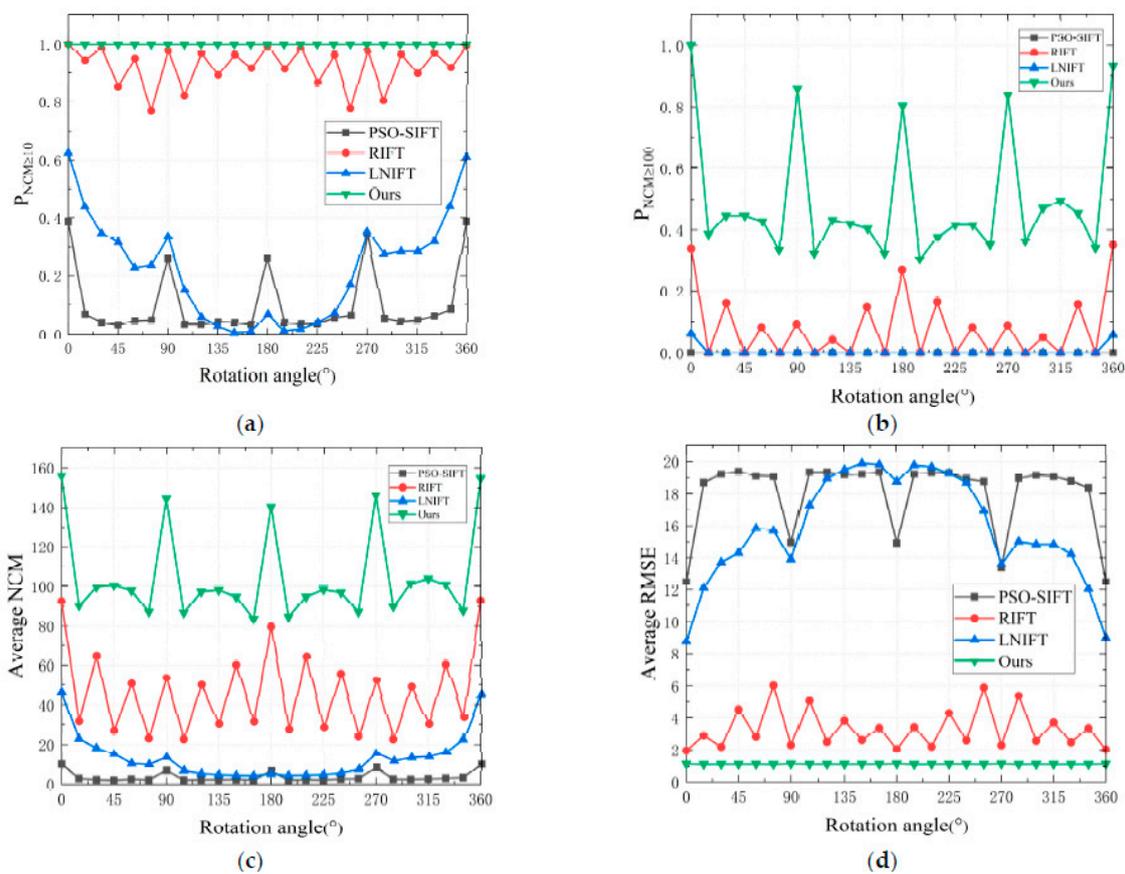


Figure 33. Changes in $P_{NCM \geq 10}$, $P_{NCM \geq 100}$, average NCM, average RMSE for Dataset 2. (a) $P_{NCM \geq 10}$; (b) $P_{NCM \geq 100}$; (c) average NCM; (d) average RMSE.

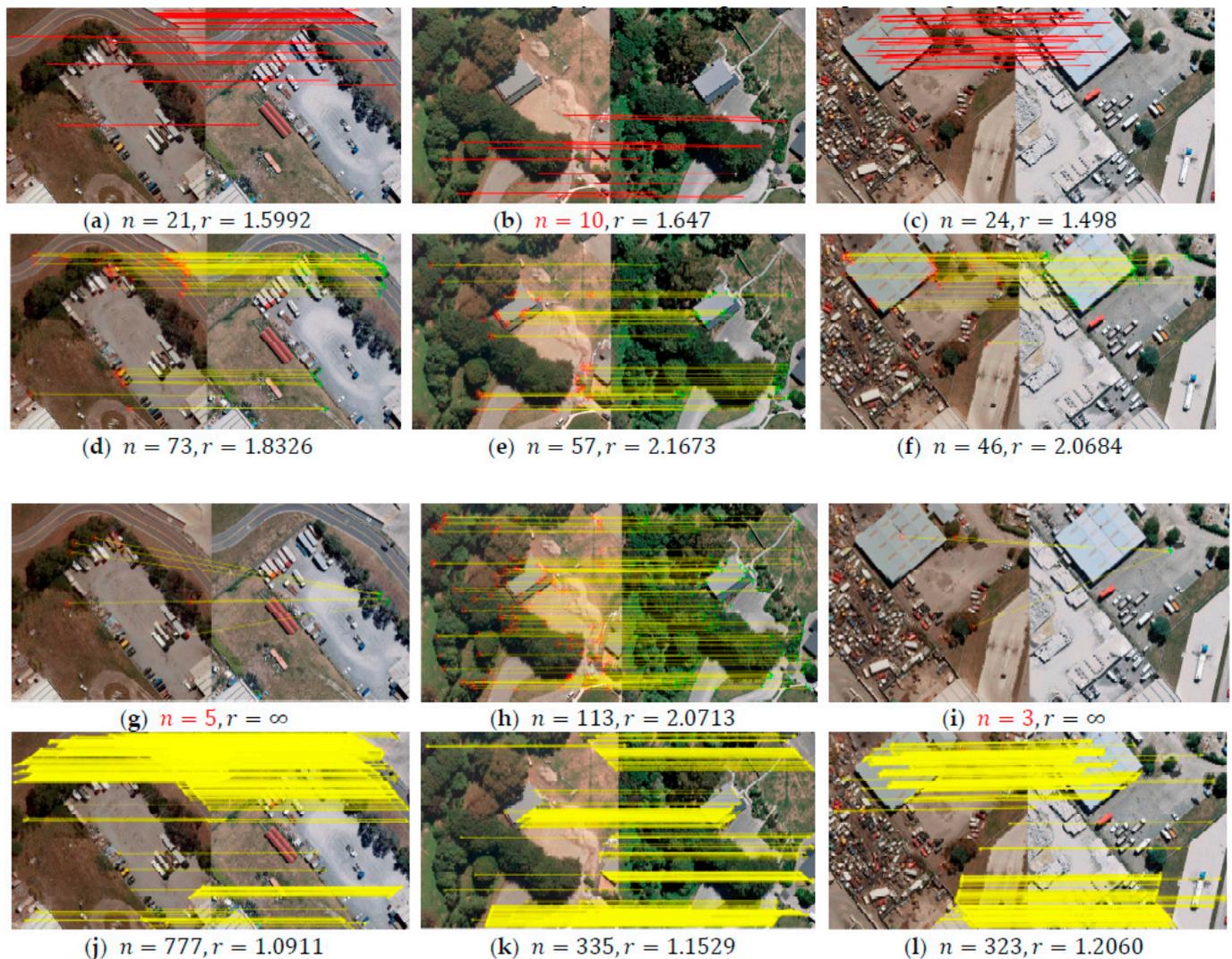


Figure 34. Registration results for Dataset 3 without rotation. (a–c) Registration results for PSO-SIFT; (d–f) registration results for RIFT; (g–i) registration results for LNIFT; (j–l) registration results for our LPHOG method. n represents the number of correct match points; r represents RMSE. $r = \infty$ indicates failure to match this image pair.

4.4.2. Quantitative Evaluation of Dataset 3

Figure 38 shows the variation in NCM with rotation angles for the first pair of sample images, the second pair of sample images and the third pair of sample images, respectively. We rotated the registration image of each sample image pair. The rotation angles were from 0° to 360° with an interval of 15° . Thus, a total of 25 registration images for each image pair were obtained (the rotation angles were $[0^\circ, 15^\circ, 30^\circ, \dots, 315^\circ, 330^\circ, 345^\circ, 360^\circ]$) [24]. We could see that the number of correctly matched point pairs with our method was much higher than that of the other three methods for all three pairs of samples. The NCM was greater than 500 for most rotation angles in the first pair and not less than 450 for all rotation angles; it was higher than 200 for most rotation angles in the second pair; and it was higher than 250 for most rotation angles in the third pair, which indicated that our method has very good rotation invariance.

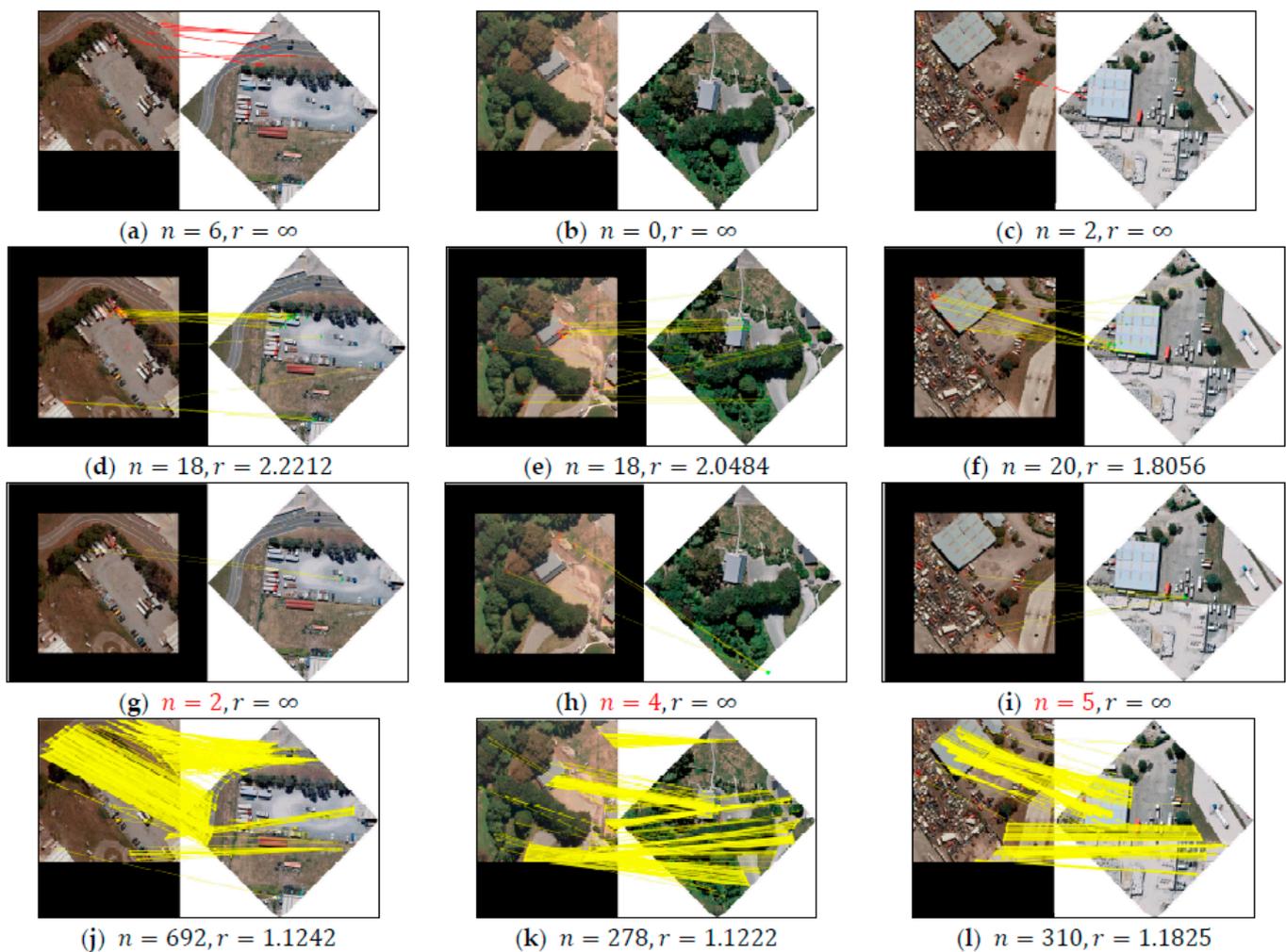


Figure 35. Registration results for Dataset 3 with 45° rotation. (a–c) Registration results for PSO-SIFT; (d–f) registration results for RIFT; (g–i) registration results for LNIFT; (j–l) registration results for our LPHOG method. n represents the number of correct match points; r represents RMSE. $r = \infty$ indicates failure to match this image pair.

Figure 39 shows the changes in $P_{NCM \geq 10}$, $P_{NCM \geq 100}$, average NCM, and average RMSE for the whole Dataset 3, respectively. In most of the rotation angle cases, the $P_{NCM \geq 10}$ of PSO-SIFT fluctuated between 0.2 and 0.4 and could jump to between 0.57 and 0.8 at 0° , 90° , 180° , 270° , and 360° , which indicated that PSO-SIFT detected many points of the image borders but could not eliminate the influence of the points of the image borders, reducing the matching performance. The $P_{NCM \geq 10}$ and $P_{NCM \geq 100}$ of LNIFT in the range of $[0^\circ, 90^\circ]$ were a little high; its $P_{NCM \geq 10}$ was higher than 0.7, and its $P_{NCM \geq 100}$ was higher than 0.57. However, both of them decreased rapidly between $[90^\circ, 180^\circ]$ (the changes in $P_{NCM \geq 10}$ and $P_{NCM \geq 100}$ of LNIFT in the range between $[180^\circ, 360^\circ]$ were approximately symmetrical with those in $[0^\circ, 180^\circ]$), which indicates that LNIFT had better matching performance in small angular differences, but poor matching performance in large angular differences. The $P_{NCM \geq 10}$ of our LPHOG method always equaled 1 for all rotation angles, and the $P_{NCM \geq 10}$ of RIFT was close to 1 for all rotation angles, indicating that the applicability of both our method and RIFT is good. However, the $P_{NCM \geq 100}$ of our method was significantly higher than the $P_{NCM \geq 100}$ of RIFT, and our method's $P_{NCM \geq 100}$ was always very close or equal to 1, indicating that the matching performance of our method not only has wide applicability but also has very good rotation invariance. Our method's average RMSE was maintained around 1.114, which was lower than that of RIFT, and much lower than that of PSO-SIFT and LNIFT, indicating that the matching accuracy of our method is very high overall. As

a whole, the average NCM values were significantly better than those of the other three methods, 2.5 times higher than those of the other three methods for every rotation angle, and the average NCM of our method was higher than 400 for most of the rotation angles, which indicates that our algorithm is significantly better than the other three algorithms.

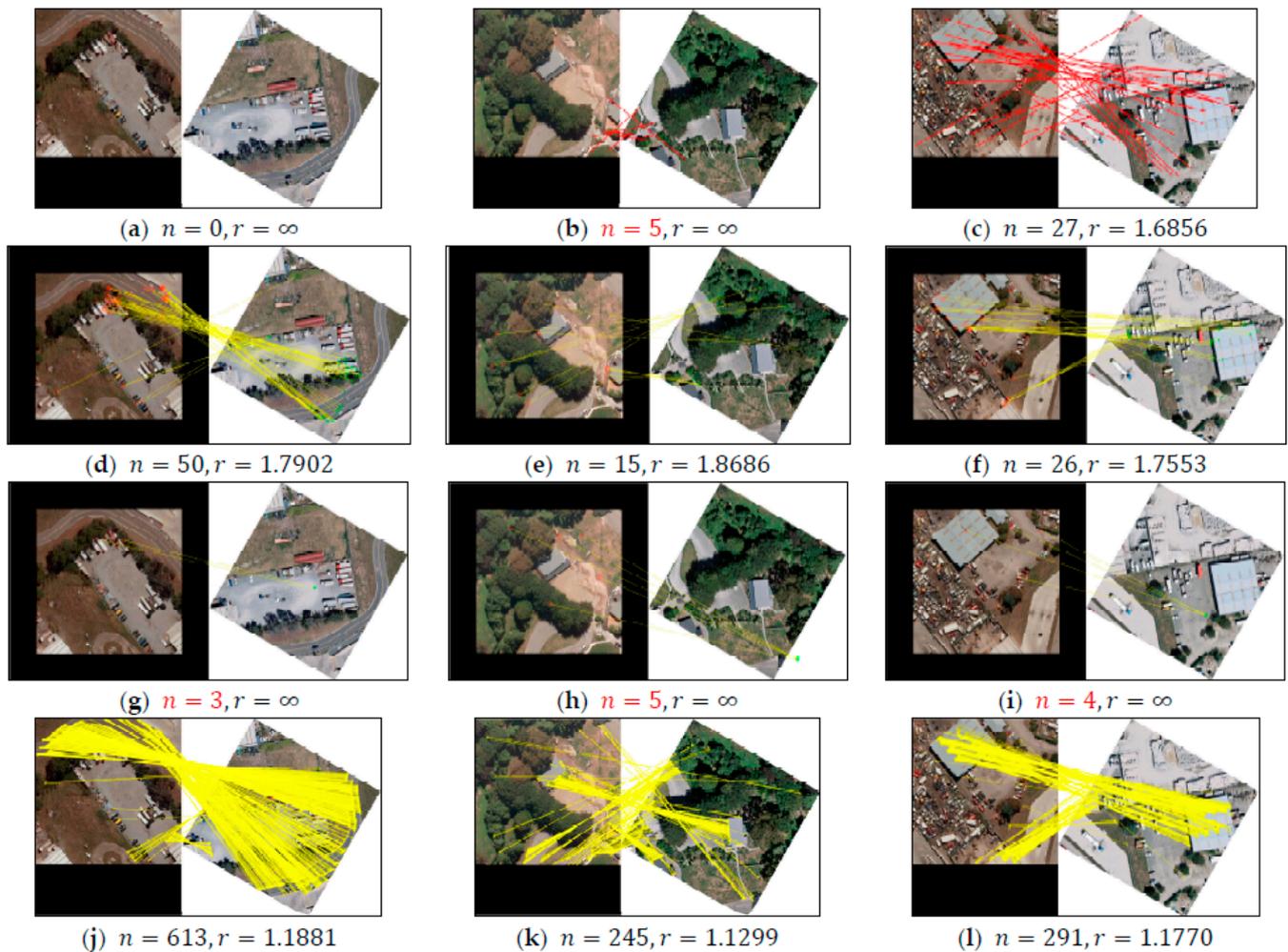


Figure 36. Registration results for Dataset 3 with 240° rotation. (a–c) Registration results for PSO-SIFT; (d–f) registration results for RIFT; (g–i) registration results for LNIFT; (j–l) registration results for our LPHOG method. n represents the number of correct match points; r represents RMSE. $r = \infty$ indicates failure to match this image pair.

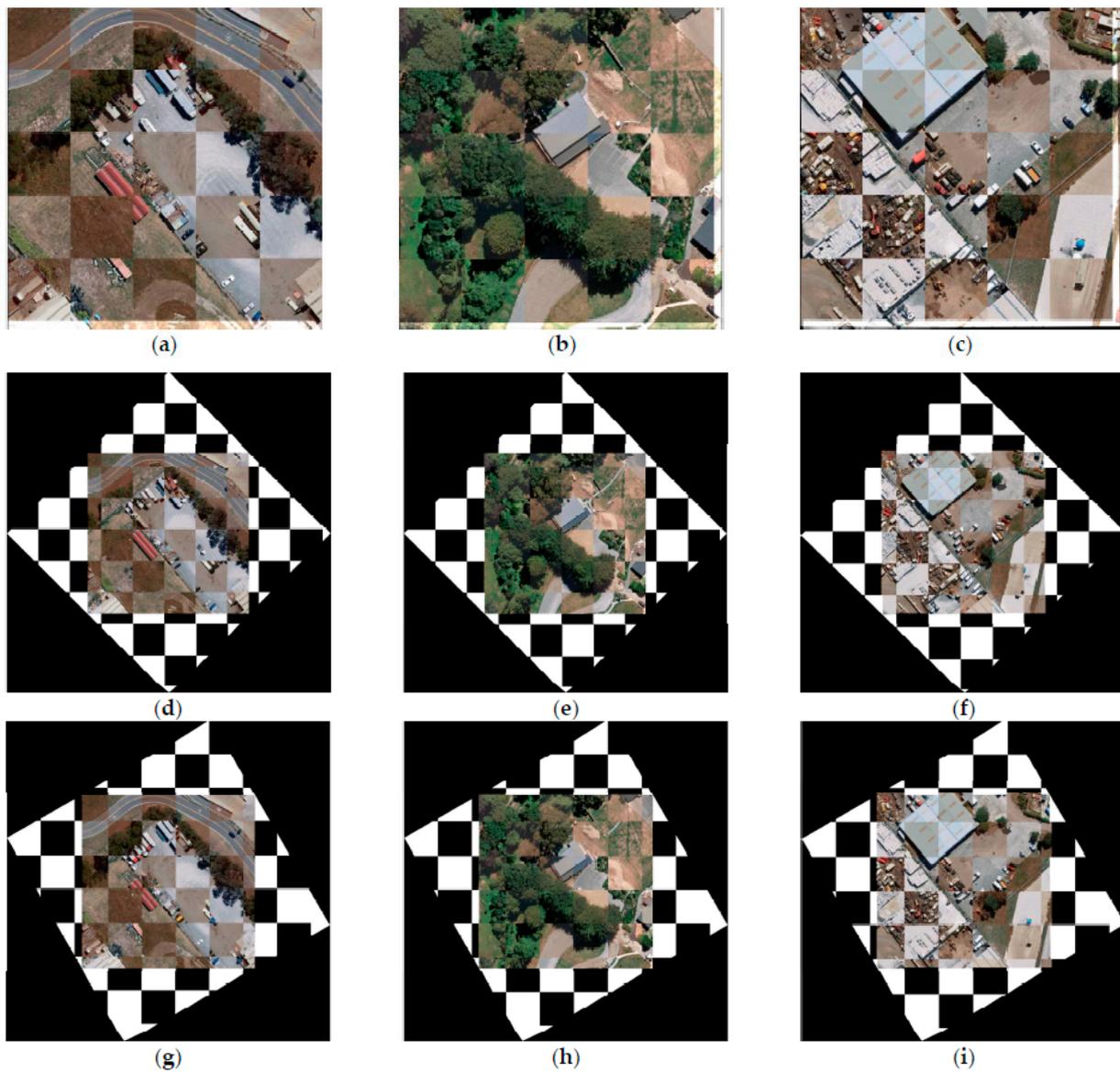


Figure 37. The checkerboard mosaic images of our LPHOG method for sample images in Dataset 3. (a–c) Without rotation; (d–f) after 45° rotation; (g–i) after 240° rotation.

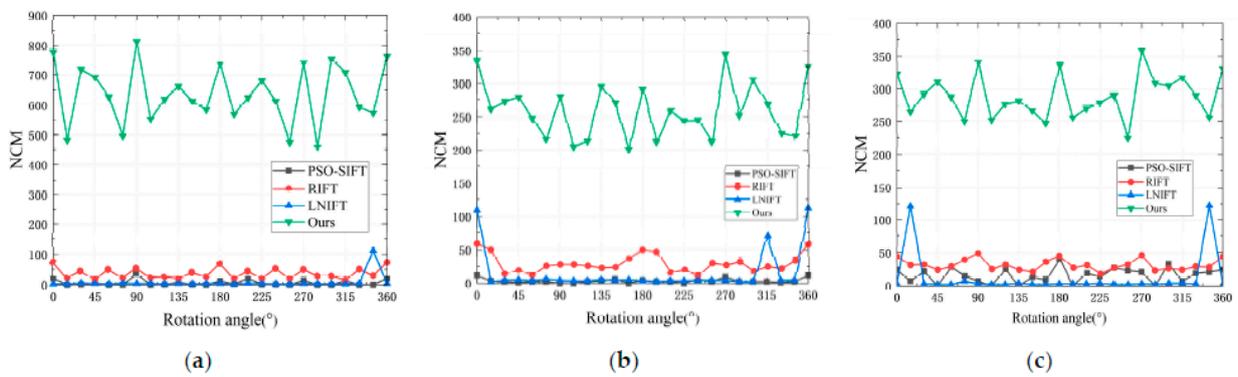


Figure 38. Changes in NCM for the sample images (corresponding to those in Figure 34, respectively) in Dataset 3. (a) Pair 1; (b) Pair 2; (c) Pair 3.

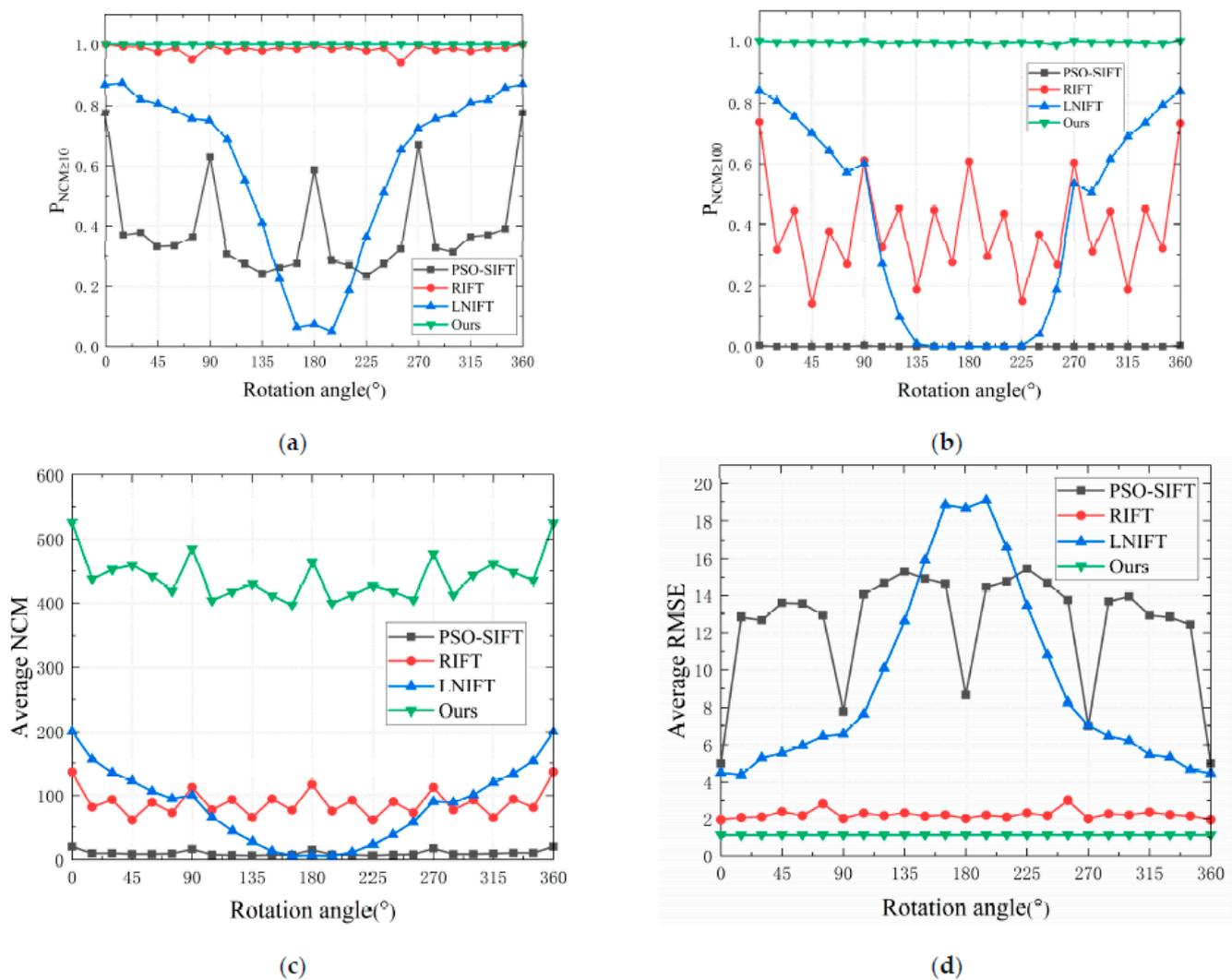


Figure 39. Changes in $P_{NCM \geq 10}$, $P_{NCM \geq 100}$, average NCM, and average RMSE for Dataset 3. (a) $P_{NCM \geq 10}$; (b) $P_{NCM \geq 100}$; (c) average NCM; (d) average RMSE.

The comparisons show that our method has wide applicability to the registration of many optical images with optical images, and adapts to many rotation angles. The number of matched pairs of points is always high, and the RMSE is always low. The rotation invariance of our method is highly robust for optical and optical image registration.

5. Discussion

In this study, we achieved robust rotation invariance. In future research, we will focus on scale invariance, the runtimes of algorithms and the effect of the number of iterations of the FSC algorithm. Figure A1 shows the performance scale invariance of our method. Figure A2 shows the effect of the number of iterations of the FSC algorithm. For the algorithm runtime, we will consider simplifying the steps of feature point extraction to reduce runtimes in the future.

6. Conclusions

In this study, a line feature- and point feature-combined heterologous image registration algorithm, LPHOG, is proposed to address the limitations of feature detection during registration, the weak rotation invariance of point feature-based methods and the complicated keypoint extraction steps in line feature-based methods.

We firstly preprocessed images according to their characteristics, and then detected the line features with the LSD algorithm. Then, we extracted the keypoints and constructed HOG-like feature descriptors on rotated images. While determining the dominant orientations of the keypoints, we abandoned the normal method of determining the dominant orientation by the intensity or gradient of the image patch around the feature points, instead using the tilt angle of the line directly.

The results and evaluations show that our method has wide applicability and adapts to many rotation angles. The matching performance of our method is very good, the number of matched pairs of points is always high, and the RMSE is always low. The rotation invariance of our method is highly robust.

Author Contributions: Conceptualization, J.H. and X.J.; methodology, J.H., X.J. and W.G.; validation, J.H., X.J., Z.H., W.G. and S.L.; formal analysis, J.H., X.J., Z.H. and M.Z.; writing—original draft preparation, J.H.; writing—review and editing, J.H., X.J., Z.H., M.Z., W.G. and S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science and Technology Department of Jilin Province, China, under grant number 20220201146GX.

Data Availability Statement: All three datasets are publicly available: <https://sites.google.com/view/yumingxiang/%E9%A6%96%E9%A1%B5> (accessed on 20 June 2023) for Dataset 1, <https://github.com/woshiybc/Multi-Scale-Unsupervised-Framework-MSUF> (accessed on 20 June 2023) for Dataset 2, http://gpcv.whu.edu.cn/data/building_dataset.html (accessed on 20 June 2023) for Dataset 3.

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

Appendix A

In order to obtain good results, we changed the size of the HOG descriptor according to the scale change one by one, because we did not construct a multiscale space and did not consider the scale factor beforehand. We achieved the scale change by down-sampling or up-sampling the optical image. In the test, the patch size of the SAR image's HOG descriptor remained 128×128 all the time. When scale = 0.5, we set the patch size of the optical image's HOG descriptor equal to 64×64 . When scale = 1, we set the patch size of the optical image's HOG descriptor equal to 128×128 . When scale = 1.5, we set the patch size of the optical image's HOG descriptor equal to 192×192 . When scale = 2, we set the patch size of the optical image's HOG descriptor equal to 256×256 . Figure A1 shows the matching results for the scale change in the sample images.

We tested the effect of scale change on registration performance for three sample image pairs. We found that the registration performance remained good, although the number of correct match points declined with the scale change. The main factor influencing the number of correct match points was the limitation feature points number. When the scale of the optical image became larger, the limitation feature points number should be set larger. In the future, we will focus on scale invariance research.

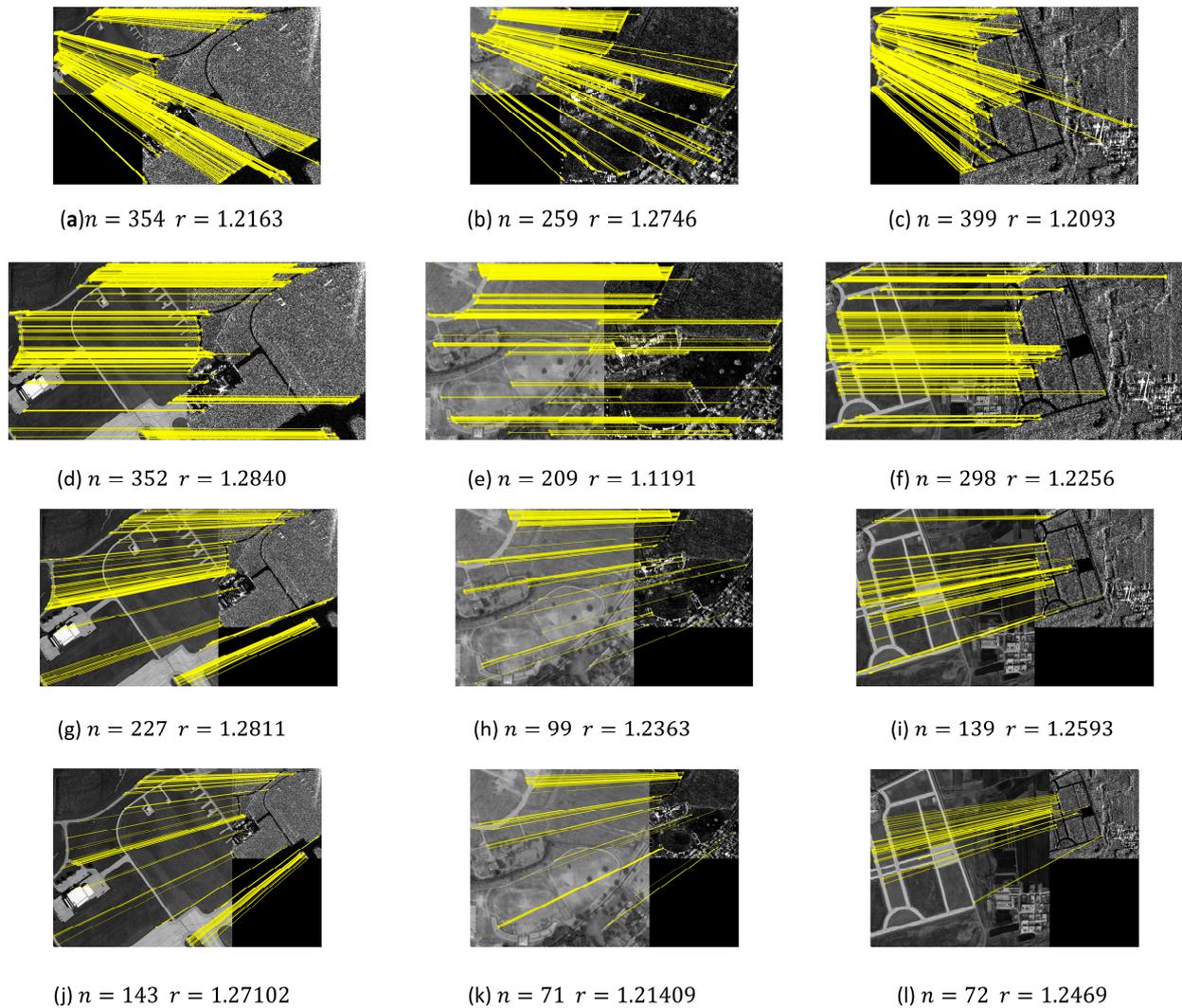


Figure A1. Registration results for scale change. (a–c) Scale = 0.5. (d–f) Scale = 1. (g–i) Scale = 1.5. (j–l) Scale = 2. n represents the number of correct match points; r represents RMSE.

Appendix B

We tested the effect of the number of iterations of the FSC algorithm on the sample images. Figure A2 shows the matching performance with different FSC iterations.

From Figure A2, we can see that small numbers of iterations could limit the registration performance, but the impact was not very strong. When the iteration number equaled 600, the images could still achieve very good registration. After increasing the number of iterations, the number of correct match points increased. The change in correct match points was not significant when the number of iterations was larger than 800, which means the convergence of the FSC algorithm is very good if the number of iterations is not small than 800. The number of iterations will be one of our research topics in the future.

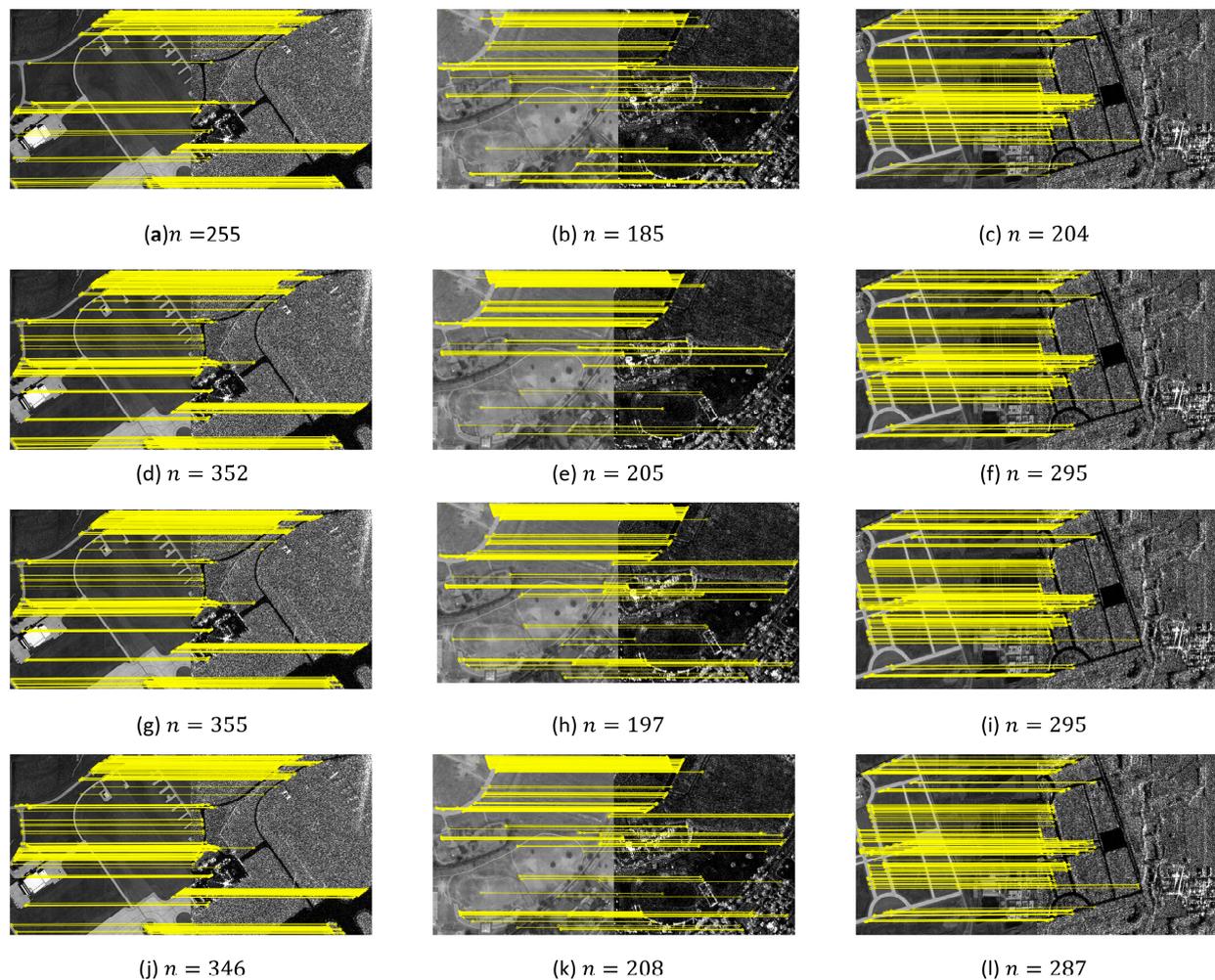


Figure A2. Matching results with changes in the number of FSC iterations. (a–c) Iterations = 600. (d–f) Iterations = 800. (g–i) Iterations = 1000. (j–l) Iterations = 1500. N represents the number of correct match points.

References

1. Wong, A.; Claudi, D.A. ARRSI: Automatic Registration of Remote-Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 1483–1493. [[CrossRef](#)]
2. Zitová, B.; Flusser, J. Image registration methods: A survey. *Image Vis. Comput.* **2003**, *21*, 977–1000. [[CrossRef](#)]
3. 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](#)]
4. Liu, S.; Jiang, J. Registration Algorithm Based on Line-Intersection-Line for Satellite Remote Sensing Images of Urban Areas. *Remote Sens.* **2019**, *11*, 1400. [[CrossRef](#)]
5. Ye, Y.; Shen, L.; Hao, M.; Wang, J.; Xu, Z. Robust Optical-to-SAR Image Matching Based on Shape Properties. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 564–568. [[CrossRef](#)]
6. 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](#)]
7. 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 September–19 October 2007; pp. 2681–2684.
8. 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.
9. Li, K.; Zhang, X. 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.
10. Moravec, H.P. Towards automatic visual obstacle avoidance. In Proceedings of the International Joint Conferences on Artificial Intelligence, Cambridge, MA, USA, 22–25 August 1977; p. 584.

11. Harris, C.G.; Stephens, M.J. A Combined Corner and Edge Detector. In Proceedings of the Alvey Vision Conference, Manchester, UK, 31 August–2 September 1988; pp. 147–151.
12. Stephen, M.S.J.; Michael, B. SUSAN—A New Approach to Low Level Image Processing. *Int. J. Comput. Vis.* **1997**, *23*, 45–78.
13. Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [[CrossRef](#)]
14. Bay, H.; Tuytelaars, T.; Gool, L.V. SURF: Speeded Up Robust Features. In Proceedings of the European Conference on Computer Vision, Graz, Austria, 7–13 May 2006; pp. 404–417.
15. Touzi, R.; Lopes, A.; Bousquet, P. A statistical and geometrical edge detector for SAR images. *IEEE Trans. Geosci. Remote Sens.* **1988**, *26*, 764–773. [[CrossRef](#)]
16. Jiao, X.; Wen, X. SAR Image Segmentation Based on Markov Random Field Model and Multiscale Technology. In Proceedings of the 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery, Tianjin, China, 14–16 August 2009; pp. 442–446.
17. Luo, S.; Tong, L.; Chen, Y. A Multi-Region Segmentation Method for SAR Images Based on the Multi-Texture Model with Level Sets. *IEEE Trans. Image Process.* **2018**, *27*, 2560–2574. [[CrossRef](#)]
18. Ye, Y.; Tang, T.; Zhu, B.; Yang, C.; Li, B.; Hao, S. A Multiscale Framework with Unsupervised Learning for Remote Sensing Image Registration. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5622215. [[CrossRef](#)]
19. Cao, S.-Y.; Yu, B.; Luo, L.; Zhang, R.; Chen, S.-J.; Li, C.; Shen, H.-L. PCNet: A structure similarity enhancement method for multispectral and multimodal image registration. *Inf. Fusion* **2023**, *94*, 200–214. [[CrossRef](#)]
20. Xu, H.; Ma, J.; Yuan, J.; Le, Z.; Liu, W. RFNet: Unsupervised Network for Mutually Reinforcing Multi-Modal Image Registration and Fusion. In Proceedings of the the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 19–20 June 2022; pp. 19679–19688.
21. Jia, X.; Bartlett, J.; Chen, W.; Song, S.; Zhang, T.; Cheng, X.; Lu, W.; Qiu, Z.; Duan, J. Fourier-Net: Fast Image Registration with Band-Limited Deformation. In Proceedings of the the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; pp. 1015–1023.
22. 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.* **2017**, *14*, 3–7. [[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. 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](#)]
25. 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*, 5621314. [[CrossRef](#)]
26. Mikolajczyk, K.; Schmid, C. A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1615–1630. [[CrossRef](#)]
27. 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](#)]
28. Dellinger, F.; Delon, J.; Gousseau, Y.; Michel, J.; Tupin, F. SAR-SIFT: A SIFT-Like Algorithm for SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 453–466. [[CrossRef](#)]
29. Kupfer, B.; Netanyahu, N.S.; Shimshoni, I. An Efficient SIFT-Based Mode-Seeking Algorithm for Sub-Pixel Registration of Remotely Sensed Images. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 379–383. [[CrossRef](#)]
30. 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](#)]
31. 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.* **2013**, *10*, 657–661. [[CrossRef](#)]
32. Shechtman, E.; Irani, M. Matching Local Self-Similarities across Images and Videos. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–8.
33. 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](#)]
34. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An efficient alternative to SIFT or SURF. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2564–2571.
35. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), San Diego, CA, USA, 20–25 June 2005; pp. 886–893.
36. Shui, P.-L.; Cheng, D. Edge Detector of SAR Images Using Gaussian-Gamma-Shaped Bi-Windows. *IEEE Geosci. Remote Sens. Lett.* **2012**, *9*, 846–850. [[CrossRef](#)]
37. Grompone von Gioi, R.; Jakubowicz, J.; Morel, J.-M.; Randall, G. LSD: A Line Segment Detector. *Image Process. Line* **2012**, *2*, 35–55. [[CrossRef](#)]
38. Meng, Z.; Han, X.; Li, X. An automatic registration method for SAR and optical images based on line extraction and control points selection. In Proceedings of the 2017 International Applied Computational Electromagnetics Society Symposium (ACES), Suzhou, China, 1–4 August 2017; pp. 1–2.

39. Sui, H.; Xu, C.; Liu, J.; Hua, F. Automatic Optical-to-SAR Image Registration by Iterative Line Extraction and Voronoi Integrated Spectral Point Matching. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6058–6072. [[CrossRef](#)]
40. Lee, J.-S. Digital Image Enhancement and Noise Filtering by Use of Local Statistics. *IEEE Trans. Pattern Anal. Mach. Intell.* **1980**, *PAMI-2*, 165–168. [[CrossRef](#)]
41. Habib, A.; Al-Ruzouq, R. Semi-automatic registration of multi-source satellite imagery with varying geometric resolutions. *Photogramm. Eng. Remote Sens.* **2005**, *71*, 325–332. [[CrossRef](#)]
42. Grompone von Gioi, R.; Jakubowicz, J.; Morel, J.M.; Randall, G. LSD: A fast line segment detector with a false detection control. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 722–732. [[CrossRef](#)]
43. Grompone von Gioi, R.; Jakubowicz, J.; Morel, J.-M.; Randall, G. On Straight Line Segment Detection. *J. Math. Imaging Vis.* **2008**, *32*, 313–347.
44. 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](#)]
45. Ji, S.; Wei, S.; Lu, M. Fully Convolutional Networks for Multisource Building Extraction From an Open Aerial and Satellite Imagery Data Set. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 574–586. [[CrossRef](#)]
46. 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](#)]

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