A Novel Fast and Robust Multimodal Images Matching Method Based on Primary Structure-Weighted Orientation Consistency

Shuo Li, Xiaolei Lv, Hao Wang, and Jian Li

***Abstract*—The matching of multimodal remote sensing images, especially high-resolution images, is a challenging task due to the nonlinear radiation distortions (NRD), the noise distribution, and the differences in structural texture differences between them. In this article, we proposed a novel fast, robust, and extensible matching method based on the primary structure-weighted orientation consistency (PSOC), which aims to extract relatively consistent primary structures and suppress texture details effectively. To construct the PSOC, we first presented a fast multiscale sigmoid Gabor filter that employs angle interpolation instead of using angle space construction. Then, we enhanced feature representation by using the local primary structure strategy and constructed the PSOC descriptor using an orientation lookup table. Finally, we used the nonmaximum suppressed 3-D normalized cross-correlation fast template matching method for the feature descriptor matching, which improved the matching success rate and reduced the matching complexity under large search radius. In experiments conducted with eight pairs of high-resolution multimodal images, the PSOC descriptor outperformed other state-of-the-art descriptors with an average improvement of 36.6% in correct match rate and an improvement of 22.7% in root mean square error (eliminating the results that these algorithms failed to match). In addition, PSOC achieves efficient matching under large search radius, and the average time complexity is about 1/3 of the other descriptors, which is important for the matching with large offsets in practical applications.**

***Index Terms*—Image matching, multimodal remote sensing images, pixel-wise dense descriptor, primary structure (PS).**

I. Introduction

MULTIMODAL remote sensing images obtained by different sensors can provide a variety of information and are widely used in image fusion [1], GCP extraction [2], change detection [3], feature information extraction [4], etc. As a prerequisite for fusion applications, the study of image matching is particularly significant. However, the differences in imaging mechanisms, nonlinear radiation distortions (NRD), noise distributions, and texture feature distributions among multimodal remote sensing images greatly increase the difficulty of matching.

The previous image matching methods can be divided into three categories: area-based matching methods, feature-based matching methods, and area-feature-based methods that combine the former. The area-based methods work by pixel values and similarity measures, such as normalized correlation methods (NCC) [5], mutual information methods [6], and frequency domain-based methods [7]. This kind of method is difficult to handle the NRD present in multimodal images. The feature-based methods attempt to mine the common features between the two images, such as the point feature [8] and line feature [9]. Scale-invariant feature transform (SIFT) and subsequent improved versions are widely used in the field of remote sensing image matching. Dellinger et al. [10] proposed the SAR-SIFT by redefining the gradient extraction part of SIFT with the ratio of exponentially weighted averages (ROEWA). To overcome the NRD and the noise differences between heterogeneous images, researchers used targeted feature descriptions to deeply optimize edge extraction in SIFT. Xiang et al. [11] proposed the OS-SIFT, which uses multiscale ROEWA for SAR and multiscale Sobel for optical in the step of gradient extraction. Zhu et al. [12] extracted highly repetitive interest points using multichannel autocorrelation of the log-Gabor detector and constructed DAISY-like feature descriptors R_2 FD_2 using rotation invariant maximum index map of the log-Gabor. Yao et al. in [13], [14], and [15] designed the histogram of absolute phase consistency gradients (HAPCG), multi-orientation tensor index feature (MoTIF), and co-occurrence filter space matching (CoFSM) based on different feature descriptions to overcome the problem of over-dependence on gradients in SIFT. However, in practical applications, SIFT-like algorithms are limited by high computational complexity and uneven distribution of matching points.

In recent years, the area-feature-based methods that combine the two methods have gradually become mainstream. They use multiple feature descriptions and template matching to mine deeper feature information to better adapt to the NRD

Manuscript received 4 June 2023; revised 11 September 2023; accepted 15 October 2023. Date of publication 20 October 2023; date of current version 31 October 2023. This work was supported by the LuTan-1 L-Band Spaceborne Bistatic SAR Data Processing Program under Grant E0H2080702. *(Corresponding author: Xiaolei Lv.)*

Shuo Li, Xiaolei Lv, and Hao Wang are with the Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, Chinese Academy of Sciences, Beijing 100094, China, also with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China, and also with the School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: lishuo193@mails.ucas.ac.cn; lvxl@aircas.ac.cn; wanghao197@mails.ucas.ac.cn).

Jian Li is with Beijing Xingtiandi Information Technology Company, Ltd., Beijing 102200, China (e-mail: jian.li@bj-xtd.com).

Digital Object Identifier 10.1109/JSTARS.2023.3325577

and improve performance. Ye et al. [16] and [17] successively proposed the histogram of orientation phase consistency (HOPC) and channel features of oriented gradients (CFOG). Zhou et al. [18] established a three-layer convolutional neural network to refine CFOG features and proposed the multiscale convolutional gradient features. Ye et al. [19] combined the first-order and second-order log-Gabor filters with multiscale strategies to construct SFOC features and used Fast-NCC$_{SFOC}$ to improve matching efficiency. Zhongli et al. [20] constructed a fast matching method named angular weighted oriented gradients (AWOG) and achieved improved matching performance. These descriptors counteract the NRD by extracting structural features. However, when the resolution increases, differences in fine structures and texture details between images become more apparent. Instead, the two images share primary structure (PS) information, such as large contours and edges. For this, we proposed the histogram of oriented primary edge structure (HOPES) [21] to match one-look SAR and optical images by a multiscale sigmoid Gabor (MSG) detector. Ye et al. [22] used relative total variation to remove texture details from the image and then extracted the main structure information by the multiscale CFOG. Descriptors based on the PS are proven advantageous.

Even though HOPES has high robustness, noise suppression ability, and matching accuracy, we have found that its time complexity is too high, and the matching performance for high-resolution urban images degrades. Therefore, this article focuses on efficiency and applicability to high-resolution multimodal images. We proposed the PS detector based on fast multiscale sigmoid Gabor filter (Fast-MSG) using angular interpolation instead of angular space construction of the original Gabor filter. Inspired by AWOG, we constructed a concise and efficient PS feature descriptor named primary structure-weighted orientation consistency (PSOC), aiming to match the multimodal high-resolution remote sensing images efficiently while maintaining robustness to noise and the NRD. By replacing the sum of squared differences (SSD) template matching commonly used for traditional descriptors with 3-DNCC fast template matching, we greatly improve the efficiency and matching performance in larger search radii. Moreover, PSOC can be extended to arbitrary gradient operators. PSOC is more like a general framework, and we have optimized its construction process to only require the edge strength maps (ESM) and Edge Direction Maps (EDM) at different scales. Therefore, Fast-MSG can be replaced with different gradient extraction operators to meet the needs of different tasks, such as Sobel [23], ROA [24], ROEWA [25], etc. The experimental results show that this method outperforms the state-of-the-art methods in terms of efficiency, correct match rate (CMR), and accuracy of multimodal image matching.

The rest of this article is organized as follows. Section II describes the proposed method. Section III provides the results and analysis of comparison and ablation experiments. Finally, Section IV concludes this article.

II. Methodology

In this section, a novel fast and robust matching method PSOC is proposed for multimodal remote sensing images. The complete PSOC matching process is shown in Fig. 1. Like the regular matching process, the PSOC matching process is divided into three parts: the extraction of keypoints (points to be matched), the construction of features, and the matching of features.

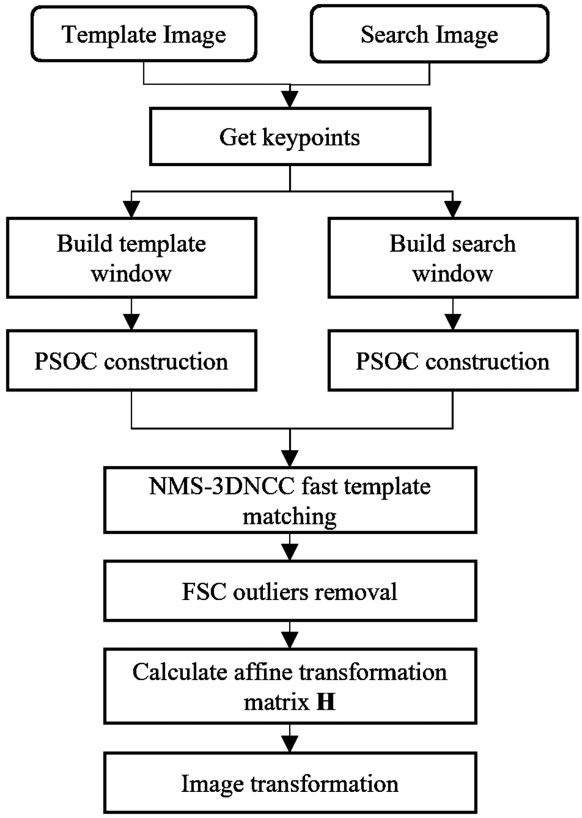

Fig. 1. PSOC matching process.

The extraction of keypoints in PSOC matching relies on the implementation of the SAR-Harris corner point extraction algorithm with nonmaximum suppression (NMS). NMS is introduced to prevent keypoint redundancy and to select only the most responsive points for matching. In the following subsections, we will present the feature construction step which involves the implementation of the fast PS extraction operator, Fast-MSG, along with the PSOC 3D density descriptor. In addition, we will describe the feature matching step, namely, the nonmaximum suppression-based 3-D fast normalized cross-correlation (NMS-3DNCC) fast template matching method.

A. PS Detector Based on Fast-MSG

Mehrotra et al. [26] demonstrated that the odd-symmetric part of the Gabor filter is an efficient and robust edge detection operator, and gave the definition of the multiscale Gabor filter as follows:

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

where ω is the frequency of the sine function and σ is the scale of the Gaussian function. As a bilateral filter, its two adjacent

windows are:

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

This leads to the ESM and EDM at the n-th scale space:

$$\begin{aligned} ESM_n &= 1 - \min_{\theta} R^{\sigma_n\theta} \\ EDM_n &= \arg\min_{\theta} R^{\sigma_n\theta} + \pi/2 \end{aligned} \tag{3}$$

where $R^{\sigma\theta} = \min(\mu_1/\mu_2, \mu_2/\mu_1)$. μ_1, μ_1 are the convolutions of $I(x,y)$ and two windows $Go_1^{\sigma\theta}$, $Go_2^{\sigma\theta}$ respectively.

To obtain the multiscale Gabor filter, we need to construct n scale-spaces by σ and angle-spaces by θ. And the algorithm complexity is $O(n^2 \cdot M^2 \cdot K^2)$, where M is the size of $I(x,y)$ and K is the size of the convolution kernel $Go(x,y)$. We use angular interpolation instead of angular space construction to obtain the ESM and EDM quickly, and the horizontal and vertical Gabor filters are obtained from (1) as follows:

$$\begin{aligned} Go^x(x,y) &= \exp\left[-\frac{x^2+y^2}{2\sigma^2}\right]\sin(\omega x) \\ Go^y(x,y) &= \exp\left[-\frac{x^2+y^2}{2\sigma^2}\right]\sin(\omega y). \end{aligned} \tag{4}$$

The adjacent windows $Go_1^{\sigma x}(x,y)$, $Go_2^{\sigma x}(x,y)$ and $Go_1^{\sigma y}(x,y)$, $Go_2^{\sigma y}(x,y)$ in the x-direction and y-direction can be obtained from (2). The horizontal component G_x^σ and the vertical component G_y^σ of the gradient at scale σ are:

$$\begin{aligned} G_x^\sigma &= \log\left(\frac{I(x,y)\otimes Go_1^{\sigma y}(x,y)}{I(x,y)\otimes Go_2^{\sigma y}(x,y)}\right) \\ G_y^\sigma &= \log\left(\frac{I(x,y)\otimes Go_1^{\sigma x}(x,y)}{I(x,y)\otimes Go_2^{\sigma x}(x,y)}\right). \end{aligned} \tag{5}$$

This leads to the ESM and EDM at the n-th scale space.

$$\begin{aligned} ESM_n &= \sqrt{(G_x^{\sigma_n})^2 + (G_y^{\sigma_n})^2} \\ EDM_n &= a\tan(G_x^\sigma / G_y^\sigma) \end{aligned} \tag{6}$$

Thus, we can obtain the Scale Edge Strength Maps (SESM) and the Scale edge direction maps (SEDM). This can be considered as a fast algorithm for (3). By angular interpolation, we reduce the time complexity to $O(n \cdot M^2 \cdot K^2)$.

As shown in Fig. 2, at a small scale, Fast-MSG performs better at localizing edges, but it is more sensitive to noise. At a wide scale, it has less edge localization capability but can suppress noise more effectively. It is generally believed that the position of low response value is noise or texture details, therefore, to extract the PS, it is necessary to reduce the value of the lower response, so we introduce a sigmoid function to extract the PS

$$\begin{aligned} \mathrm{PS}_n(x,y) &= \frac{1}{1+e^{\gamma(c-s_n(x,y))}} \\ \mathrm{PS} &= \min\{\mathrm{PS}_1, \mathrm{PS}_2, \ldots, \mathrm{PS}_n\} \end{aligned} \tag{7}$$

where c is the corresponding cutoff value of the filter, γ is the gain factor controlling the cutoff sharpness, and the larger the value of γ, the stronger the suppression ability for weak edges and noise. $s_n(x,y)$ is a normalized term

$$s_n(x,y) = \frac{1}{N}\left(\frac{\mathrm{ESM}_n(x,y)}{\varepsilon + \mathrm{ESM}_{\max}(x,y)}\right). \tag{8}$$

However, while image texture details and noise are suppressed, there is also a suppression of the PS where contrast is not apparent, which is particularly noticeable with high-resolution SAR images. Fig. 3(a) illustrates an image containing an airport and buildings. Due to the obvious radiation difference, the contrast of the runway is low, resulting in the PS map being suppressed by the bright edges of the houses, as depicted in Fig. 3(b). To mitigate this phenomenon, the local primary structure (LPS) strategy is employed by calculating the PS of local blocks to compose the final PS map, as demonstrated in Fig. 3(c).

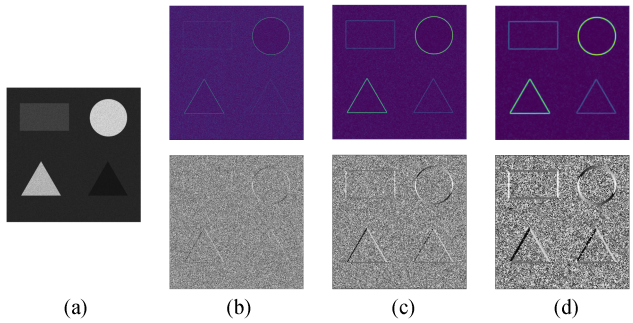

Fig. 2. ESM and EDM of different scale. (a) Simulated SAR image. (b) n=1. (c) n=3. (d) n=5.

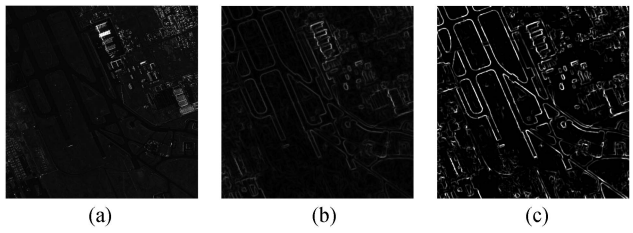

Fig. 3. Comparison of global PS strategy and local PS strategy. (a) SAR image. (b) Global PS. (c) LPS.

B. PSOC Density Descriptor

The construction flow of the PSOC descriptor is shown in Fig. 4. In the previous section, we obtained PS and SEDM by constructing Fast-MSG. The orientation map of PS is derived from the smallest scale in SEDM and is quantized into N directions by unifying it to the range of 0–π. The red box in Fig. 4 shows how the directions are divided. Due to the potential for gradient direction reversal between different modal images, the EDM's first and last orientation intervals must be unified into a single orientation interval and then transformed into the

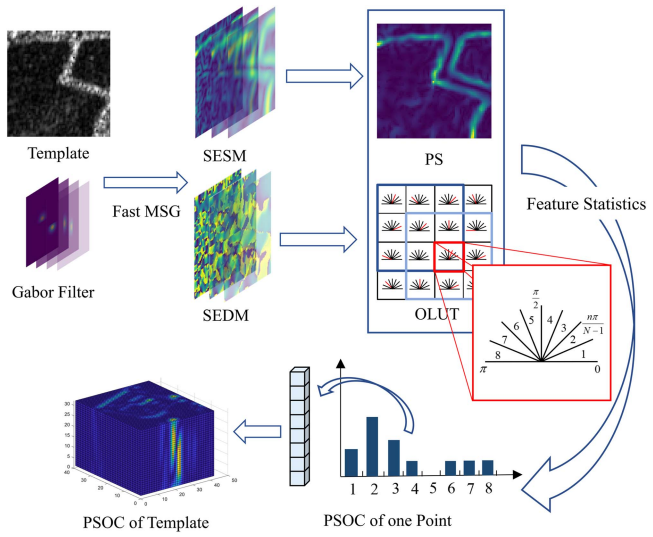

Fig. 4. Construction process of the PSOC density descriptor.

orientation lookup table (OLUT)

$$\begin{aligned}\mathrm{EDM} &= \mathrm{EDM} + \pi, \mathrm{EDM} < 0\\ \mathrm{EDM} &= \mathrm{EDM} + \frac{N-2}{N-1}\pi, \mathrm{EDM} < \frac{\pi}{N-1}\\ \mathrm{OLUT}(x,y) &= \mathrm{floor}\left(\frac{(N-1)\cdot \mathrm{EDM}(x,y)}{\pi}\right).\end{aligned} \tag{9}$$

By combining the PS map and OLUT, we establish a 3×3 neighborhood window at each point $P(x, y)$ in the PS map and determine and record the orientations of the pixels within the window. The corresponding intensities in the PS are then incorporated into the orientation histogram to derive the PSOC feature vector of point P. By repeating this process for each pixel, the full PSOC density descriptor can be obtained. Assuming the template radius is r, then the template area size is $(2r + 1) \times (2r + 1)$. Due to the use of 3×3 neighborhood, the final dimension of PSOC features is $(2r + 1) \times (2r + 1) \times N$.

As shown in Fig. 4, the construction of PSOC requires only the computation of SESM and SEDM to obtain PS and OLUT. As the research progresses, gradient operators with better performance will definitely appear in the future. At the same time, we hope that PSOC can be applied to image matching of more models. Therefore, PSOC can be extended to any multiscale edge extraction operator, meaning that it can be extended to handle larger and more complex image datasets.

C. Nonmaximum Suppression-Based 3-D Fast Normalized Cross-Correlation (NMS-3DNCC)

3-D dense descriptors, such as CFOG, AWOG, and HOPES, often use the SSD as the criterion for feature matching. However, SSD usually faces the issues as follows.

1) The SSD-based matching may fail if the image energy $\sum f^2(x, y)$ varies with position. For example, if there is a bright spot in the region to be matched, it can interfere with the matching result.

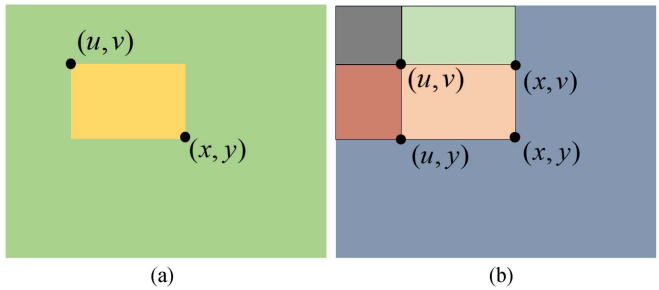

Fig. 5. Calculation of integral image. (a) Image. (b) Integral image.

2) The result of SSD matching varies with the image magnitude and is not light invariant.

As a common matching method in the field of image matching, NCC is usually used for 2-D image block matching. Lewis [27] proposed a fast NCC algorithm for matching the target image and the template image

$$r(u,v) = \frac{\sum_x \sum_y [f(x,y) - \bar{f}_{u,v}][t(x-u,y-v) - \bar{t}]}{\left\{\sum_x \sum_y [f(x,y) - \bar{f}_{u,v}]^2 \sum_x \sum_y [t(x-u,y-v) - \bar{t}]^2\right\}^{0.5}} \tag{10}$$

where $\bar{t}$ represents the mean value of the template image, and $\bar{f}_{u,v}$ represents the mean value of $f(x, y)$ in the region under the template. First, consider the numerator in the (10). Assuming that there is an image $t'(x, y) = t(x, y) - \bar{t}$, then

$$\begin{aligned}\gamma_{\mathrm{num}}(u,v) = &\sum_x \sum_y f(x,y) t'(x-u,y-v)\\ &- \bar{f}_{u,v} \sum_x \sum_y t'(x-u,y-v).\end{aligned} \tag{11}$$

Since $t'(x, y)$ has zero mean, so zero sum the term $\bar{f}_{u,v} \sum t'(x - u, y - v)$ is also zero. Thus, the numerator is

$$\gamma_{\mathrm{num}}(u,v) = \sum_x \sum_y f(x,y) t'(x-u,y-v). \tag{12}$$

The equation can be regarded as a convolution between f and $t'(-x, -y)$, so we can use the Fourier transform to convert it to a frequency domain multiplication form to speed up the operation

$$\mathcal{F}^{-1}\left\{\mathcal{F}(f)\mathcal{F}^*(t')\right\} \tag{13}$$

where F is the Fourier transform. The complex conjugate accomplishes reversal of the feature via the Fourier transform property $\mathcal{F}f^*(-x) = F^*(\omega)$.

The calculation of the denominator in (10) is accelerated by integral images, as shown in Fig. 5. A 2-D integral image is defined as follows:

$$II(x,y) = \sum_x \sum_y I(i,j) \tag{14}$$

Then, we have

$$\sum_{i=u}^{x}\sum_{j=v}^{y} I(i,j) = II(x,y) + II(u,v) - II(x,v) - II(u,v)$$
$$\sum_{i=u}^{x}\sum_{j=v}^{y} I^2(i,j) = II^2(x,y) + II^2(u,v) - II^2(x,v) - II^2(u,v). \quad (15)$$

Examining the denominator

$$\sum_x\sum_y \left(f(x,y) - \bar{f}_{u,v}\right)^2 = \sum_x\sum_y f^2(x,y) - 2\bar{f}_{u,v}\sum_x\sum_y f(x,y); + \sum_x\sum_y \bar{f}_{u,v}^2 \quad (16)$$

where

$$\sum_x\sum_y \bar{f}_{u,v}^2 = N_x N_y \left(\frac{1}{N_x N_y}\sum_x\sum_y f(x,y)\right)^2. \quad (17)$$

Therefore, (16) can be expressed as

$$\sum_x\sum_y \left(f(x,y) - \bar{f}_{u,v}\right)^2 = \sum_x\sum_y f^2(x,y) - \frac{1}{N_x N_y}\left(\sum_x\sum_y f(x,y)\right)^2. \quad (18)$$

The integral images of (15) can be used to accelerate the calculation and achieve constant time calculation independent of window radius.

Based on this, we can extend the 2-D image to 3-D features, to enable the matching of the PSOC 3-D dense descriptors. The 3-D promotion form of NCC is given by (19) shown at the bottom of the this page. The numerator part of the formula calculates the correlation between the two 3-D images at the center of the window (u, v, w). The larger this value is, the higher the similarity of the two 3-D images in this window. Similar to Fast-NCC, we can use 3-D Fourier transform $\mathcal{F}^{-1}\{\mathcal{F}(A)\mathcal{F}^*(T)\}$ to reduce the computational complexity. The denominator part of the formula is a normalization factor that is used to eliminate the effect of the brightness and contrast of the image. The calculation of the denominator can be accelerated using 3-D integral images

$$II(x,y,z) = \sum_x\sum_y\sum_z I(i,j,k). \quad (20)$$

Here, we have selected three images with high matching difficulty to test the similarity measure maps (SMM) of different descriptors. The first two subfigures of Fig. 6(a)–(c) shows the difference between 3DNCC and SSD matching methods. To make the peak comparison of different matching strategies more intuitive, we normalized the result of both matching methods. In addition, since the best match point for SSD is at the minimum peak, we subtract the normalized value from 1 to make the maximum peak the best match point. In Fig. 6(a), the bright points in the matching region cause the SSD matching to fail, while the SSD matching algorithms in Fig. 6(b) and (c) also exhibit stronger multipeakedness, which in turn reduces the confidence of the matching results. AWOG does not have noise rejection ability because it uses $dx = [-1, 0, 1]$ and $dy = [-1, 0, 1]^T$ as the edge descriptor, so it cannot accurately describe features in images with strong noise distribution, thus causing matching failure. Due to the use of the primary edge structure, HOPES is also able to match three sets of images, but the peak's sharpness is weaker than that of the PSOC descriptor, and multipeakedness is observed in Fig. 6(b). However, the peak performance of matching for CFOG and HOPC descriptors is poor, resulting in a low matching success rate.

To reduce the effect of multipeakedness and improve the matching confidence, the NMS strategy is used to select the final matched points. Here, we call it NMS-3DNCC. As shown in Fig. 7, we first select the N points (x_i, y_i) with maximum peak values as the seed points, and use the point as the top-left point (x_i^{tl}, y_i^{tl}) to make a square box. Calculate the area of the overlap between the square S_1 formed by the maximum point (x_1^{tl}, y_1^{tl}) and the square S_i formed by another seed point (x_i^{tl}, y_i^{tl}). As shown in Fig. 7(b), if the area ratio between the overlapping region and the square is larger than discriminant threshold N_t, the seed point is considered to belong to the maximum point cluster and needs to be rejected. As shown in Fig. 7(c), if the area ratio is smaller than N_t, the seed point is considered as a subpeak point. The final matched point will be determined as the primary peak point in cases where subpeak points exist and the ratio of primary to secondary peak exceeds the threshold t, or when only the primary peak is present.

III. Experiments and Results

In this section, we evaluated the proposed matching method by comparing it to three state-of-the-art methods: HOPES [21], AWOG [20], CFOG [17], HOPC [16], MoTIF [14], COFSM [15], and HAPCG [13]. At the same time, we added the PSOC with SSD to compare the effect of different matching methods on PSOC. We conducted experiments using eight pairs of multimodal images and perform ablation experiments to verify the importance of multiscale and the LPS strategy. We will provide details on the dataset, parameter settings, evaluation metrics, and results of the experiments in the following sections.

$$r(u,v,w) = \frac{\sum_{x,y,z}\left[f(x,y,z) - \bar{f}_{u,v,w}\right]\left[t(x-u,y-v,z-w) - \bar{t}\right]}{\left\{\sum_{x,y,z}\left[f(x,y,z) - \bar{f}_{u,v,w}\right]^2 \sum_{x,y,z}\left[t(x-u,y-v,z-w) - \bar{t}\right]^2\right\}^{0.5}}. \quad (19)$$

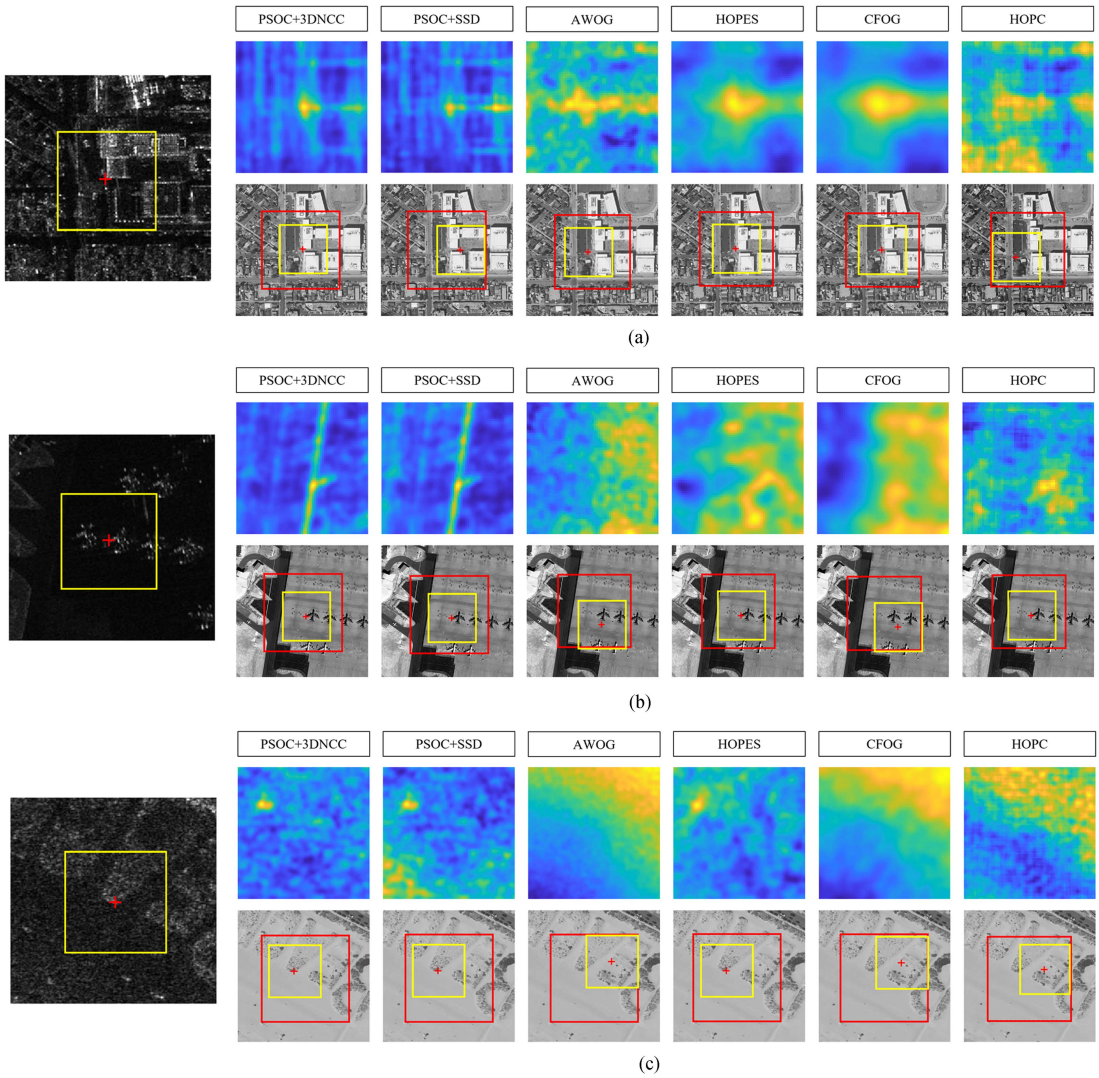

Fig. 6. SMM of PSOC (3DNCC), PSOC (SSD), AWOG, HOPES, CFOG, and HOPC.

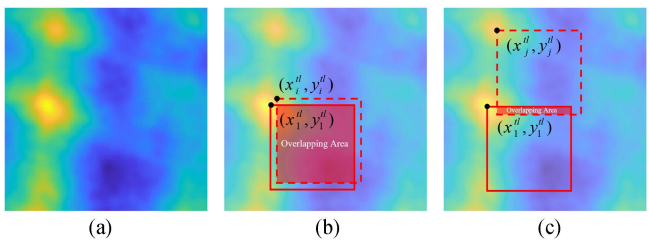

Fig. 7. Nonmaximum suppressed SSD fast template matching: (**a**) If there are repetitive structural features in the region, the SMM may have multiple peaks. (**b**) If the area ratio is larger than N_t, the seed point is rejected. (**c**) If the area ratio is smaller than N_t, the seed point is considered as a subpeak point.

A. *Datasets and Parameter Settings*

We selected eight pairs of images, as given in Table I, which include SAR, optical, Lidar intensity map, vector map, and NIR. These images cover common regions, such as cities, suburbs, and airports. All datasets have a resolution of 1 m, except for Pair H, which is limited by the resolution of the NIR images. We orthorectified the target images by rational polynomial coefficients model and digital elevation maps, and the reference images were reprojected using the same projection parameters as target images to achieve a coarse match. The coarse-matched images were then cropped to obtain the test data. At the same time, we added random rotations and nonproportional stretching to Pairs (A)–(F) to simulate the deformation distortion that may occur in the actual matching.

The experimental parameters are set as follows: the scale of Fast-MSG is set to 3, $\sigma_1 = 2$, $\sigma_i/\sigma_{i-1} = 1.6$, the filter radius is 11, and $c = 0.5$, $\gamma = 6$ in (7). The number of orientations in OLUT is 8. Same as this, the other three algorithms are set to 8 orientations, and the other parameter settings follow the recommended settings in the original paper. The template window radius is 55, and the search window radius is 15. Feature points are extracted by Harris, and the outliers are eliminated by the FSC algorithm [28]. We run our experiments on an i9-13900 K @5.4 GHz processor using MATLAB 2021B.

TABLE I
INFORMATION OF THE TEST IMAGES

Pair	Sensor	Size	Resolution	Region
A	SAR Optical	4000 × 2600	1m	Suburbs
B	SAR Optical	1700 × 1700	1m	Suburbs
C	SAR Optical	1300 × 1300	1m	Airport
D	SAR Optical	1000 × 1000	1m	Industrial Area
E	LiDAR Optical	1300 × 1300	1m	Urban
F	Map Optical	1300 × 1300	1m	Urban
G	SAR LiDAR	2200 × 2200	1m	Airport
H	SAR NIR	1500 × 1500	3m	Urban

B. Evaluation Criteria and Results

To assess the performance of the proposed matching algorithm, both subjective and objective evaluation metrics are utilized. Subjective evaluation metrics include the examination of enlarged subimages of checkerboard mosaic images. Meanwhile, objective evaluation metrics are as follows.

1) *Number of Correct Matches (NCM):* The NCM after the outlier removal step.
2) *Correct matching rate (CMR):*

$$\text{CMR} = \frac{\text{NCM}}{N_{\text{total}}} \tag{21}$$

where N_{total} is the number of keypoints, that is, the total number of points to be matched.

3) *Root mean squared error (RMSE):*

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\left(x'_i - x_i^{gt}\right)^2 + \left(y'_i - y_i^{gt}\right)^2\right)}. \tag{22}$$

We selected 10–20 pairs of corresponding points manually as the ground truth, including the keypoints (x_i, y_i) of the template image and the matched points (x_i^{gt}, y_i^{gt}) of the search image. $(x'_i, y'_i) = \mathbf{H} \times (x_i, y_i)$, $\mathbf{H}$ is the affine transformation matrix.

4) *Time:* Algorithm execution time.
5) *Time per point (TPP):*

$$\text{TPP} = \frac{\text{Time}}{N_{\text{total}}}. \tag{23}$$

It indicates the time taken to complete each match. It is important to note that SIFT-like algorithms, including MoTIF, CoFSM, and HAPCG, demonstrate a distinctive approach from other algorithms. These algorithms extract an extensive array of feature points and construct features by minimizing the feature distance violent search, resulting in matching outcomes. Consequently, we do not include their TPP values in our analysis.

The statistics are given in Table II and Fig. 9. It can be seen that in most test data, PSOC achieves the best RMSE and CMR while having higher computational efficiency. Meanwhile,

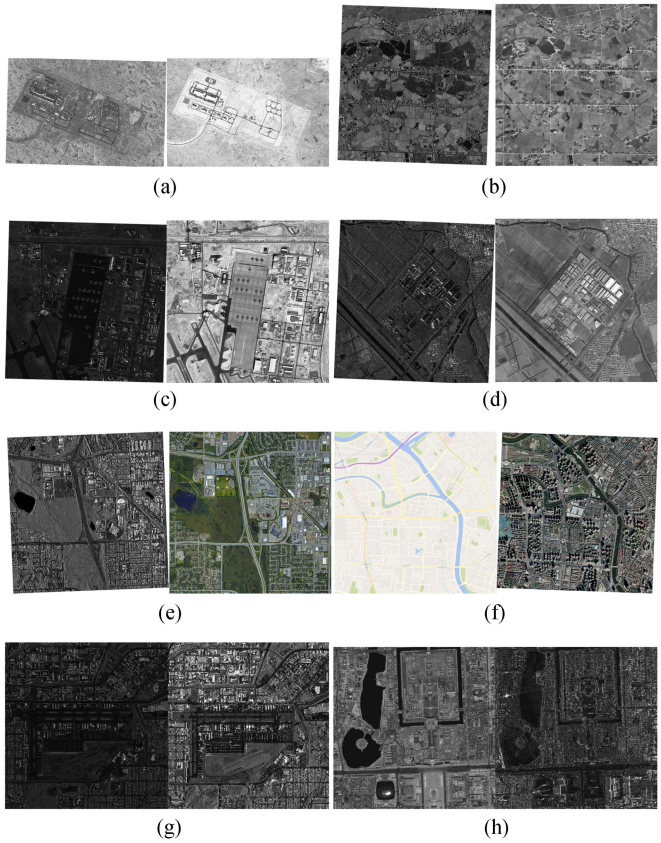

Fig. 8. Image pairs used in the experiments.

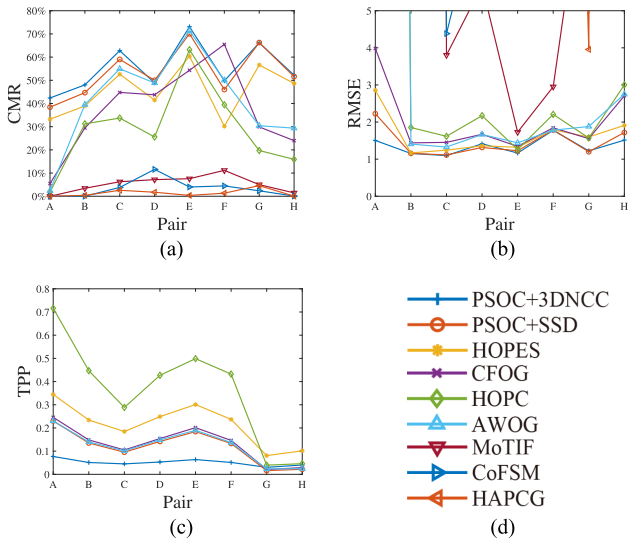

Fig. 9. CMR, RMSE, and TPP of different algorithms.

PSOC with 3DNCC has better CMR and RMSE than PSOC with SSD. When the search radius is small, SSD has lower computational complexity and a faster matching time, and in the image pairs with larger initial offsets, such as Pairs A–F. With the larger search radius, 3DNCC shows better computational complexity due to its fast algorithm achieving the computation

TABLE II
NUMBERS OF KEYPOINTS, NCM, CMR, RMSE, TOTAL TIME, AND TPP OF DIFFERENT DESCRIPTORS

Pair	Template Radius	Search Radius	Performance	**PSOC+3DNCC**	PSOC+SSD	HOPES	CFOG	HOPC	AWOG	MoTIF	CoFSM	HAPCG
A	55	65	Keypoints	328	328	328	328	328	328	\	\	\
			NCM	139	126	109	18	6	8	\	\	\
			CMR	**42.38%**	38.41%	33.23%	5.49%	1.83%	2.44%	\	\	\
			RMSE	**1.5038**	2.2249	2.8563	3.9839	108.7676	115.4398	\	\	\
			Time	25.2616	75.5171	112.9534	80.6240	234.8801	75.3718	\	\	\
			TPP	**0.0770**	0.2302	0.3444	0.2458	0.7161	0.2298	\	\	\
B	55	55	Keypoints	506	506	506	506	506	506	204	40277	59953
			NCM	243	226	197	149	158	200	7	4	229
			CMR	**48.02%**	44.66%	38.93%	29.45%	31.23%	39.53%	3.43%	0.01%	0.38%
			RMSE	**1.1534**	1.1560	1.1681	1.4387	1.8529	1.4056	\	\	99.4100
			Time	25.8414	68.3652	118.5032	75.1051	226.2043	70.7494	32.4810	431.1280	99.7840
			TPP	**0.0511**	0.1351	0.2342	0.1484	0.4470	0.1398	\	\	\
C	55	45	Keypoints	344	344	344	344	344	344	511	18813	31756
			NCM	216	203	181	154	116	189	32	717	817
			CMR	**62.79%**	59.01%	52.62%	44.77%	33.72%	54.94%	6.26%	3.81%	2.57%
			RMSE	**1.0958**	1.1138	1.2428	1.4524	1.6154	1.3272	3.8026	4.3842	7.7340
			Time	15.5285	32.8905	63.5017	36.1527	99.2375	34.8341	6.2460	116.1270	42.3180
			TPP	**0.0451**	0.0956	0.1846	0.1051	0.2885	0.1013	\	\	\
D	55	55	Keypoints	176	176	176	176	176	176	251	11775	18838
			NCM	86	88	73	77	45	86	18	1365	327
			CMR	48.86%	**50.00%**	41.48%	43.75%	25.57%	48.86%	7.17%	11.59%	1.74%
			RMSE	1.4144	**1.3148**	1.3596	1.6696	2.1719	1.6637	5.6809	8.2576	10.5677
			Time	9.3452	24.9334	43.7849	27.1348	75.1425	26.1173	16.9020	53.8560	22.6240
			TPP	**0.0531**	0.1417	0.2488	0.1542	0.4269	0.1484	\	\	\
E	65	55	Keypoints	394	394	394	394	394	394	449	26381	33689
			NCM	288	276	238	214	249	281	34	1053	124
			CMR	**73.10%**	70.05%	60.41%	54.31%	63.20%	71.32%	7.57%	3.99%	0.37%
			RMSE	**1.1649**	1.2267	1.3201	1.3454	1.2611	1.4411	1.7269	5.7524	31.2115
			Time	25.0962	72.5894	118.6082	79.3087	196.5363	74.7237	20.0490	149.0440	40.3920
			TPP	**0.0637**	0.1842	0.3010	0.2013	0.4988	0.1897	\	\	\
F	55	55	Keypoints	1037	1037	1037	1037	1037	1037	394	6250*	15739*
			NCM	516	477	313	679	409	519	44	279*	210*
			CMR	49.76%	46.00%	30.18%	**65.48%**	39.44%	50.05%	11.17%	4.46%*	1.33%*
			RMSE	**1.7776**	1.7854	1.7909	1.8439	2.2069	1.7805	2.9516	23.6154*	37.1086*
			Time	53.1711	138.0641	245.6363	151.7421	448.2640	142.2774	23.7800	29.6910*	19.6810*
			TPP	**0.0513**	0.1331	0.2369	0.1463	0.4323	0.1372	\	\	\
G	55	15	Keypoints	786	786	786	786	786	786	439	34380	41829
			NCM	520	521	445	235	155	239	22	811	1916
			CMR	66.16%	**66.28%**	56.62%	29.90%	19.72%	30.41%	5.01%	2.36%	4.58%
			RMSE	1.2229	**1.1989**	1.5973	1.5532	1.5710	1.8802	9.6678	5.9475	3.9566
			Time	23.8119	12.5776	63.3980	16.3905	31.0918	18.4254	20.7050	267.4350	55.2020
			TPP	0.0303	**0.0160**	0.0807	0.0209	0.0396	0.0234	\	\	\
H	65	15	Keypoints	446	446	446	446	446	446	274	5945*	25774
			NCM	233	230	217	107	71	131	4	5*	15
			CMR	**52.24%**	51.57%	48.65%	23.99%	15.92%	29.37%	1.46%	0.08%*	0.06%
			RMSE	**1.5125**	1.7180	1.9112	2.7086	3.0023	2.7509	\	\	\
			Time	17.9114	9.7157	45.2552	12.8698	20.5133	10.8280	18.5230	11.7150*	35.5300
			TPP	0.0402	**0.0218**	0.1015	0.0289	0.0460	0.0243	\	\	\

* Original data execution failed, downsampled to $1/2$ of the original image size.
Bold term represents the best performance.

time independent of the window radius. Nevertheless, MoTIF, CoFSM, and HAPCG exhibit erratic matching performance, are unable to complete matching in the majority of the data and consume a substantial amount of computing resources. In Pair F, we had to downsample the image by a factor of $\frac{1}{2}$ to enable the CoFSM and HAPCG algorithms to run successfully. We faced a similar challenge with the CoFSM algorithm in Pair H.

Due to space limitations, here we present only the visual matching results of two datasets. The matching results of Pair A are shown in Fig. 10. In addition, Fig. 11 presents the checkboard mosaic images and enlarged subimages of the registered images of Pair A. It can be seen that PSOC with 3DNCC has a higher matching success rate and a more uniform point distribution. Fig. 11(a) demonstrates that PSOC with 3DNCC produces consistent edge articulation at ① and ②. Conversely, PSOC with SSD and HOPES exhibit a small offset at ②, whereas CFOG produces an offset at both ① and ②. The HOPC and AWOG algorithms fail to register due to a low number of matched points. The AWOG's gradient extraction operator has low resistance to noise, making it challenging to cope with the higher noise levels present in this data, which we believe is the primary reason for its failure to match. CFOG and HOPC algorithms also face similar issues. Pair A's size means that extracting a large number of feature points and performing a violent search put a significant strain on the resources of algorithms, such as MoTIF, CoFSM, and HAPCG, which prevent them from completing successfully. As a result, we cannot provide their results.

The matching results of Pair H are shown in Fig. 12. In addition, Fig. 13 presents the checkboard mosaic images and enlarged subimages of the registered images of Pair H. Fig. 12(g)

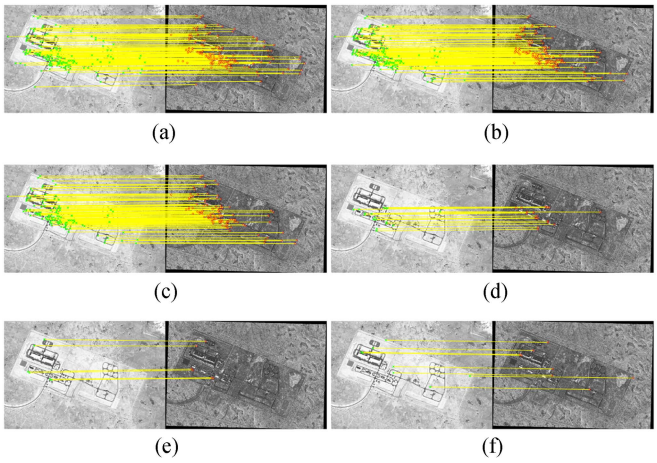

Fig. 10. Matching results of Pair A. (a) PSOC with 3DNCC. (b) PSOC with SSD. (c) HOPES. (d) CFOG. (e) HOPC. (f) AWOG.

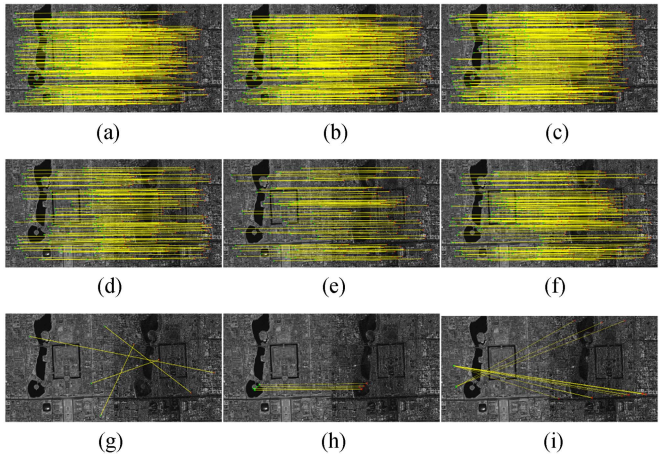

Fig. 12. Matching results of Pair H. (a) PSOC with 3DNCC. (b) PSOC with SSD. (c) HOPES. (d) CFOG. (e) HOPC. (f) AWOG. (g) MoTIF. (h) CoFSM. (i) HAPCG.

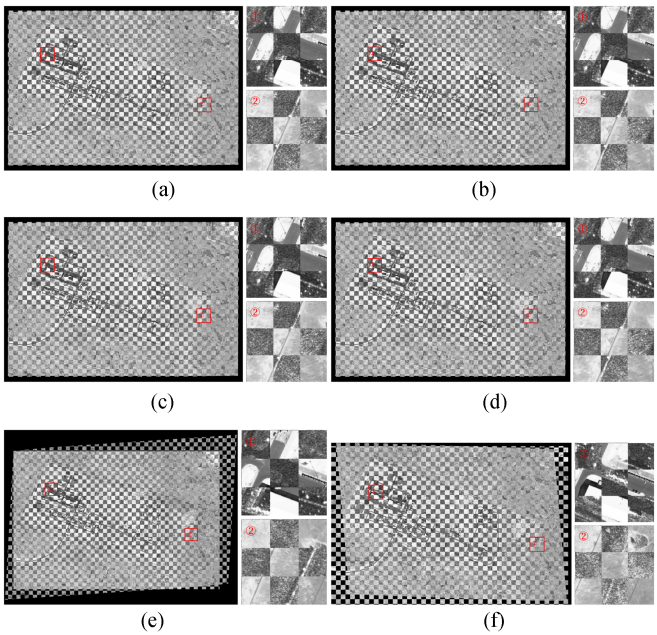

Fig. 11. Checkboard mosaic images and enlarged subimages of Pair A. (a) PSOC with 3DNCC. (b) PSOC with SSD. (c) HOPES. (d) CFOG. (e) HOPC. (f) AWOG.

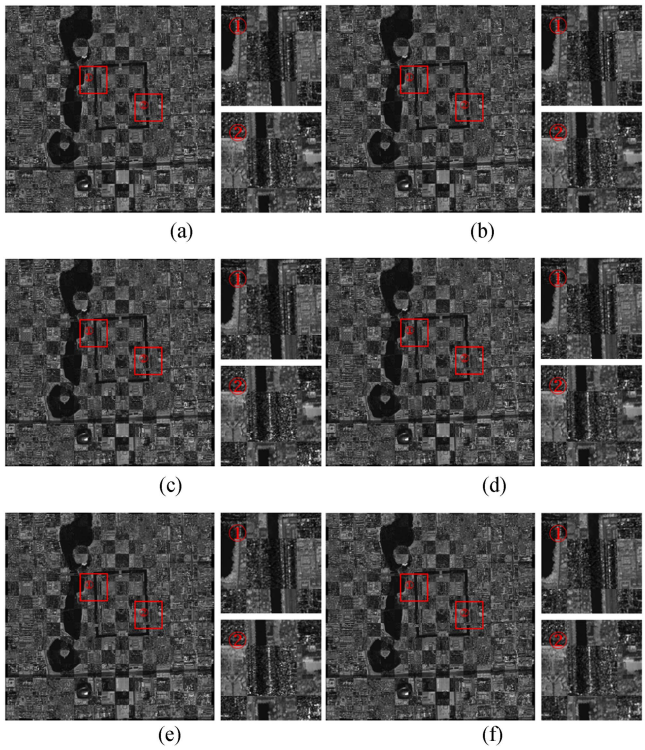

Fig. 13. Checkboard mosaic images and enlarged subimages of Pair H. (a) PSOC with 3DNCC. (b) PSOC with SSD. (c) HOPES. (d) CFOG. (e) HOPC. (f) AWOG.

and (i) reveals that the MoTIF and HAPCG matching algorithms produce erroneous results, while the distribution of CoFSM matching points is too concentrated, resulting in significant deviations in the final matching results. Hence, we do not provide the checkboard mosaic images of the three algorithms. Conversely, PSOC and HOPES exhibit the same edge articulation in their checkboard mosaic images, CFOG and AWOG produce an offset at ①, and HOPC produces an offset at ②.

We tested the performance under different template radii with Pair C, as shown in Fig. 14(a)–(c). In this experiment, to enhance the matching results with a smaller search radius, we conducted a comparison using coarse-matched data, without adding distortions, such as rotation and stretching. The performance of all descriptors improves with increasing template size, but PSOC consistently outperforms others by a substantial margin and achieves high CMR even with small template sizes. PSOC with 3DNCC has higher CMR and slightly better RMSE than PSOC with SSD, but its computational complexity is higher. PSOC with 3DNCC achieves about 66.7% CMR at a template window radius of 45, and the equivalent template radius of HOPES is 65; however, the equivalent template radius of CFOG and HOPC is about 105 to 110. Compared with PSOC, CFOG takes 3.26

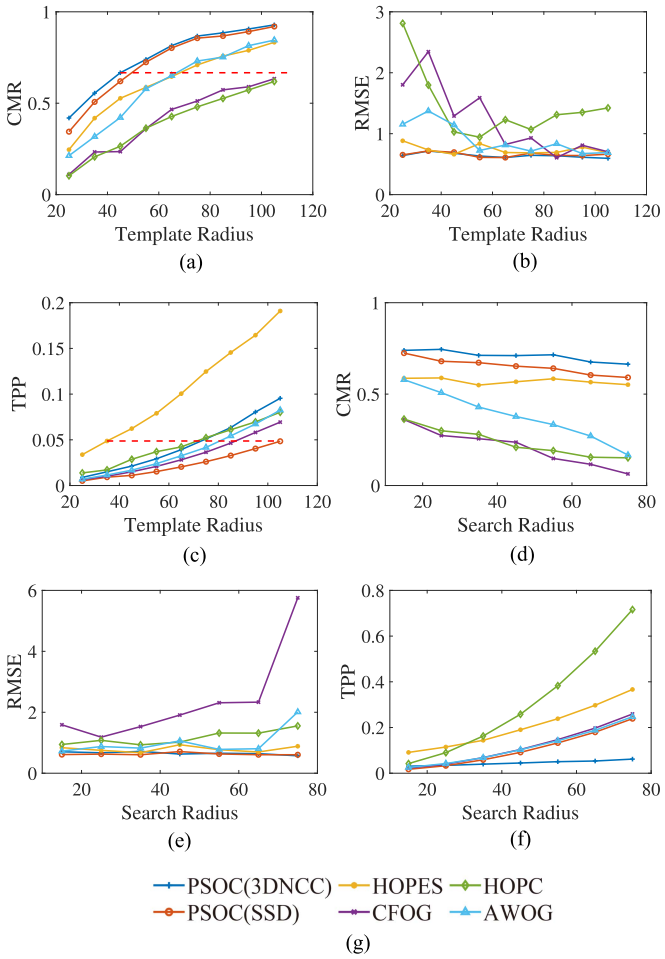

Fig. 14. (a)–(c) CMR, RMSE, and TPP of the varying template windows. (d)–(f) Search windows.

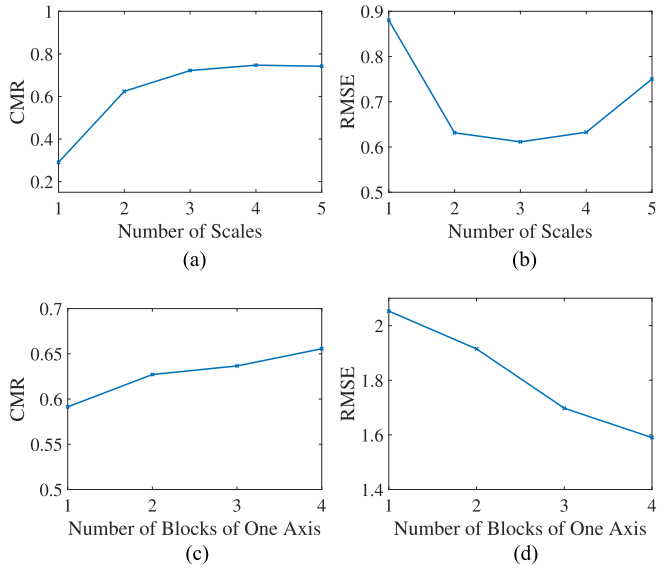

Fig. 15. Effect of scale on (a) CMR and (b) RMSE, and the effect of LPS on (c) CMR and (d) RMSE, where Number $=1$ means without multiscale or LPS.

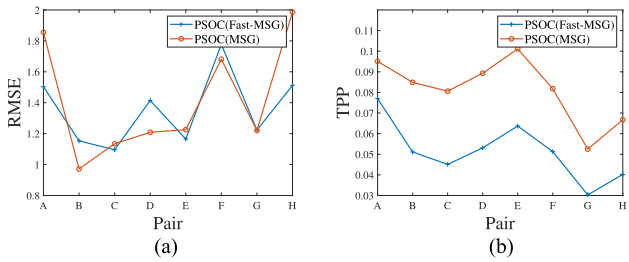

Fig. 16. Effect of angular interpolation on matching performance, where PSOC (Fast-MSG) uses angular interpolation and PSOC (MSG) does not use angular interpolation. (a) RMSE. (b) TPP.

times longer, HOPC takes 3.77 times longer, and HOPES takes 4.73 times longer. The matching time of PSOC with 3DNCC at a template window radius of 70 is equivalent to HOPES at a template window radius of 35. These all demonstrate the great efficiency benefits of PSOC.

PSOC with 3DNCC does not have a time advantage at a large template radius, its advantage is reflected in the large search radius. Fig. 14(d)–(f) illustrate the impact of different search radii. As the search radius increases, HOPC, CFOG, and AWOG exhibit significant reductions in their CMR values and gradual increases in RMSE values, with CFOG even experiencing failure at a search radius of 75. PSOC's CMR performance has been consistently high, with even better results achieved with 3DNCC as compared with SSD, while maintaining an almost consistent RMSE (or even better). More critically, the application of 3DNCC allows PSOC to maintain a very high matching efficiency under a large search radius, which is especially important in the application. This highlights the robustness, stability, and efficiency of the PSOC descriptor in comparison to other descriptors when dealing with large image offsets.

The ablation experiments on scale operation and LPS show the importance of these two works, as Fig. 15. The CMR is only 30% at scale 1 (without multiscale), and the highest matching performance is achieved at scale 3. As the scale continues to increase, the decline of feature location capability leads to a decline in matching accuracy. While in Figs. 15(c) and (d), the addition of the LPS strategy increases the CMR and decreases the RMSE.

To ascertain the impact of angle interpolation used in constructing the Fast-MSG operator on the matching accuracy, we constructed the PSOC descriptor with the MSG detector by (3). It obtains ESM and EDM by constructing scale and angle space. Next, we conducted matching tests on the data and recorded the results, as depicted in Fig. 16. Based on the experimental results, the PSOC with MSG operators achieve superior matching accuracy in four out of eight datasets. In addition, the RMSE for PSOC with MSG is only 4.05% ahead of PSOC with Fast-MSG in average. However, its matching time lags behind PSOC with Fast-MSG by 58.33% in average. Therefore, we consider the accuracy loss to be acceptable compared with the computing time gain from Fast-MSG.

IV. Conclusion

In this article, to solve the problem of fast matching between high-resolution multimodal remote sensing images, we proposed a novel multimodal matching descriptor, PSOC. The efficient implementation of PSOC is achieved through the integration of several key components, including the utilization of multiscale information, the fast extraction of multiscale PS features using the PS detector based on Fast-MSG, the improvement of feature representation through the LPS strategy, the accelerated calculation of weighted direction vectors by OLUT, and the fast descriptor matching method 3DNCC. The experiments show that PSOC has the best matching efficiency; furthermore, it effectively improves the matching success rate and accuracy. It can obtain better matching results even when the matching template is small, and it can still maintain the stability of the matching results using a large search window.

However, PSOC still faces significant challenges, such as matching high resolution SAR images, within complex urban environments involving tall buildings and undulating terrain. This poses a great challenge to existing matching algorithms and it is the next challenge to be solved.

Acknowledgment

The author would like to thank Yao et al. from Wuhan University for providing the source code of the HAPCG, MoTIF, and CoFSM algorithms, and Ye et al. from Southwest Jiaotong University for providing the source code of the CFOG and HOPC algorithms.

References

[1] S. C. Kulkarni and P. P. Rege, "Pixel level fusion techniques for SAR and optical images: A review," *Inf. Fusion*, vol. 59, pp. 13–29, 2020.

[2] T. Bürgmann, W. Koppe, and M. Schmitt, "Matching of terrasar-x derived ground control points to optical image patches using deep learning," *ISPRS J. Photogrammetry Remote Sens.*, vol. 158, pp. 241–248, 2019.

[3] L. Wan, T. Zhang, and H. You, "Multi-sensor remote sensing image change detection based on sorted histograms," *Int. J. Remote Sens.*, vol. 39, no. 11, pp. 3753–3775, 2018.

[4] F. Tupin and M. Roux, "Detection of building outlines based on the fusion of SAR and optical features," *ISPRS J. Photogrammetry Remote Sens.*, vol. 58, no. 1-2, pp. 71–82, 2003.

[5] W. Shi, F. Su, R. Wang, and J. Fan, "A visual circle based image registration algorithm for optical and SAR imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2012, pp. 2109–2112.

[6] S. Suri and P. Reinartz, "Mutual-information-based registration of terrasar-x and ikonos imagery in urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 2, pp. 939–949, Feb. 2010.

[7] C. D. Kuglin, "The phase correlation image alignment methed," in *Proc. Int. Conf. Cybern. Soc.*, 1975, pp. 163–165.

[8] L. Yu, D. Zhang, and E.-J. Holden, "A fast and fully automatic registration approach based on point features for multi-source remote-sensing images," *Comput. Geosci.*, vol. 34, no. 7, pp. 838–848, 2008.

[9] X. Shi and J. Jiang, "Automatic registration method for optical remote sensing images with large background variations using line segments," *Remote Sens.*, vol. 8, no. 5, 2016, Art. no. 426.

[10] F. Dellinger, J. Delon, Y. Gousseau, J. Michel, and F. Tupin, "Sar-sift: A sift-like algorithm for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 453–466, Jan. 2015.

[11] Y. Xiang, F. Wang, and H. You, "OS-SIFT: A robust sift-like algorithm for high-resolution optical-to-SAR image registration in suburban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3078–3090, Jun. 2018.

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

[13] Y. Yao, Y. Zhang, Y. Wan, X. Liu, and H. Guo, "Heterologous images matching considering anisotropic weighted moment and absolute phase orientation," *Geomatics Inf. Sci. Wuhan Univ.*, vol. 46, no. 11, pp. 1727–1736, 2021.

[14] Y. Yao, B. Zhang, Y. Wan, and Y. Zhang, "Motif: Multi-orientation tensor index feature descriptor for SAR-optical image registration," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 43, pp. 99–105, 2022.

[15] Y. Yao, Y. Zhang, Y. Wan, X. Liu, X. Yan, and J. Li, "Multi-modal remote sensing image matching considering co-occurrence filter," *IEEE Trans. Image Process.*, vol. 31, pp. 2584–2597, 2022.

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

[17] Y. Ye, L. Bruzzone, J. Shan, F. Bovolo, and Q. Zhu, "Fast and robust matching for multimodal remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9059–9070, Nov. 2019.

[18] L. Zhou, Y. Ye, T. Tang, K. Nan, and Y. Qin, "Robust matching for SAR and optical images using multiscale convolutional gradient features," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2021, Art. no. 4017605.

[19] Y. Ye, B. Zhu, T. Tang, C. Yang, Q. Xu, and G. Zhang, "A robust multimodal remote sensing image registration method and system using steerable filters with first-and second-order gradients," *ISPRS J. Photogrammetry Remote Sens.*, vol. 188, pp. 331–350, 2022.

[20] F. Zhongli, Z. Li, W. Qingdong, L. Siting, and Y. Yuanxin, "A fast matching method of SAR and optical images using angular weighted orientated gradients," *Acta Geodaetica et Cartographica Sinica*, vol. 50, no. 10, 2021, Art. no. 1390.

[21] S. Li, X. Lv, J. Ren, and J. Li, "A robust 3D density descriptor based on histogram of oriented primary edge structure for SAR and optical image co-registration," *Remote Sens.*, vol. 14, no. 3, 2022, Art. no. 630.

[22] Y. Ye, C. Yang, J. Zhang, J. Fan, R. Feng, and Y. Qin, "Optical-to-SAR image matching using multiscale masked structure features," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6509405.

[23] I. E. Sobel, *Camera Models and Machine Perception*. Stanford, CA, USA: Stanford Univ., 1970.

[24] R. Touzi, A. Lopes, and P. Bousquet, "A statistical and geometrical edge detector for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 6, pp. 764–773, Nov. 1988.

[25] R. Fjortoft, A. Lopes, P. Marthon, and E. Cubero-Castan, "An optimal multiedge detector for SAR image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 3, pp. 793–802, May 1998.

[26] R. Mehrotra, K. R. Namuduri, and N. Ranganathan, "Gabor filter-based edge detection," *Pattern Recognit.*, vol. 25, no. 12, pp. 1479–1494, 1992.

[27] J. P. Lewis, "Fast template matching," in *Proc. Vis. Interface*, Quebec City, QC, Canada, 1995, vol. 95, no. 120123, pp. 15–19.

[28] Y. Wu, W. Ma, M. Gong, L. Su, and L. Jiao, "A novel point-matching algorithm based on fast sample consensus for image registration," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 43–47, Jan. 2015.

Shuo Li received the B.S. degree in electronic and information engineering from Chongqing University, Chongqing, China, in 2019. He is currently working toward the Ph.D. degree in signal and information processing with the Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China.

He is currently with the University of Chinese Academy of Sciences, Beijing. His current research interests include remote sensing image registration and satellite photogrammetry.

Xiaolei Lv received the B.S. degree in computer science and technology and the Ph.D. degree in signal processing from Xidian University, Xi'an, China, in 2004 and 2009, respectively.

From 2009 to 2010, he was with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. From 2011 to 2013, he was with the Department of Civil and Environmental Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA. Since April 2013, he has been a Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. His main research interests include sparse signal processing, radar imaging (synthetic aperture radar/inverse synthetic aperture radar), interferometric synthetic aperture radar, and ground moving-target indication.

Hao Wang received the B.S. degree in telecommunication engineering from Shandong University, Jinan, China, in 2019. He is currently working toward the Ph.D. degree in signal and information processing with the Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China.

His current research interests include remote sensing image 3-D reconstruction and change detection.

Jian Li received the B.S. degree in engineering survey from Liaoning Technical University, Fuxin, China, in 1998, the M.S. degree in photogrammetry and remote sensing from the Chinese Academy of Geological Science, Beijing, China, in 2001, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2005.

He is currently with Beijing Xingtiandi Information Technology Co., Ltd., Beijing. His current research interests include remote sensing image 3-D reconstruction and photogrammetry.