MSRSI-TPMF: A Tie Points Matching Framework of Multisource Remote Sensing Images

Qian Cheng, Xin Li, Taoyang Wang, Boyang Jiang, He Fu, Yunming Wang, and Feida Zhang

***Abstract*—Remote sensing sensor platforms are typically located at a significant distance from the ground, ranging from several hundred meters to hundreds of kilometers. This means that, compared to natural images, remote sensing images (RSI) have larger coverage areas and more complex information. The larger size and data volume of RSI presents challenges for computer vision matching algorithms (MAs), making it difficult to apply them directly to RSI matching. Moreover, a matching framework for multisource RSIs capable of large-scale processing by integrating multiple MAs with the entire RSI as input is presently lacking. This study proposes a tie points (TPs) matching framework of multisource remote sensing images based on the geometric and radiation characteristics of RSI. First, RSI is divided into different grids and undergoes local geometry correction. Next, matching between slice images is performed by MAs. Finally, TPs are generated by mapping matched points in multiple slice images to the whole RSI using a geometric processing model. Six representative MAs including artificial feature MAs and deep learning algorithms are integrated into the framework to match TPs from different RSI. Results demonstrate the extraction of TPs for multisource RSI, validating the framework's efficacy. In addition, a large-scale TPs matching test for deep learning MA is performed by using 13 synthetic aperture radar images (10-m resolution) with TPs root mean square error of 0.368 pixels, further confirming the framework's reliability.**

***Index Terms*—Block adjustment (BA), framework, image matching, remote sensing, tie points (TPs).**

I. Introduction

REMOTE sensing image (RSI) matching is a basic task in RSI processing, which refers to the process of mapping an image to another image or images of the same scene acquired under different conditions such as phase, angle, and illumination through spatial transformation and establishing spatial correspondences between the two or more images [1]. RSI matching, as the core foundation of visual processing and understanding tasks such as RSI stitching [2], fusion [3], 3D reconstruction [4], 4D reconstruction [5], [6], change detection and target localization, plays an important role in the fields of natural disaster emergency response and damage assessment. With the rapid development of remote sensing technology, sensors of different platforms can provide abundant data for human observation of the Earth, such as multi/hyperspectral, infrared, and synthetic aperture radar (SAR) [7]. How to effectively process and analyze multisensor, multiresolution, and multitemporal remote sensing data has become a hot research topic in the field of remote sensing, and multisource RSI matching is one of the core problems that need to be solved urgently [8].

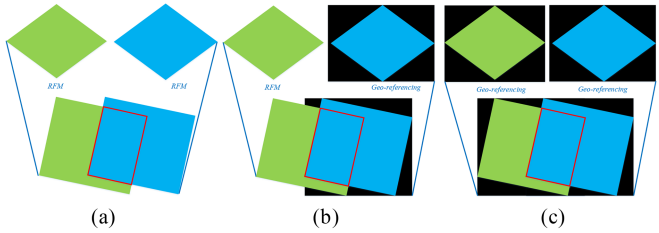


Fig. 1. Three matching cases of RSI matching. (a) Case 1: RFM-RFM. (b) Case 2: RFM-Geo-referencing. (c) Case 3: Geo-referencing-Geo-referencing.

In fact, the matching of RSIs can be divided into three categories according to their geographical attributes (see Fig. 1). Case 1 involves nongeocoded images (L1 data) with the rational polynomial model (RFM) model. In Case 2, one scene of the image is geocoded (L2 data), and the other is equipped with the RFM model. Finally, in Case 3, both images are geocoded.

Case 1 refers to the matching of RSI tie points (TPs). Automatically and reliably obtaining accurate and uniformly distributed TPs and performing block adjustment (BA) calculation is a prerequisite for applications such as high-precision mapping, 3-D information extraction, and 3-D urban model construction by using RSIs [9]. The purpose of matching in Case 2 and Case 3 is to rectify the images. Due to the phenomenon of degraded positioning model in the participating images, it becomes impossible to recover the three-dimensional

Manuscript received 11 August 2023; revised 10 October 2023; accepted 1 November 2023. Date of publication 8 November 2023; date of current version 19 December 2023. This work was supported by the Foundation Strengthening Fund Project under Grant 2021-JCJQ-JJ-0251. *(Qian Cheng and Xin Li are co-first authors.) (Corresponding author: Taoyang Wang)*

Qian Cheng, Taoyang Wang, and Yunming Wang are with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China (e-mail: cheng_qian@whu.edu.cn; wangtaoyang@whu.edu.cn; wangyunming@whu.edu.cn).

Xin Li and Boyang Jiang are with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: lixin_1995@whu.edu.cn; jiangboyang@whu.edu.cn).

He Fu is with the ZhongSe Lantu Technology Company, Ltd., Beijing 101300, China (e-mail: 178395625@qq.com).

Feida Zhang is with the School of Earth Science and Resources, China University of Geosciences, Beijing 100083, China (e-mail: zhangfeida@email.cugb.edu.cn).

Our framework is available at https://github.com/TobyChengV1/MSRSI-TPMF.

Digital Object Identifier 10.1109/JSTARS.2023.3331251

information. Currently, although there is extensive research on RSI matching, it mainly focuses on algorithm innovation and the tested images are usually small in size. Besides, the applicability to Case 1, Case 2, and Case 3 has not been explained.

In existing research on regional image processing, such as SAR image DOM production [10], optical image block network adjustment [11], [12], and night light image DOM production [13], although matching algorithms (MAs) for TPs (such as SIFT, SAR-SIFT) are mentioned, the details are not described in detail. Stable MAs like SIFT, SAR-SIFT, etc., are commonly used in the engineering processing of multisource RSIs. However, the improvement strategies vary in different studies, and it is necessary to modify the applicability of the algorithms according to different data types.

In conclusion, the matching of TPs in RSIs faces several challenges.

1) There is a lack of a unified strategy for TP matching that can be applied to multisource RSIs. Existing methods have low generality and different data require different strategies.
2) Advanced algorithms are difficult to directly apply to TP matching in RSIs. Existing methods mainly focus on algorithm innovation, and the tested images are often small size, without considering the issue of large size and wide-range RSIs. There is a gap between visual and RSIs, which limits the application of advanced algorithms in the field of remote sensing.
3) Deep learning matching methods are currently limited to the theoretical research stage, and there exists a significant gap between their development and practical engineering applications.

To address these challenges, this article proposes a TPs matching framework for multisource RSIs (MSRSI-TPMF). The framework adopts a grid division and local geometric correction strategy, which is not only suitable for multisource RSIs but also effectively reduces the impact of scale and rotation differences on matching. The matching problem of wide-range RSI is converted into multiple slice image matching using a quantified module, allowing access to different computer vision MAs. Finally, the final TPs are obtained by projecting the slice image homonymous points mapping onto the whole image using a geometry processing model.

The main contributions of this article include the following three aspects.

1) In this article, RFM is selected as an effective means to connect different images. By dividing the grid, the spatial distribution of TPs is ensured. Combined with geometry correction, it reduces the impact of scale and rotation differences on matching. This approach is applicable to multisource RSIs.
2) This article proposes a complete framework for remote sensing TP matching. By decomposing the TP matching process and incorporating a matching module, different MAs can be directly applied to TP matching in RSIs.
3) For the first time, this article applies deep learning matching methods to large-scale engineering, validating the potential of engineering applications.

TABLE I
STATISTICAL RESULTS OF EXPERIMENTAL IMAGE SIZE BY DIFFERENT METHODS

Method	Year	Size(pixel)	Test data
literature [32]	2022	256×256	satellite–aerial image pairs
literature [33]	2022	256×256	GF3 FSII
literature [34]	2022	1240×1400	AHB dataset
literature [35]	2022	600×600	Multimodal Remote Sensing
literature [36]	2022	13056×11008	Google Earth
literature [37]	2022	800×800	UAV, PAN, SAR, CIAP
literature [31]	2022	512×512	Multimodal Remote Sensing
literature [38]	2023	750×750	Multimodal Remote Sensing
literature [39]	2023	27620×29200	Jilin-1, Gaofen-1, Gaofen-2
literature [40]	2023	855×831	Multimodal Remote Sensing

II. RELATED WORK

A. Image Matching Method

RSI matching originated from image matching in traditional computer vision, the differences between them are becoming increasingly smaller. Currently, a substantial amount of research has been conducted on both image matching and RSI matching. In general, image-matching methods can be categorized into three groups [14], [15], intensity-based methods (IBM), feature-based methods (FBM), and learning-based methods (LBM).

IBMs find the correspondence between images by optimizing the similarity metrics in the corresponding window pairs of images [16]. The similarity metrics include the normalized cross correlation [17], the mutual information [18], etc.

FBMs register images by extracting local features of key points and calculating the distance of features, such as scale-invariant feature transform (SIFT) [19], speeded up robust features [20], and oriented fast and rotated brief [21], channel features of orientated gradients (CFOG) [22], histogram of the orientation of weighted phase [23], and steerable filters of first- and second-order channels [24].

LBMs are initially applied only to extract advanced features to obtain keypoints and feature descriptions, such as learning invariant feature transform [25], deep local feature (DELF) [26], detection and description network [27], SuperGlue [28], and multiscale convolutional gradient features [29]. With the development of technology, end-to-end learning methods have been proposed. End-to-end learning methods aim to directly predict the match points. For example, cross-modal feature description matching network [30], multiscale framework with unsupervised learning [31], etc.

In current RSI matching, slice images are commonly used for experimentation. However, in practical applications, wide-range RSIs need to be processed. There still exists a significant gap between MAs and practical applications. Table I provides

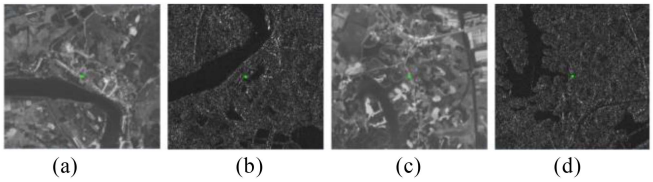


Fig. 2. Example of rotation and scale difference. (a) and (c) Tianhui-1 5-m resolution. (b) and (d) Gaofen-3 UFS 3-m resolution.

statistics on the sizes of experimental images used in some RSI matching papers from 2022.

B. Strategies for RSI Matching

In the case of RSI TPs matching (Case 1), scholars have predicted homologous points' location based on epipolar constraints to achieve fast image matching [41], [42]. However, the standard epipolar line may not apply to all satellite images. For example, the epipolar line of linear push-room RSI are similar to hyperbolas, and SAR images do not have standard epipolar line due to their special slant-range imaging characteristics [43]. Therefore, the image matching based on the epipolar constraint becomes quite complex, and it is difficult to be used among different data sources. Another matching strategy is the pyramid image matching method, which involves a process from coarse to fine, resulting in high reliability. However, generating pyramid images for RSI, which are generally GB-level data, is time-consuming and can affect matching efficiency [44]. Additionally, the conventional coarse-to-fine matching strategy relies heavily on the distribution of initial matching points and matching accuracy [45]. The block search strategy based on RFM is another method that uses the four corners of a rectangular block from the reference image to narrow down the corresponding search area on another image [10]. However, differences in rotation and scale caused by variations in view angles in the image window can pose challenges to matching, as shown in Fig. 2.

III. Method

MSRSI-TPMF is proposed in this article, considering that the computer vision MAs are difficult to directly apply to RSI matching. The framework divides the RSI TPs matching problem into three modules, grid division and geometry correction, image matching, and postprocessing. The computer vision MA can be applied to RSI TPs matching by image matching module. The workflow of the proposed framework is shown in Fig. 3.

A. Grid Division and Geometry Correction for RSI

1) Overlapping Area Acquisition: Overlap in geospatial location is a prerequisite between image pairs involved in image matching. However, the L1 image data itself does not contain geographic information, and a geometric processing model is needed for spatial conversion. First, the four corners of the matching image are projected onto the ground to obtain the corresponding geographic polygon. Then, the intersection polygons

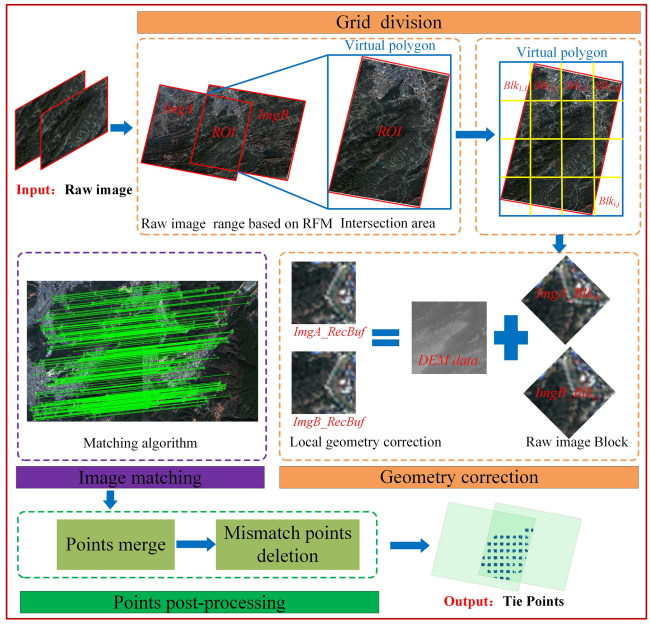


Fig. 3. Workflow of the proposed framework.

(denoted as ROI) between the matching images were obtained by combining space intersection. In particular, RFM [46] can be constructed according to the auxiliary file of rational polynomial coefficients (RPC) provided by L1 image products, which is defined as follows:

$$\begin{cases} x = \dfrac{\mathrm{Num}_L(U,V,W)}{\mathrm{Den}_L(U,V,W)} = \dfrac{\sum_{i=0}^{3}\sum_{j=0}^{3}\sum_{k=0}^{3} p_{1ijk}U^iV^jW^k}{\sum_{i=0}^{3}\sum_{j=0}^{3}\sum_{k=0}^{3} p_{2ijk}U^iV^jW^k} \\ y = \dfrac{\mathrm{Num}_S(U,V,W)}{\mathrm{Den}_S(U,V,W)} = \dfrac{\sum_{i=0}^{3}\sum_{j=0}^{3}\sum_{k=0}^{3} p_{3ijk}U^iV^jW^k}{\sum_{i=0}^{3}\sum_{j=0}^{3}\sum_{k=0}^{3} p_{4ijk}U^iV^jW^k} \end{cases} \tag{1}$$

where intermediate coordinates (x,y) and (U,V,W) are obtained by normalizing the image coordinates (l,s) and ground coordinates (B,L,H) to enhance the stability of mathematical calculations using RFM. p_{1ijk}, p_{2ijk}, p_{3ijk}, p_{4ijk} ($i = 1, 2, 3$; $j = 1, 2, 3$; $k = 1, 2, 3$) are RPCs.

Here, (l,s) and (B,L,H) can be converted into each other through positive RFM and inverse RFM [46]. The calculation method is as follows:

$$\begin{cases} (l,s,H) = h^{+}_{\mathrm{RFM}}(B,L) \\ (B,L,H) = h^{-}_{\mathrm{RFM}}(l,s) \end{cases} \tag{2}$$

where h^{+}_{RFM} is positive RFM, h^{-}_{RFM} is inverse RFM.

It is well known that scale differences will affect the matching results. Therefore, geometric correction is adopted to correct the matching images to the uniform resolution due to the difference in resolution between different images. Image resolution can be obtained by projecting two image points onto the ground and calculating the ratio of the ground distance to the image distance. The corrected resolution $\mathrm{IRS}_{\mathrm{Final}}$ is calculated as follows:

$$\mathrm{IRS}_{\mathrm{Final}} = \mathrm{Min}\left[\mathrm{IRS}_1, \ldots, \mathrm{IRS}_M\right] \tag{3}$$

where M represents the number of matching images. Min means to get the minimum value. $M \geq 2$.

2) Grid Division: Considering the large size of the whole RSI, it is usually processed into blocks. Moreover, the spatial distribution of matching points (TPs) between images will affect the BA results. The method first obtains the minimum enclosing rectangle corresponding to ROI, which is called virtual polygon (VP). Then the VP is divided to get several blocks. The width and height of VP are calculated as follows:

$$\begin{cases} VP\_W = (\text{long\_max} - \text{long\_min}) \ / \text{IRS}_{\text{Final}} \\ VP\_H = (\text{lat\_max} - \text{lat\_min}) \ / \text{IRS}_{\text{Final}} \end{cases} \tag{4}$$

where VP_W and VP_H represent the width and height of VP respectively. long_max and long_min represent the maximum and minimum longitude of ROI. lat_max and lat_min indicate the maximum and minimum latitude ROI, respectively.

The number of grid divisions in the column direction and row direction can be calculated by given parameters of block pixel width (Block*X*) and block pixel height (Block*Y*).

$$\begin{cases} nBLK\_X = VP\_W/BlockX \\ nBLK\_Y = VP\_H/BlockY \end{cases} \tag{5}$$

where *n*BLK_*X* and *n*BLK_*Y* represent the number of grid divisions in the column direction and row direction, respectively.

Blocks are divided based on VP, so there is a possibility that blocks may not intersect with the ROI, it is necessary to determine block's effectiveness. We define the block in row *j* and column *i* as $\text{BLK}_{i,j}$, where, $i \in [1, \text{nBLK\_X}]$, $j \in [1, \text{nBLK\_Y}]$.

The geographical range of $\text{BLK}_{i,j}$ can be transformed according to the row and column coordinates, [long_min, lat_max], and $\text{IRS}_{\text{Final}}$. According to the geographical range of $\text{BLK}_{i,j}$, its corresponding geographical polygon is easy to construct and then overlapping rate can be calculated

$$\text{ORATE}_{\text{BLK}_{i,j}} = \frac{\text{Area}\left(\text{Intersect}\left(\text{BLK}_{i,j}, ROI\right)\right)}{\text{Area}\left(\text{BLK}_{i,j}\right)} \tag{6}$$

where $\text{ORATE}_{\text{BLK}_{i,j}}$ represents the overlapping rate of $\text{BLK}_{i,j}$ Intersect $(*)$ represents space intersection, and Area $(*)$ represents polygon area calculation.

The block effectiveness can be determined as follows:

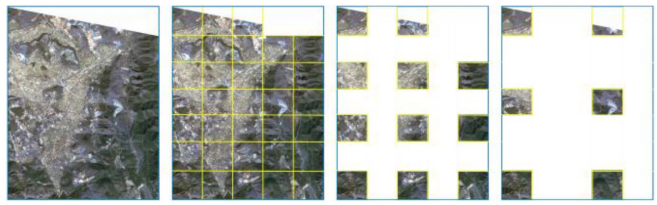
$$\begin{cases} if\ \text{ORATE}_{\text{BLK}_{i,j}} \geq \alpha, & \text{BLK}_{i,j}^{\text{FLAG}} = 1 \\ else & \text{BLK}_{i,j}^{\text{FLAG}} = 0 \end{cases} \tag{7}$$

where α represents the overlapping rate threshold of block, $\text{BLK}_{i,j}^{\text{FLAG}}$ is the attribute of the block in row *j* and column *i*.

If block meets the condition $\text{BLK}_{i,j}^{\text{FLAG}} = 1$, defined as BLK_{i_g,j_g},$g \in [1, K]$, $i_g \in [1, n\text{BLK\_X}]$,$i_g \in [1, n\text{BLK\_Y}]$. Where BLK_{i_g,j_g} represents the *g*th filtered block, that is, the block in row i_g and column j_g, *K* represents the number of blocks after filtering.

In fact, properly reducing the number of grids can reduce the amount of computation while the space can be evenly distributed. Fig. 4 shows the result that step is set as 1, 2, 3.

3) Geometry Correction: According to the BLK_{i_g,j_g} and its corresponding geographical range, combined with DEM data and the $\text{IRS}_{\text{Final}}$, geometric correction is carried out for the corresponding regions. The relationship of geometric correction can be simplified as (8). The details of geometric correction can be found in [47]



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

Fig. 4. Schematic of different steps. (a) VP. (b) Step = 1. (c) Step = 2. (d) Step = 3.

$$\begin{cases} ImgA\_RecBuf\ \left[\text{BLK}_{i_g,j_g}\right] \\ \quad = \mathcal{L}\left(\text{BLK}_{i_g,j_g}, \text{IRS}_{\text{Final}}, \text{ImgA}, \text{DEM}\right) \\ ImgB\_RecBuf\ \left[\text{BLK}_{i_g,j_g}\right] \\ \quad = \mathcal{L}\left(\text{BLK}_{i_g,j_g}, \text{IRS}_{\text{Final}}, \text{ImgB}, \text{DEM}\right) \end{cases} \tag{8}$$

where ImgA_$RecBuf[\text{BLK}_{i_g,j_g}]$ and ImgB_$RecBuf$ $[\text{BLK}_{i_g,j_g}]$ represent the ImgA and ImgB block corresponding to block BLK_{i_g,j_g} after geometric correction, respectively. $\mathcal{L}$ represents the transformation relationship of geometric correction.

B. Image Matching

The wide-range RSI has been converted into multiple slice images through grid division and geometric correction. Consequently, different MAs can be integrated to achieve the extraction of the homologous points between slice images. The relationship is as follows:

$$LMP\ \left[\text{BLK}_{i_g,j_g}\right] = \mathcal{M}\left(\text{ImgA\_}RecBuf\ \left[\text{BLK}_{i_g,j_g}\right], \text{ImgB\_}RecBuf\ \left[\text{BLK}_{i_g,j_g}\right]\right) \tag{9}$$

where $LMP[\text{BLK}_{i_g,j_g}]$ indicates matched points in Block BLK_{i_g,j_g}. $\mathcal{M}$ represents the MA. The result of matched points here is the pixel coordinate of the corresponding block without geographic information, it needs to be converted to the original RSI through postprocessing with the help of RFM.

C. Points Postprocessing

The initial matching points of the whole image can be obtained by merging the matching points of each block, and the relationship is as follows:

$$CMP = \sum F\ \left(LMP\left[\text{BLK}_{i_g,j_g}\right]\right) \tag{10}$$

where *CMP* is the initial matching points of the whole image, and Σ represents the merging operation of all the blocks. *F* is the functional relationship of pixel coordinates mapped from the block to the whole image, as shown in Fig. 5.

The relationship of pixel coordinates mapped from the block to the whole image can be divided into three steps. Step 1, converting pixel coordinates of blocks to pixel coordinates of VP. Step 2, longitude and latitude coordinates corresponding to pixel coordinates in Step 1 are calculated based on VP range, [long_min, lat_max], and $\text{IRS}_{\text{Final}}$, and DEM data is utilized

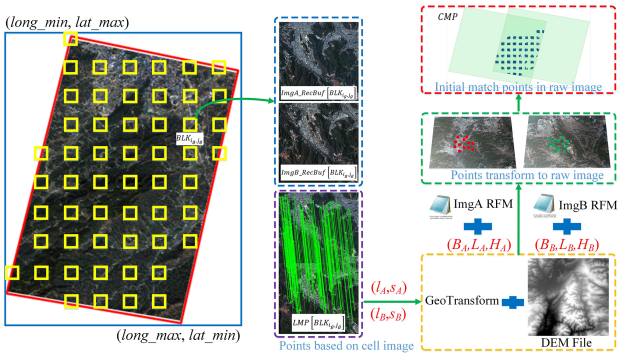


Fig. 5. Schematic of pixel coordinates mapped from block to the whole image.

to complete the conversion from block pixel coordinates to three-dimensional geodetic coordinates. Step 3, with the help of RPC files, the inverse RFM in (2) can be used to obtain the 3-D geodetic coordinates in Step 2 corresponding to pixel coordinates on the original image.

Although the MA in the image matching module will eliminate the mismatched points, the relation is only satisfied in the local space because the matching is based on slice images. Now that, the whole image should be taken as the object to eliminate mismatched points, the relationship is as follows:

$$FMP = P(CMP) \tag{11}$$

where *FMP* represents the point after global mismatched elimination, and *P* represents the mismatched point elimination model.

Based on the least square principle, this article compensates for the systematic error of the RFM through the geometric constraint relationship between images, which can be described as

$$\begin{cases} l + \Delta l = l + a_0 + a_1 l + a_2 s = F_l\ (B, L, H) \\ s + \Delta s = s + b_0 + b_1 l + b_2 s = F_s\ (B, L, H) \end{cases} \tag{12}$$

where (l,s) represents the image coordinate of homologous points. (a_0, a_1, a_2) and (b_0, b_1, b_2) are the compensation parameters. (B, L, H) is the ground coordinates of the corresponding homologous points on the ground, which can be obtained by directly solving the model parameters after adjustment [48], [49], [10]. Finally, the root mean square error (RMSE) of all points is calculated as follows:

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{n} \left[(l_i - T(l_i', s_i'))^2 + (s_i - T(l_i', s_i'))^2 \right] / N} \tag{13}$$

where $(l_i,\ s_i)$ and $(l_i',\ s_i')$ are the sets of homologous points involved in calculation, respectively. *T* represents the transformation model, here is the geometric constraint relationship compensation based on RFM. *N* represents the number of points participating in the calculation.

According to (12) and (13), combined with the iterative solution strategy, the mismatched point elimination model *P* of (11) is obtained.

Here, the systematic error of RFM compensation model (12) was solved based on the initial points, then all point residuals were calculated, and the point whose residual was greater than the given threshold was iteratively eliminated. Until the condition is satisfied, and the RMSE is calculated. In this article, we called BA mismatched point elimination (BAMPE). The given threshold is set to 1.5 pixels as default in this article.

Algorithm 1: MSRSI-TPMF Algorithm.

Input: the RSI *ImgA*, *ImgB* and DEM data
Output: final tie points *FMP*
Parameter: Block pixel width (*BlockX*), block pixel height (*BlockY*), overlapping rate threshold α
Calculate intersection polygons ROI and its corresponding virtual polygon VP
Calculate IRS_{Final} (3).
Calculate VP_W and VP_H (4).
Calculate $nBLK\_X$ and $nBLK\_Y$ (5).
Repeat:
  Extract $BLK_{i,j}$.
  Calculate $ORATE_{BLK_{i,j}}$ (6).
  if $ORATE_{BLK_{i,j}} \geq \alpha$
    $BLK_{i,j}^{FLAG} = 1$ marked as BLK_{i_g,j_g}
    geometry correction (8).
    image matching (9).
  end
points merge (10).
mismatched points eliminate (11).

D. TPs Matching for Multiple Images

Step 1, Initial task construction. Select two images out of *N* images and combine them, denoted as $\mathrm{TSK}_{i,j}^{\mathrm{FLAG}=0}$, where $i \in [1,N]$, $j \in [1,N]$, and $i \neq j$. FLAG $\in [0, 1]$ is the task flag whose initialization value is 0. The initial number of tasks is TskCount, which is determined by the following equation:

$$\mathrm{TskCount} = C_N^2 = \frac{N(N-1)}{2}. \tag{14}$$

Step 2, Optimization of tasks based on spatial relationships. Select the *i*th and *j*th image ($\mathrm{TSK}_{i,j}^{\mathrm{FLAG}}$) and judge whether they intersect in space. If an overlap relationship exists, set $\mathrm{TSK}_{i,j}^{\mathrm{FLAG}=1}$. Spatial overlap calculations involve projecting the four corner points of the master image onto elevation surface Z to obtain ground points, which are then reprojected onto the slave image to perform polygon intersection. A schematic diagram of this process is shown in Fig. 6.

Step 3, TP matching is carried out for the tasks in (2) that meet the conditions of $\mathrm{TSK}_{i,j}^{\mathrm{FLAG}=1}$.

Step 4, Perform BA for all images based on the matched TPs.

IV. Experiments and Analysis

In order to verify the feasibility of MSRSI-TPMF proposed in this article, four sets of experiments were designed for analysis and verification.

1) Quantitative experiments were conducted to explore the impact of scale and rotation differences on matching.

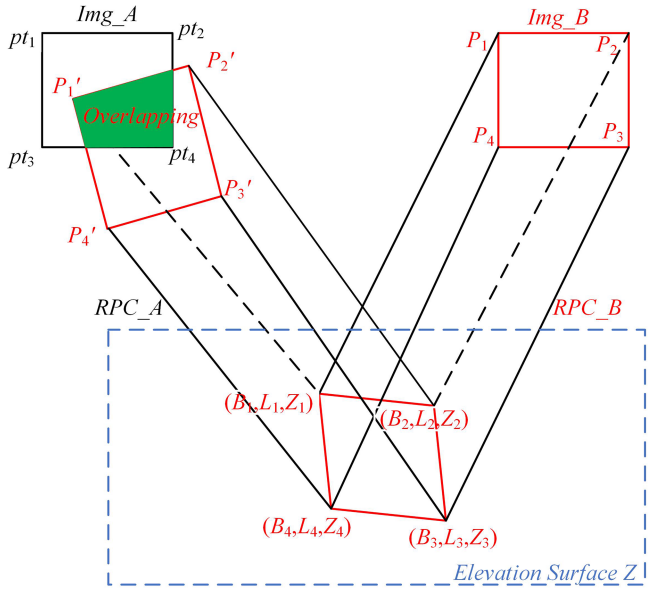


Fig. 6. Overlapping rate calculation method based on RFM.

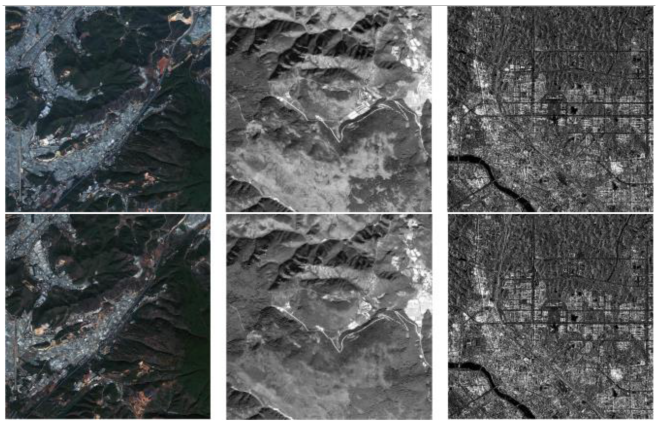


Fig. 7. Test Data 1 slice images. (a) Pair-A (900 × 900). (b) Pair-B (1024 × 1024). (c) Pair-C (2000 × 2000).

2) Different mismatched points elimination methods were performed and their effects were analyzed.
3) A variety of different MAs were used to match the TPs, verifying the generality of the framework.
4) Multiscene RSI was organized to match TPs in order to verify the engineering processing capability of the framework.

A. Description of Experimental Datasets

Test Data 1: Three sets of remote sensing slice image pairs, including Gaofen-2 (GF-2), Tianhui-1 (TH-1), and Gaofen-3 (GF-3), were geometry corrected and adopted for the experiment, and the information shown in Table II and Fig. 7.

Test Data 2: Two groups of image pairs, including GF-2 and Gaofen-7 (GF-7), were used to test the influence of different mismatched point elimination methods. The image information is shown in Table III.

Test Data 3: It includes SAR (GF-3) image pair in Japan and optical /SAR (GF-7/GF-3) images in Yunnan province. The framework was adopted to conduct the TPs matching experiment by different MAs. Table IV is the image information.

Test Data 4: In this study, GF-3 SAR images in Hubei Province and Gaofen-1 (GF-1) optical images in Hebei Province were used as the test data to verify the potential of this framework in large-scale area applications. The spatial distribution of images is shown in Fig. 8.

Hubei Province test area: The test area was located between 28.87° and 33.79° latitude, 108.21° and 112.35° longitude. It is a mountainous/hilly region in the west of Hubei Province, China. The elevation range is 26 to 3090 m. Thirteen images were available for the area. The acquisition times were from October 2017 to November 2017.

Hebei Province test area: The test area was located between 40.34° and 42.11° latitude, 114.62° and 116.28° longitude. It is a mountainous region in the north of Hebei Province, China. The elevation range is 204 to 2288 m. Seventeen images were available for the area. The acquisition times were from June 2017 to June 2020.

TABLE II
TEST DATA 1 IMAGES INFORMATION

pairs	Date	Size	Image Source	GSD/m	Location
pair-A	2019.01.23/2019.10.31	900×900	GF-2-MSS	4.00	hilly
pair-B	2015.10.21/2015.10.21	1024×1024	TH-1 PAN	5.00	hilly
pair-C	2017.12.23/2018.01.21	2000×2000	GF-3 FSII	10.00	urban

TABLE III
TEST DATA 2 IMAGE INFORMATION

Data	Date	Size	Image Source	GSD/m	Location
GF-2	2019.01.23	7300×6908	GF-2-MSS	4.00	hilly, urban
	2020.03.12	7300×6908			
GF-7	2021.04.14	31267×30996	GF-7-FWD	0.80	hilly, urban
	2021.04.14	35863×40009	GF-7-BWD	0.65	

TABLE IV
TEST DATA 3 IMAGE INFORMATION

Data	Date	Size	Image Source	GSD/m	Location
GF-3	2017.12.23	28427×26636	GF-3-FSII	10.00	urban
	2018.01.21	27383×25420			
GF-7/GF-3	2020.08.26	8967×10033	GF-7 MUX	2.80	mountain
	2021.12.04	16886×24973	GF-3 UFS	3.00	

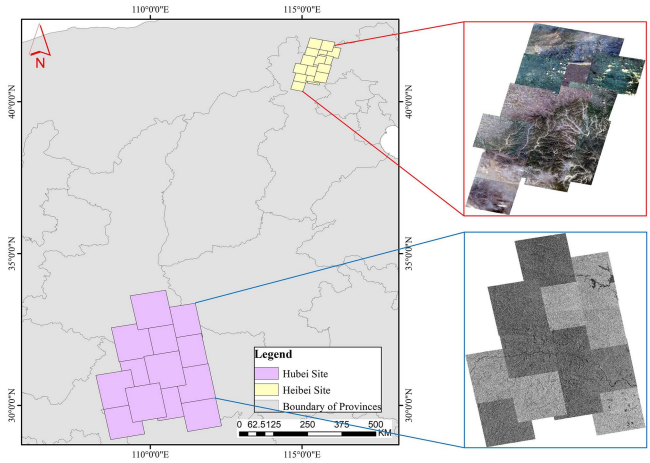


Fig. 8. Distribution of Hubei Province and Hubei Province test area.

B. Influence of Scale and Rotation to Image Match

The pair-A, pair-B, and pair-C in Test Data 1 were performed with two times, three times, and four times of down sampling (DS) respectively. Four plans were used to compare the result of image match by DS images. The details of the plans were as follows.

Plan 1, no scale and rotation difference experiment, the original image 1 (Size) and image 2 (Size) were DS by k times (k = 23,4), and the DS image 1 (Size/k) and DS image 2 (Size/k) were matched.

Plan 2, scale difference experiment, matching the original image 1 (Size) and the DS image 2 (Size/k) in Plan 1.

Plan 3, rotation difference experiment, matching the DS image 1 (Size/k) in Plan 1 and the rotated DS image 2 (Size/k + Rotation) in Plan 1 (15° counterclockwise rotation in this article).

Plan 4, scale and rotation difference experiment, matches the original image 1 (Size) and the rotated DS image 2 (Size/k + Rotation) in Plan 3.

SIFT method and SAR-SIFT method, as representative FBMs in the field of optical and SAR image matching, are widely used for optical and SAR image matching [10], [11], [12]. In this experiment, pair-A and pair-B were matched by the SIFT algorithm, SAR-SIFT algorithm was used for pair-C (SAR image). The number of initial matches (NIM), number of the correct matches (NCM), and RMSE of the experimental results were counted respectively, then the corresponding correct matching ratio (CMR) is calculated. In this experiment, the match within 1.5 pixels was defined as the correct match [24]. CMR was defined as CMR = NCM/NIM, the RMSE can be calculated according to (13). The difference is that T is an affine transformation model. Table V shows the statistical results.

In Table V, red marks indicate that this index is the best. For NCM and CMR, the larger the value is, the better the effect is. On the contrary, the smaller the value is, the higher the accuracy is for RMSE. Table V shows that the best results can be obtained in NCM, CMR, and RMSE when there is no scale and rotation difference (Plan1). To analyze the impact of different plans on the matching results, the best and worst times of NCM, CMR, and RMSE in pair-A, pair-B, and pair-C for different plans under different DS scales were calculated (the sum of the best and worst times of each index is 9). The results are shown in Table VI.

The statistical results in Table VI showed that the best times of NCM, CMR, and RMSE are 9, 8, and 6, respectively, when there is no scale and rotation difference (Plan1). At the same time, the worst statistics of NCM, CMR, and RMSE are all 0, indicating that the best matching results can be obtained without scale and rotation differences. In addition, it can be seen that when there is a scale and rotation difference (Plan4), the statistics of the worst times of NCM, CMR, and RMSE are 9, 9, and 5, respectively. It shows that the matching result will be worse when scale and rotation difference exists at the same time. In terms of RMSE, the best time in the scale and rotation difference (Plan4) experiment is 3, and the worst time is 5, showing an abnormal phenomenon. Moreover, it shows that the best results of RMSE all appeared in pair-C. According to the results in Table V, the NCM statistical values of pair-C at different DS scales (k = 23,4) are 41, 14, and 6, respectively. The reliability of the results is difficult to guarantee due to a few points participating in the calculation of RMSE. Pair-C is matched by the SAR-SIFT algorithm, which indirectly indicates that the SAR-SIFT algorithm has poor robustness to scale and rotation.

In brief, it shows that when there are scale and rotation differences among the images involved in matching, even if a more robust SIFT algorithm is adopted, the matching accuracy will still be reduced. Therefore, effective reduction of the impact of scale and rotation can improve the accuracy of matching and the number of matching points.

C. Influence of Different Mismatch Points Method on BA

Grid division strategy is adopted for image matching and it can achieve good results on local blocks in this article. However, the whole image may not meet the optimal results. It is necessary to eliminate the mismatched TPs based on the whole image to improve the matching accuracy. In this experiment, the SIFT algorithm was selected to match and obtain initial points (IPs) due to the use of GF-2 and GF-7 optical images. The results of IPs, BAMPE method, quadratic polynomial (QPMPE) method, and random sample consensus (RANSAC) method were counted. Here, QPMPE is performed by fitting the QPMPE coefficient based on the IPs, and the pixel whose residual is greater than the given threshold is iteratively eliminated. Until the condition is satisfied, and the RMSE is calculated. The process is the same as (11), except that the transformation model is a QPMPE model. The fitting model of RANSAC is the perspective model. In Section IV-B, the images involved in matching are slice images that reduce the difference in rotation and scale, therefore, an affine transformation is selected and the threshold is set at 1.5 pixel [24]. The difference is that the QPMPE model was selected and the iterative threshold was added to 3 pixels in this experiment based on the wide-range RSIs' characteristic. To maintain generality, the RANSAC threshold is also set to 3.

TABLE V
STATISTICS OF RESULTS OF DIFFERENT PLAN IN TEST DATA 1

DS	Pair	Plan	Size (R represents 15°rotation)	NIM/points	NCM/points	CMR/%	RMSE/pixels
2 times	pair-A	Plan1	450×450/450×450	341	254	74.5	0.473
		Plan2	900×900/450×450	421	191	45.4	0.662
		Plan3	450×450/450×450, R	264	178	67.4	0.526
		Plan4	900×900/450×450, R	345	119	34.5	0.681
	pair-B	Plan1	512×512/512×512	1088	864	79.4	0.498
		Plan2	1024×1024/512×512	1315	623	47.4	0.645
		Plan3	512×512/512×512, R	803	594	74.0	0.542
		Plan4	1024×1024/512×512, R	1014	367	36.2	0.719
	pair-C	Plan1	1000×1000/1000×1000	1525	351	23.0	0.865
		Plan2	2000×2000/1000×1000	1008	122	12.1	0.881
		Plan3	1000×1000/1000×1000, R	761	153	20.1	0.823
		Plan4	2000×2000/1000×1000, R	487	41	8.4	0.774
3 times	pair-A	Plan1	300×300/300×300	89	60	67.4	0.626
		Plan2	900×900/300×300	279	38	13.6	0.658
		Plan3	300×300/300×300, R	71	46	64.8	0.646
		Plan4	900×900/300×300, R	303	23	7.6	0.705
	pair-B	Plan1	341×341/341×341	509	380	74.7	0.527
		Plan2	1024×1024/341×341	829	172	20.7	0.708
		Plan3	341×341/341×341, R	401	295	73.6	0.543
		Plan4	1024×1024/341×341, R	695	116	17.0	0.737
	pair-C	Plan1	666×666/666×666	1357	342	25.2	0.851
		Plan2	2000×2000/666×666	674	51	7.6	0.860
		Plan3	666×666/666×666, R	594	151	25.4	0.787
		Plan4	2000×2000/666×666, R	211	14	6.6	0.749
4 times	pair-A	Plan1	225×225/225×225	85	64	75.3	0.573
		Plan2	900×900/225×225	218	22	10.1	0.601
		Plan3	225×225/225×225, R	57	41	71.9	0.671
		Plan4	900×900/225×225, R	207	9	4.3	0.596
	pair-B	Plan1	256×256/256×256	320	247	77.2	0.505
		Plan2	1024×1024/256×256	642	103	16.0	0.760
		Plan3	256×256/256×256, R	218	168	77.1	0.556
		Plan4	1024×1024/256×256, R	539	61	11.3	0.784
	pair-C	Plan1	500×500/500×500	836	215	25.7	0.806
		Plan2	2000×2000/500×500	350	25	7.1	0.875
		Plan3	500×500/500×500, R	363	84	23.1	0.769
		Plan4	2000×2000/500×500, R	119	6	5.0	0.413

TABLE VI
NUMBER OF BEST/WORST STATISTICS OF DIFFERENT PLANS/TIMES

index/Plan	Plan1		Plan2		Plan3		Plan4	
	best	worst	best	worst	best	worst	best	worst
NCM	9	0	0	0	0	0	0	9
CMR	8	0	0	0	1	0	0	9
RMSE	6	0	0	3	0	1	3	5

TABLE VII
STATISTICS OF RESULTS OF DIFFERENT PLAN IN TEST DATA 2

Data	Plan	Number of points	CMR (%)	BA result (pixel)		
				Max XY	RMSE XY	change
GF-2	IPs	13162	—	86.957	1.853	—
	BAMPE	13143	99.86	1.202	0.373	1.480(↓)
	QPMPE	12504	95.00	1.080	0.360	1.493(↓)
	RANSAC	6810	51.74	1.034	0.351	1.502(↓)
GF-7	IPs	15158	_	19.056	0.343	—
	BAMPE	15153	99.97	1.197	0.280	0.063(↓)
	QPMPE	3115	20.55	1.115	0.274	0.069(↓)
	RANSAC	175	1.15	0.797	0.211	0.132(↓)

Fig. 9 shows the distribution of different mismatch point removal methods on GF-2 and GF-7 data.

According to Table VII, the accuracy of the BA of IPs is 1.853 and 0.343 pixels on GF-2 data and GF-7 data, respectively. The result shows the effectiveness of the block-matching method in this article. However, the existence of some mismatched points leads to large maximum errors, such as the maximum XY error of GF-2 data and GF-7 data reaching 86.957 and 19.056 pixels, respectively. The accuracy of BA is improved by using BAMPE, QPMPE, and RANSAC to eliminate mismatched points. Although QPMPE and RANSAC can eliminate mismatched points, the CMR is low. Especially for GF-7

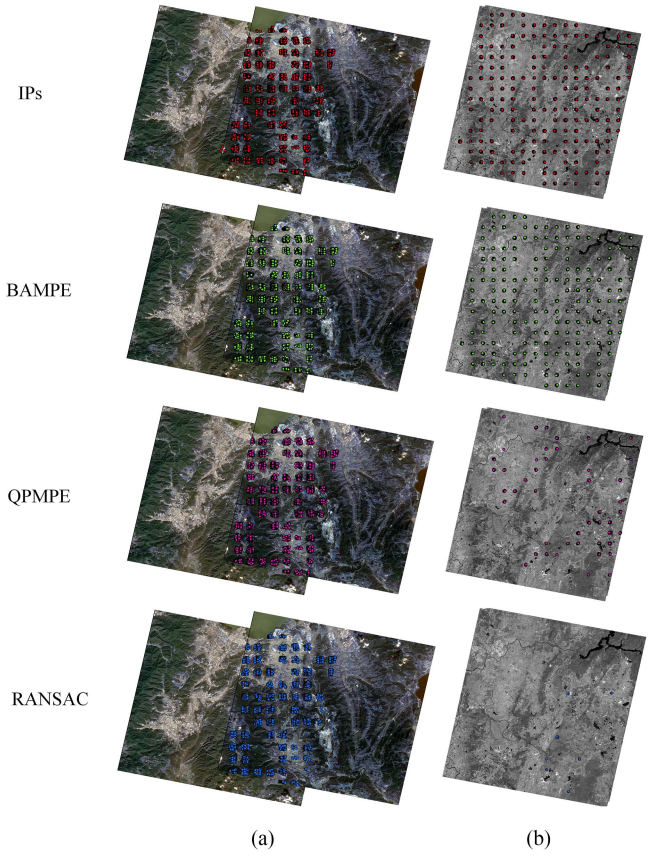


Fig. 9. TPs distribution of different mismatch point removal methods. (a) GF-2 Data. (b) GF-7 Data.

data, the CMR is only 20.55% and 1.15%, respectively. Low CMR is difficult to ensure spatial distribution, as shown in Fig. 9. The CMR of the BAMPE is 99.86% and 99.97% in the GF-2 data and GF-7 data, respectively. Compared with the QPMPE method and RANSAC method, only a few points need to be deleted in the initial matching to obtain better matching accuracy, which reflects the effectiveness of the BAMPE method.

From the perspective of the data source, the image angle of both fields in GF-2 data is 3.45°, and its deformation can be simply replaced by a QPMPE in a wide-range image. However, the angles of the forward and backward cameras equipped with GF-7 are $+26°$ and $-5°$, respectively [50]. At this time, the deformation of the image is not limited to the plane, and the QPMPE method cannot completely fit its deformation.

Further analysis shows that the BAMPE method is based on the combination of the RFM model and affine transformation, which is essentially a high-order polynomial model. Although low-order polynomials can replace high-order models to some extent, there will be an underfitting phenomenon. Briefly, the method of eliminating mismatched points in this article can ensure the number of matching points and accuracy. Considering that the framework adopts the grid division strategy, which can ensure a sufficient number of points and the uniform distribution of points in space. Besides, it found that it is extremely important to select a suitable model to eliminate the mismatched points

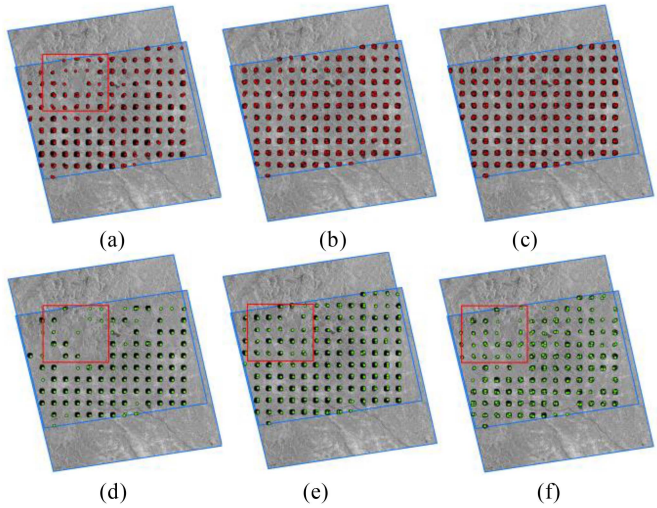


Fig. 10. TPs results by different methods for GF-3 images in Japan. (a) SIFT. (b) SAR-SIFT. (c) SARPointNet. (d) CFOG. (e) RIFT. (f) MoTIF.

through this experiment. When the intersection angle of the images participating in the matching is large, the low-order polynomial is difficult to effectively fit the projected geometric deformation, and the high-order model needs to be considered.

D. Framework Generality Verification

In this section, six representative MAs (SIFT [19], SAR-SIFT [33], SARPointNet [51], CFOG [22], RIFT [52], and MoTIF [53]) including artificial feature MA and deep learning algorithm were integrated with the framework first. Then different MAs were performed TPs matching from different RSI in Test Data 3. The results of different methods were evaluated qualitatively and quantitatively at the same time. Here, the SIFT method is a representative MA for optical images. SAR-SIFT method is a representative MA for SAR images. SARPointNet method is the first algorithm to apply deep learning-based matching methods to the field of SAR matching. CFOG and RIFT algorithms are high-cited algorithms with citation frequencies of 175 and 187, respectively. MoTIF is the latest algorithm published in 2022.

1) GF-3 Image Matching Experiment in Japan: The matching TPs results of the GF-3 image in Japan are shown in Table VIII. Fig. 10 shows the spatial distribution of points. Red marks indicate that this index is the best in Table VII. According to NCM, it is shown in Table VIII that SIFT, SAR-SIFT, SARPointNet, CFOG, RIFT, and MoTIF can effectively obtain enough TPs, which demonstrates the effectiveness of those methods. Besides, RIFT can get the most NIM and NCM, CFOG has the highest CMR, SARPointNet has the highest initial accuracy of 1 pixel, and after removing mismatched points, CFOG has the highest accuracy of 0.250 pixels. However, the implementation of the algorithm comes from the open source website and there are three coding methods (C++, Python, and MATLAB). It cannot be directly applied to wide-range RSI matching. Combined with the framework proposed in this article, different algorithms and different coding methods can effectively serve RSI TPs matching, which verifies the generality

TABLE VIII
STATISTICS OF TPS RESULTS BY DIFFERENT METHOD FOR GF-3 IMAGES IN JAPAN

method	code	NIM	NCM	CMR(%)	IPs RMSE (Pixel)			BAMPE RMSE (Pixel)		
					X	Y	XY	X	Y	XY
SIFT	C++	4459	4282	96.03	5.253	3.738	6.447	0.266	0.262	0.374
SAR-SIFT	C++	5988	5757	96.14	3.084	1.293	3.344	0.425	0.423	0.600
SARPointNet	Python	13389	10500	78.42	0.739	0.716	1.029	0.526	0.519	0.739
CFOG	matlab	14719	14193	96.42	1.979	1.806	2.679	0.206	0.141	0.250
RIFT	matlab	39031	32808	84.05	0.926	0.783	1.213	0.524	0.538	0.751
MoTIF	matlab	16420	15491	94.34	2.296	2.751	3.584	0.262	0.252	0.364

TABLE IX
STATISTICS OF TP RESULTS BY DIFFERENT METHODS FOR GF-7/GF-3 IMAGES IN YUNNAN PROVINCE

method	code	NIM	NCM	CMR(%)	IPs RMSE (Pixel)			BAMPE RMSE (Pixel)		
					X	Y	XY	X	Y	XY
CFOG	matlab	8831	6953	78.73	4.712	5.476	7.224	0.524	0.381	0.648
RIFT	matlab	11142	2904	26.06	8.427	11.334	14.124	0.537	0.507	0.739
MoTIF	matlab	2320	380	16.38	42.282	52.980	67.784	0.423	0.380	0.569

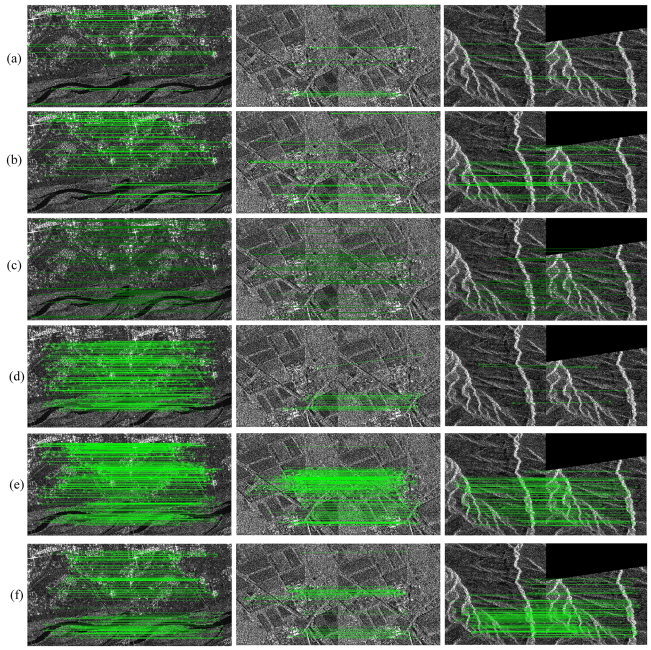


Fig. 11. Details of matching results by different algorithms in Japan. (a) SIFT (b)SAR-SIFT. (c) SARPointNet. (d) CFOG. (e) RIFT. (f) MoTIF.

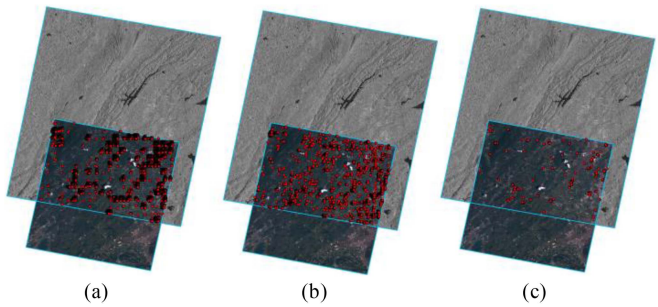


Fig. 12. TPs results by different methods for GF-3/GF-7 images in Yunnan Province. (a) CFOG. (b) RIFT. (c) MoTIF.

of the framework. In addition, SARPointNet is a deep-learning algorithm, this framework can provide a good idea for the engineering of deep learning algorithm.

In Fig. 10, SAR image MAs, such as SAR-SIFT and SAR-PointNet, can achieve better spatial distribution of TPs for SAR images. However, the other four methods have different degrees of "holes," as shown in Fig. 10 red boxes area. Compared with the results of the three multimode MAs (CFOG, RIFT, and MoTIF), the spatial distribution of RIFT is the most balanced. Fig. 11 shows the matching details of local corrected slice images.

2) GF-7/GF-3 Multimodal Image Matching Experiment in Yunnan Province: The matching TP results of GF-7/GF-3 images in Yunnan Province are shown in Table IX, and Fig. 12 shows the spatial distribution of points. SIFT, SAR-SIFT, and SARPointNet algorithms failed due to the multimodal images, there are no result statistics.

Table IX shows that the IPs RMSE of multimodal matching is poor. The results of the CFOG, RIFT, and MoTIF methods are 7.224, 14.124, and 67.784 pixels, respectively. At the same time, the CMR of CFOG, RIFT, and MoTIF methods are 78.73%, 26.06%, and 16.38%, respectively, indicating the difficulty of multimodal data matching. Although the TPs RMSE can reach the subpixel level after the removal of mismatched points, the overall spatial distribution is poor, as shown in Fig. 12.

In Fig. 12, the TP results of the RIFT method can basically cover the overlapping region, while MoTIF has the worst effect. According to Table IX, although the NCM and CMR of the CFOG method are large, their spatial distribution is weaker than that of the RIFT algorithm. In multimodal matching, in order to obtain a better spatial distribution of TPs, the threshold of mismatched points (1.5 pixels in this article) can be appropriately raised when mismatched points are eliminated. Fig. 13 shows the local details of CFOG, RIFT, and MoTIF on the experimental images.

The framework integrates SIFT, SAR-SIFT, SARPointNet, CFOG, RIFT, MoTIF, and TPs matching experiments are carried out for different images. The results show that the TPs can be extracted effectively and the generality of the framework is verified. Meanwhile, the experimental results show that the RIFT algorithm can achieve good results on both homologous

TABLE X
BUNDLE ADJUSTMENT RESULTS OF SARPOINTNET TPS BY DIFFERENT METHODS/PIXELS

Type	Number of points	Max X	Max Y	Max XY	RMSE X	RMSE Y	RMSE XY	Time
IPs	100223	509.521	119.152	509.522	19.682	0.923	19.703	29 min
BAMPE	71097	65.869	1.518	65.869	0.747	0.360	0.829	33 min
Final TPs	56012	0.838	0.833	0.854	0.267	0.254	0.368	/

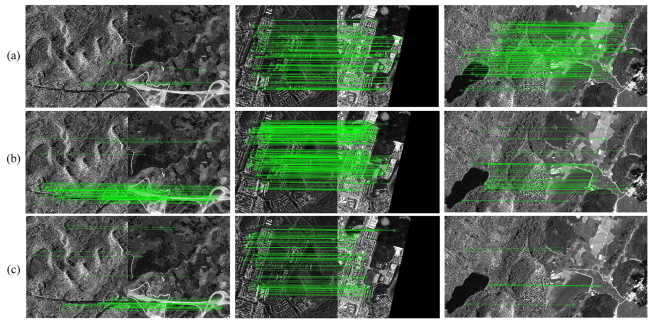


Fig. 13. Details of matching results by different algorithms in Yunnan Province. (a) CFOG. (b) RIFT. (c) MoTIF.

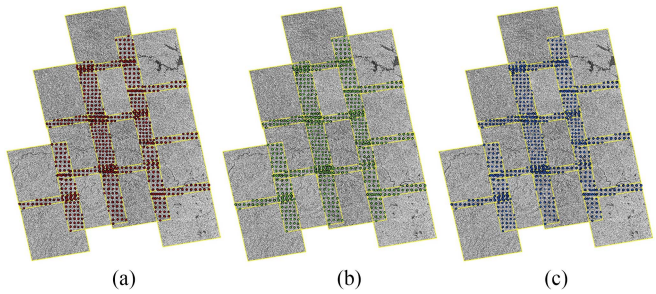


Fig. 14. Spatial distribution of TPs by different methods. (a) IPs. (b) BAMPE. (c) Final PTs.

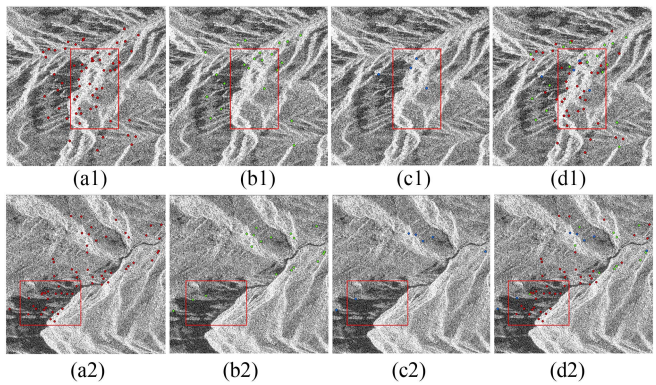


Fig. 15. Details of SARPointNet TPs by different methods. (a1), (b1), (c1) Details of IPs, BAMPE, final PTs located in the layover region. (d1) Result of (a1), (b1), (c1) stacking. (a2), (b2), (c2) Details of IPs, BAMPE, final PTs located in the shadow region. (d2) Result of (a2), (b2), (c2) stacking.

and multimodal images, which can provide guidance for the selection of practical application methods.

E. MSRSI-TPMF Experiments for Area RSI

1) Hubei Province GF-3 SAR Images Experiment: In the engineering processing of TPs matching, artificial feature MAs such as SIFT and SAR-SIFT are widely used, while deep learning methods are only being researched and there are few reports on their practical applications. In order to explore the potential application of deep learning MAs in large-scale area processing, SARPointNet method was used combined with the framework of this article proposed to conduct TPs matching experiments on GF-3 SAR images in the western Hubei Province.

Through BA method, the experiment evaluated the accuracy of initial TPs, the accuracy of refined TPs using BAMPE method, and the final TPs which removed gross errors through global BA from the refined TPs.

The statistical results showed that geometry correction took 22 min, image matching took 6 min, the TP merging took 1 min, and the points refined using the BAMPE method took 4 min. Therefore, the original matching points took 29 min, and the BAMPE method took 33 min. The experimental environment was an Intel(R) Core(TM) i7-10750H CPU @ 2.60 GHz 2.59 GHz 16.0 GB RAM, NVIDIA GeForce GTX1650. Experimental parameters, BlockX = 256 BlockY = 256, Step = 4, BAMPE Threshold = 1.5.

Table X is the BA results of SARPointNet TPs by different methods. It shows that for regional SAR images, without removing mismatched points, the RMSE of XY by BA is 19.703 pixels, which does not meet the requirements for seamless mosaic. After removing mismatched points using the BAMPE method, the RMSE of XY can be improved to 0.829 pixels, and the accuracy of the TPs is significantly improved, demonstrating the effectiveness of this method.

However, further analysis shows that the Max XY error of the TPs processed by BAMPE is 65.869 pixels. Considering that the BAMPE method is only for two images, while the BA of the regional images needs all TPs involved in the matching to meet the global optimum condition, so gross error point removal is still necessary for regional images by global BA. After the gross error was detected and eliminated by global BA, the RMSE of XY is 0.368 pixels, and MAX XY is 0.854 pixels, which meets the requirement of seamless mosaic. Fig. 14 shows the spatial distribution results of TPs obtained by different methods, and Fig. 15 shows the details of Fig. 14.

Local geometric correction strategy adopted in this framework can effectively reduce the rotation and scale difference, however, the layover and shadow caused by the SAR side-view imaging characteristics lead to mismatch points in this region, as shown in the red box in Fig. 15. In Fig. 15, it showed that the BAMPE method can effectively eliminate the mismatch points located in the region of layover and shadow. In addition, the

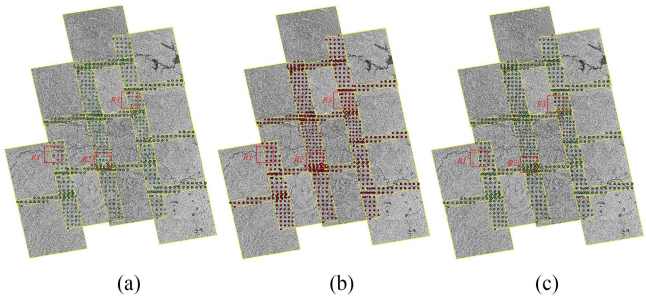


Fig. 16. Spatial distribution of TPs of SAR-SIFT and SARPointNet methods by BAMPE. (a) SAR-SIFT results. (b) SARPointNet results. (c) SAR-SIFT (green dots) and SARPointNet (red dots) stack results.

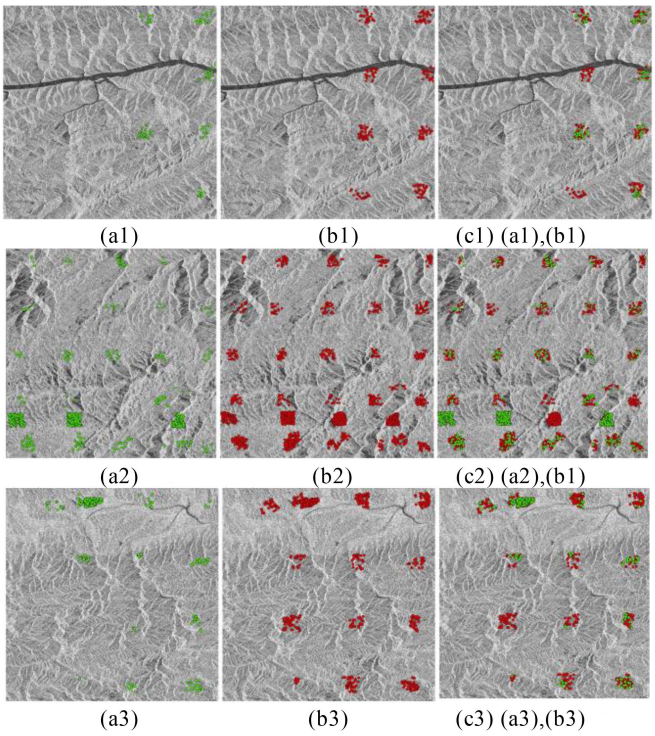


Fig. 17. Details of the connection points of SAR-SIFT and SARPointNet. (a1)SAR-SIFT in R1. (b1)SARPointNet in R1. (c1) (a1),(b1) stacking. (a2)SAR-SIFT in R2. (b2)SARPointNet in R2. (c2) (a2),(b1) stacking. (a3) SAR-SIFT in R3. (b3) SARPointNet in R3. (c3) (a3),(b3) stacking.

mismatch points can be further suppressed by means of global BA.

As a comparison, the experiments were conducted under the same experimental conditions integrating the SAR-SIFT matching method for large-scale area TPs matching. Here, image matching took 15 min, TP merging took 1.5 min, and BAMPE method optimization took 6 min. The geometry correction time remains the same (22 min), the original matching point took 38.5 min, and the BAMPE method took 44.5 min.

Fig. 16 shows the spatial distribution of TPs for SAR-SIFT and SARPointNet method, and Fig. 17 shows the details of the box area in Fig. 16 (R1, R2, R3).

As can be seen from Fig. 16, SAR-SIFT and SARPointNet can obtain better spatial distribution and a sufficient number of

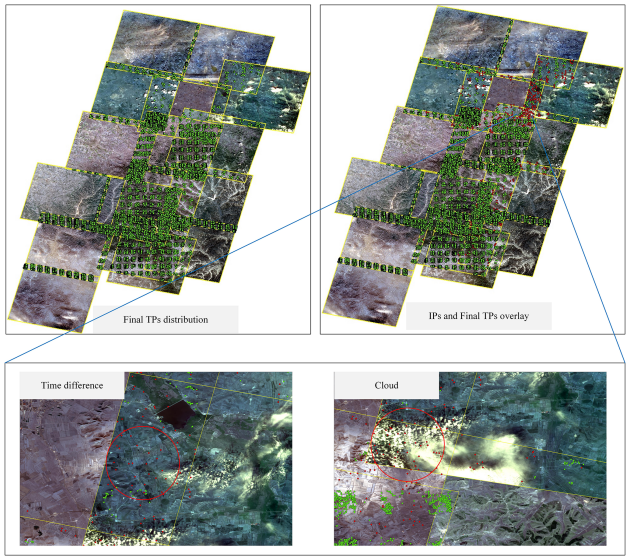


Fig. 18. Optical region images TPs results in Hebei Province.

TPs. Besides, the comparison in detail shows that the spatial distribution of SARPointNet TPs is more uniform, and there are fewer areas with "holes," which further verifies the potential of the deep learning matching method. It shows that SARPointNet/SAR-SIFT takes 33 /44.5 min in Tables X and XI, which reflects the advantages of the deep learning MA. Although the number of TPs acquired by SARPointNet is 71097, which is less than that of SAR-SIFT at 138307, Fig. 16 shows that the method can still achieve a better spatial distribution.

In summary, this framework can realize the large-scale area processing application of deep learning MA and artificial feature MA. Besides, the BAMPE method can effectively eliminate the mismatched points, which verifies the reliability of the framework in this article. The regional image comparison experiments show that the deep learning MA can achieve the same accuracy as the feature MA and can shorten the matching time. Combined with the framework of this article, it can provide guidance for the application of deep learning MAs in large-scale area image processing.

2) Hebei Province GF-1 Optical Images Experiment: Through the GF-3 SAR images experiment in Hubei Province, it was verified that the framework can perform regional processing on SAR images. In this experiment, 17 GF-1 optical images in Hebei Province were adopted, and MSRSI-TPMF proposed in this article was used to integrate the SIFT algorithm for TPs matching.

Finally, mismatched points were eliminated through BAMPE and global BA. The results are shown in Table XII. It can be seen that for the optical region images, the overall accuracy of BA is not high without removing the mismatched points, and the RMSE of TPs is 5.135 pixels, which cannot meet the requirements of a seamless mosaic. After eliminating the mismatched points, the RMSE of TPs is 0.225 pixels, which significantly improves the accuracy. The experimental results show that the CMR is 99.30% and the NCM is 130054. Points

TABLE XI
BUNDLE ADJUSTMENT RESULTS OF SAR-SIFT TPS BY DIFFERENT METHODS/PIXEL

Type	Number of points	Max X	Max Y	Max XY	RMSE X	RMSE Y	RMSE XY	Time
IPs	167471	459.185	134.405	459.185	19.047	0.591	19.056	38.5 min
BAMPE	138307	320.706	1.667	320.708	1.242	0.253	1.268	44.5 min
Final TPs	119104	1.482	0.691	1.485	0.214	0.220	0.307	/

TABLE XII
STATISTICS OF TP RESULTS IN HEBEI PROVINCE/PIXEL

Type	Number of points	CMR(%)	Max X	Max Y	Max XY	RMSE X	RMSE Y	RMSE XY
IPs	130969	—	119.997	156.219	166.468	3.173	4.037	5.135
Final TPs	130054	99.30	1.159	1.184	1.204	0.152	0.166	0.225

can be guaranteed even if the mismatched points are removed when the number of IPs is sufficient and evenly distributed.

Fig. 18 is the result of TPs in Hebei Province, where the green dot represents the final TPs, and the red dot represents the mismatched points. It can be seen from the details that for optical images, due to the influence of factors such as large time differences and clouds, there are many mismatched points. The mismatched points in this area can be effectively eliminated by our method in this article.

V. CONCLUSION

This article introduces the MSRSI-TPMF based on the geometric and radiation characteristics of RSI. The framework combines six representative MAs (SIFT, SAR-SIFT, SARPointNet, CFOG, RIFT, and MoTIF) to enable the integration of MAs with multiple algorithms and coding methods. The proposed geometry correction strategy effectively reduces the impact of scale and rotation differences on the matching point results, and the effectiveness of BAMPE method is validated through experimental analysis. A large-scale TPs matching test for deep learning MA was performed by using 13 scenes SAR images (10-m resolution) with TPs RMSE of 0.368 pixels, further validating the reliability of the framework.

In summary, the proposed matching framework can achieve automatic TPs matching for optical/SAR and other multisource RSIs by embedding artificial feature MAs and deep learning MAs. This lays the technical foundation for convenient, fast, and large-scale application of state-of-the-art MAs in remote sensing.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive comments and suggestions. The link to matching algorithms used in this article is as follows.

SIFT: https://github.com/yyxgiser/Image-Registration/tree/master/SIFT-opencv249

SAR-SIFT:https://github.com/yyxgiser/Image-Registration/tree/master/SAR-SIFT-opencv249

CFOG: https://github.com/yeyuanxin110/CFOG

RIFT: https://github.com/LJY-RS/RIFT-multimodal-image-matching

MoTIF: https://skyearth.org/publication/project/MoTIF

REFERENCES

[1] A. A. Cole-Rhodes, K. L. Johnson, J. LeMoigne, and I. Zavorin, "Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient," *IEEE Trans. Image Process.*, vol. 12, no. 12, pp. 1495–1511, Dec. 2003, doi: 10.1109/TIP.2003.819237.

[2] X. Li, N. Hui, H. Shen, Y. Fu, and L. Zhang, "A robust mosaicking procedure for high spatial resolution remote sensing images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 109, pp. 108–125, Nov. 2015, doi: 10.1016/j.isprsjprs.2015.09.009.

[3] M. P. S. Tondewad and M. M. P. Dale, "Remote sensing image registration methodology: Review and discussion," *Procedia Comput. Sci.*, vol. 171, pp. 2390–2399, 2020, doi: 10.1016/j.procs.2020.04.259.

[4] A. H. Ahmadabadian, S. Robson, J. Boehm, M. Shortis, K. Wenzel, and D. Fritsch, "A comparison of dense matching algorithms for scaled surface reconstruction using stereo camera rigs," *ISPRS J. Photogrammetry Remote Sens.*, vol. 78, pp. 157–167, Apr. 2013, doi: 10.1016/j.isprsjprs.2013.01.015.

[5] G. Schindler and F. Dellaert, "4D cities: Analyzing, visualizing, and interacting with historical urban photo collections," *J. Multimedia*, vol. 7, no. 2, pp. 124–131, Apr. 2012, doi: 10.4304/jmm.7.2.124-131.

[6] G. Kyriakaki et al., "4D reconstruction of tangible cultural heritage objects from web-retrieved images," *Int. J. Heritage Digit. Era*, vol. 3, no. 2, pp. 431–451, Jun. 2014, doi: 10.1260/2047-4970.3.2.431.

[7] J. Zhang, "Multi-source remote sensing data fusion: Status and trends," *Int. J. Image Data Fusion*, vol. 1, no. 1, pp. 5–24, Mar. 2010, doi: 10.1080/19479830903561035.

[8] H. Sui, C. Liu, Z. Gang, Z. Jiang, and C. Xu, "Overview of multi-modal remote sensing image matching methods," *Acta Geodaetica et Cartographica Sinica*, vol. 51, no. 9, pp. 1848–1861, 2022.

[9] L. Zhang, H. Ai, B. Xu, Y. Sun, and Y. Dong, "Automatic tie-point extraction based on multiple-image matching and bundle adjustment of large block of oblique aerial images," *Acta Geodaetica et Cartographica Sinica*, vol. 46, no. 5, pp. 554–564, 2017, doi: 10.11947/j.AGCS.2017.20160571.

[10] T. Wang et al., "Large-scale orthorectification of GF-3 SAR images without ground control points for China's land area," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jan. 2022, Art. no. 5221617, doi: 10.1109/TGRS.2022.3142372.

[11] B. Yang, M. Wang, W. Xu, D. Li, J. Gong, and Y. Pi, "Large-scale block adjustment without use of ground control points based on the compensation of geometric calibration for ZY-3 images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 134, pp. 1–14, Dec. 2017, doi: 10.1016/j.isprsjprs.2017.10.013.

[12] Y. Jiang et al., "A stable block adjustment method without ground control points using bound constrained optimization," *Int. J. Remote Sens.*, vol. 43, no. 12, pp. 4708–4722, Jun. 2022, doi: 10.1080/01431161.2022.2119109.

[13] X. Li et al., "Planar block adjustment for China's land regions with LuoJia1-01 nighttime light imagery," *Remote Sens.*, vol. 11, no. 18, Sep. 2019, Art. no. 2097, doi: 10.3390/rs11182097.

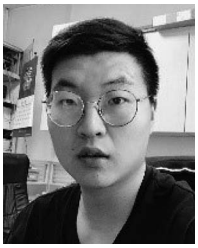
[14] X. Jiang, J. Ma, G. Xiao, Z. Shao, and X. Guo, "A review of multimodal image matching: Methods and applications," *Inf. Fusion*, vol. 73, pp. 22–71, Sep. 2021, doi: 10.1016/j.inffus.2021.02.012.

[15] X. Xu, X. Li, X. Liu, H. Shen, and Q. Shi, "Multimodal registration of remotely sensed images based on Jeffrey's divergence," *ISPRS J. Photogrammetry Remote Sens.*, vol. 122, pp. 97–115, Dec. 2016, doi: 10.1016/j.isprsjprs.2016.10.005.

[16] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, Oct. 2003, doi: 10.1016/S0262-8856(03)00137-9.
[17] Y. Jiang, "Optical/SAR image registration based on cross-correlation with multi-scale and multi-direction Gabor characteristic matrixes," in *Proc. IET Int. Radar Conf.*, 2013, pp. 0435–0435. doi: 10.1049/cp.2013.0343.
[18] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008, doi: 10.1109/TPAMI.2007.1166.
[19] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004, doi: 10.1023/B:VISI.0000029664.99615.94.
[20] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understanding*, vol. 110, no. 3, pp. 346–359, Jun. 2008, doi: 10.1016/j.cviu.2007.09.014.
[21] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 2564–2571. doi: 10.1109/ICCV.2011.6126544.
[22] 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, doi: 10.1109/TGRS.2019.2924684.
[23] Y. Zhang et al., "Histogram of the orientation of the weighted phase descriptor for multi-modal remote sensing image matching," *ISPRS J. Photogrammetry Remote Sens.*, vol. 196, pp. 1–15, Feb. 2023, doi: 10.1016/j.isprsjprs.2022.12.018.
[24] 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, Jun. 2022, doi: 10.1016/j.isprsjprs.2022.04.011.
[25] K. M. Yi, E. Trulls, V. Lepetit, and P. Fua, "LIFT: Learned invariant feature transform," in *Computer Vision – ECCV 2016*, vol. 9910, B. Leibe, J. Matas, N. Sebe, and M. Welling Eds. Cham, Switzerland: Springer International Publishing, 2016, pp. 467–483. doi: 10.1007/978-3-319-46466-4_28.
[26] H. Noh, A. Araujo, J. Sim, T. Weyand, and B. Han, "Large-scale image retrieval with attentive deep local features," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 3476–3485. doi: 10.1109/ICCV.2017.374.
[27] M. Dusmanu et al., "D2-Net: A trainable CNN for joint description and detection of local features," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 8084–8093. doi: 10.1109/CVPR.2019.00828.
[28] P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, "SuperGlue: Learning feature matching with graph neural networks," Mar. 28, 2020. Accessed: Aug. 06, 2023. [Online]. Available: http://arxiv.org/abs/1911.11763
[29] 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, Sep. 2022, Art. no. 4017605, doi: 10.1109/LGRS.2021.3105567.
[30] L. Li, M. Liu, L. Ma, and L. Han, "Cross-modal feature description for remote sensing image matching," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 112, Aug. 2022, Art. no. 102964, doi: 10.1016/j.jag.2022.102964.
[31] Y. Ye, T. Tang, B. Zhu, C. Yang, B. Li, and S. Hao, "A multiscale framework with unsupervised learning for remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Apr. 2022, Art. no. 5622215, doi: 10.1109/TGRS.2022.3167644.
[32] J. Zhao, D. Yang, Y. Li, P. Xiao, and J. Yang, "Intelligent matching method for heterogeneous remote sensing images based on style transfer," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6723–6731, Aug. 2022, doi: 10.1109/JSTARS.2022.3197748.
[33] 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, doi: 10.1109/TGRS.2014.2323552.
[34] L. Shen, J. Zhu, C. Fan, X. Huang, and T. Jin, "A novel affine covariant feature mismatch removal for feature matching," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Aug. 2022, Art. no. 5403713, doi: 10.1109/TGRS.2021.3104146.
[35] D. Quan et al., "Deep feature correlation learning for multi-modal remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Aug. 2022, Art. no. 4708216, doi: 10.1109/TGRS.2022.3187015.
[36] Z. Li, J. Yue, and L. Fang, "Adaptive regional multiple features for large-scale high-resolution remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jan. 2022, Art. no. 5617313, doi: 10.1109/TGRS.2022.3141101.
[37] X. Gong et al., "Feature matching for remote-sensing image registration via neighborhood topological and affine consistency," *Remote Sens.*, vol. 14, no. 11, May 2022, Art. no. 2606, doi: 10.3390/rs14112606.
[38] B. Zhu, C. Yang, J. Dai, J. Fan, Y. Qin, and Y. Ye, "R_2FD_2: 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, Apr. 2023, Art. no. 5606115, doi: 10.1109/TGRS.2023.3264610.
[39] Z. Zhao, H. Long, and H. You, "An optical remote sensing image matching method based on the simple and stable feature database," *Appl. Sci.*, vol. 13, no. 7, Apr. 2023, Art. no. 4632, doi: 10.3390/app13074632.
[40] J. Fan, Q. Xiong, Y. Ye, and J. Li, "Combining phase congruency and self-similarity features for multimodal remote sensing image matching," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, Jan. 2023, Art. no. 4001105, doi: 10.1109/LGRS.2023.3239191.
[41] X. Yuan and X. Liu, "A novel matching method for high resolution satellite imagery based on rational function model," *Geomatics Inf. Sci. Wuhan Univ.*, vol. 34, no. 6, pp. 671–674, 2009.
[42] J. Dai, W. Song, Y. Jia, and Q. Zhang, "A new automatically matching algorithm for multi-source high resolution optical satellite images," *Acta Geodaetica et Cartographica Sinica*, vol. 42, no. 1, pp. 80–86, 2013.
[43] K. Gutjahr, R. Perko, H. Raggam, and M. Schardt, "The epipolarity constraint in stereo-radargrammetric DEM generation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5014–5022, Aug. 2014, doi: 10.1109/TGRS.2013.2286409.
[44] Z. Zhu, Z. Shen, and J. Luo, "Parallel remote sensing image registration based on improved SIFT point feature," *J. Remote Sens.*, vol. 15, no. 5, pp. 1024–1039, 2011.
[45] H. Yang, S. Zhang, and Y. Wang, "Robust and precise registration of oblique images based on scale-invariant feature transformation algorithm," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 4, pp. 783–787, Jul. 2012, doi: 10.1109/LGRS.2011.2181485.
[46] C. V. Tao and Y. Hu, "A comprehensive study of the rational function model for photogrammetric processing," *Photogrammetric Eng.*, vol. 67, no. 12, pp. 1347–1357, 2001.
[47] D. Gong, X. Tang, and L. Zhang, "The transformation from tilted image to horizontal image of line camera based on the rational function model," *J. Geomatics Sci. Technol.*, vol. 29, no. 4, pp. 240–248, 2012, doi: 10.3969/j.issn.1673-6338.2012.04.002.
[48] G. Zhang et al., "Block adjustment for satellite imagery based on the strip constraint," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 933–941, Feb. 2015, doi: 10.1109/TGRS.2014.2330738.
[49] H. Wang et al., "Layover compensation method for regional spaceborne SAR imagery without GCPs," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 10, pp. 8367–8381, Oct. 2021, doi: 10.1109/TGRS.2020.3045505.
[50] S. Li, G. Zhang, and X. Fan, "Rigorous sensor model of gaofen-7 satellite laser altimeter based on coupled footprint camera," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1815–1826, Feb. 2023, doi: 10.1109/JSTARS.2023.3242736.
[51] X. Li et al., "SARPointNet: An automated feature learning framework for spaceborne SAR image registration," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6371–6381, Aug. 2022, doi: 10.1109/JSTARS.2022.3196383.
[52] J. Li, Q. Hu, and M. Ai, "RIFT: Multi-modal image matching based on radiation-variation insensitive feature transform," *IEEE Trans. Image Process.*, vol. 29, pp. 3296–3310, Dec. 2019, doi: 10.1109/TIP.2019.2959244.
[53] 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. XLIII-B2-2022, pp. 99–105, May 2022, doi: 10.5194/isprs-archives-XLIII-B2-2022-99-2022.

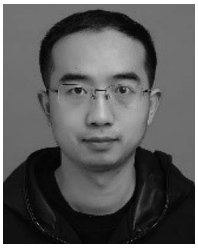
Qian Cheng received the master's degree in surveying and mapping science and technology from the School of Geomatics, Liaoning Technical University, Fuxin, China, in 2020. He is currently working toward the Ph.D. degree in photogrammetry and remote sensing with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China.

His main research interests include geometry processing of spaceborne synthetic aperture radar and 3-D reconstruction.



Xin Li received the master's degree in surveying and mapping engineering from the Wuhan University, Wuhan, China, in 2020. He is currently working toward the Ph.D. degree in photogrammetry and remote sensing with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China.

His main research interests include geometry processing of spaceborne synthetic aperture radar and deep learning in remote sensing.



Taoyang Wang received the B.E. and Ph.D. degrees in photogrammetry and remote sensing from the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China, in 2007 and 2012, respectively.

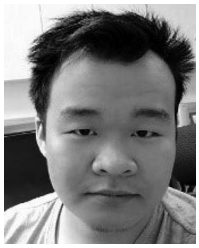
His doctoral dissertation concerned the block adjustment of high-resolution satellite remote sensing imagery. He was with the School of Remote Sensing and Information Engineering, Wuhan University, since 2014, and where he became an Associate Research Fellow, in 2015. His research interests include space photogrammetry, geometry processing of spaceborne optical/SAR/InSAR imagery, and target detection and recognition based on satellite video.

Boyang Jiang is currently working toward the Ph.D. degree in photogrammetry and remote sensing with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China.

His research interests include geometry processing of spaceborne imagery and block adjustment.

He Fu received the M.E. degree in surveying and mapping engineering from the Jilin University, Changchun, China, in 2014.

He is currently working with ZhongSe Lantu Technology Company, Ltd., Beijing, China. His main research interests include geometry processing of spaceborne optical/SAR imagery.



Yunming Wang received the B.E. degree in surveying and mapping from the Henan Polytechnic University, Jiaozuo, China, in 2017, and the M.E. degree in surveying and mapping, in 2020, from the Wuhan University, Wuhan, China, where he is currently working toward the Ph.D. degree in photogrammetry and remote sensing with the School of Remote Sensing and Information Engineering.

His research interest mainly focuses on 3-D reconstruction.

Feida Zhang is currently working toward the M.E. degree in surveying and mapping engineering with the School of Earth Science and Resources, China University of Geosciences, Beijing, China.

His research mainly focuses on multimodal image matching.