

# Adaptive Regional Multiple Features for Large-Scale High Resolution Remote Sensing Image Registration

Zezhou Li, Jun Yue, Leyuan Fang, *Senior Member, IEEE*

**Abstract**—The efficient and accurate registration of multi-temporal images is essential for many remote sensing applications. With the increasing of the imaging resolution and field in satellites, the acquired large-scale remote sensing images have brought the serious challenges, since there exist parallax shifts and large background variations among different local regions of the acquired images. To address these issues, this paper proposes an adaptive regional multiple features (ARMF) matching method for the registration of multi-temporal large-scale high resolution remote sensing images. Specifically, since large background variations in fixed-size regions of multiple-temporal images will cause the insufficient features and the failure of features matching, the ARMF introduces an adaptive regions searching strategy, which utilizes the pyramid-amplification technique to adaptively select the regions that can find the sufficient matched features. Then, the ARMF extracts multiple types of features (i.e., gradient feature, phase feature, and line feature) from the adaptive searched region that can more effectively represent the characteristics of the large regions. Finally, we utilize the feature matching error as the rule to adaptively select the suitable features as the descriptors of the region. The experimental results on large-scale multi-temporal image data obtained from Google Earth demonstrated the proposed method can outperform several state-of-the-arts remote sensing registration approaches.

**Index Terms**—Image registration, multi-features, large-scale remote sensing images, pyramid-shape region adaptive amplification.

## I. INTRODUCTION

MULTI-temporal remote sensing image registration is a process of establishing correspondences and geometric transformation model to align two or more remote sensing images of the same scene at different times or from different viewpoints [1]. Recently, registration has become a key technology for prerequisite of many remote sensing applications,

This work was supported in part by the National Natural Science Foundation of China under Grant 61922029 and Grant 62101072, in part by Open Fund of State Key Laboratory of Remote Sensing Science Grant OFSLRSS202104, in part by the Science and Technology Plan Project Fund of Hunan Province under Grant 2019RS2016, and in part by the Key Research and Development Program of Hunan under Grant 2021SK2039. (Zezhou Li and Jun Yue contributed equally to this work.) (Corresponding author: Leyuan Fang.)

Z. Li is with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China, and also with the State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China. (e-mail:lzz23@hnu.edu.cn.)

L. Fang is with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China, and also with the Peng Cheng Laboratory, Shenzhen 518000, China. (fangleyuan@gmail.com.)

J. Yue is with the Department of Geomatics Engineering, Changsha University of Science and Technology, Changsha 410114, China. (e-mail:jyue@pku.edu.cn.)

such as change detection [2], [3], geographic surveying, and image fusion [4]–[6]. The geographic information can be directly used to achieve remote sensing image pre-registration, which usually has an offset of dozen of pixels [7]. Therefore, how to automatically achieve high-precision remote sensing image registration has become a hot topic in the remote sensing field.

In recent years, many automatic methods have been introduced for the image registration of multi-temporal remote sensing images [8]–[14]. These methods can be roughly divided into two categories: feature-based and area-based [15]. Feature-based methods first extract salient features from two considered images, and then match them according to their similarity to achieve the registration. These methods focus more on how to improve the universality and robustness of image features. Commonly used image features include point features [16]–[22], edge features [23]–[26], and region features [27], [28]. Scale invariant feature transform (SIFT) [16] is a typical local features extraction method, that has been widely used in the remote sensing image registration. Although the SIFT can extract robust features, there are some problems in the quantity and distribution of extracted features in remote sensing images. Therefore, a feature selection strategy called the uniform robust SIFT (UR-SIFT) [17] based on stability and saliency constraint isolation feature quality is introduced. Then, to discriminate distributed descriptors, the adaptive binning SIFT (AB-SIFT) [18] is introduced to improve the local distortion performance of the descriptors. A two-step alignment method [29] using fused images for registration makes image pixel clustering and anomaly detection performance better. With the development of deep learning methodology, a multi-temporal remote sensing image registration based on multi-layer feature fusion of deep residual network [30] is proposed, which fused different levels of features in the deep residual network. After that, the method in [31] based on the combination of deep features and traditional local SIFT feature is introduced to address the problem that features are negatively affected by noise and differences between remote sensing images.

Area-based methods usually adopt a template window of a predefined size to detect the correspondence between different temporal images, and then utilize similarity-based metrics to match the corresponding patches in the images [32]. Similarity measure is the key component for the area-based image registration. The normalized cross correlation (NCC) is a very popular similarity measure, because of its invariance to linear

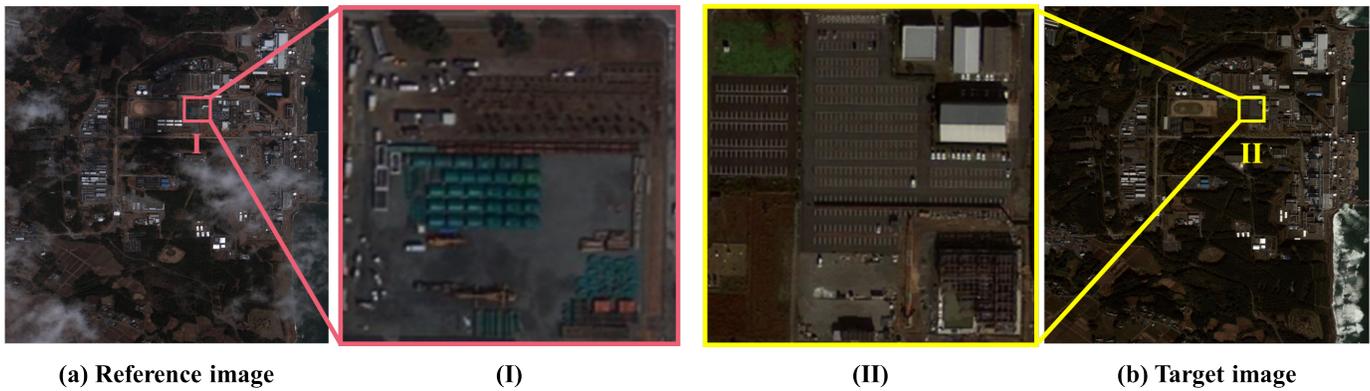


Fig. 1. The obvious background and objects changes in local regions of multi-temporal large-scale high resolution remote sensing images. (I) and (II) are the local regions of reference image (a) and target image (b).

intensity changes. However, NCC is susceptible to nonlinear radiation differences [33]. The mutual information (MI) is more robust to complex radiation measurement changes, but it is computationally expensive since it needs to calculate the joint histogram of each window to be matched [34]. In addition, the MI ignores the spatial information of neighboring pixels, thereby reducing the quality of image matching [35]. For multi-temporal remote sensing images, the scene characteristics of the image have been dramatically changed or disappeared. To alleviate these problems, a method called the phase correlation [36] is introduced to consider the robustness for the scaling and rotation changes and achieve excellent stability. The template window of the area-based registration method is related to the control points and the search space is geometrically limited. This method is easy to fail, especially when there are obvious distortions or geometric differences in multi-temporal remote sensing images. For the feature-based methods, the extracted feature points are matched through feature descriptors and spatial relationships, which can be applied when the image has obvious distortions and geometric differences. Therefore, compared with the area-based method, the feature-based method is more effective in multi-temporal high-resolution remote sensing images.

Recently, with the rapid development of remote sensing imaging technology, the resolution and size of acquired multiple-temporal images have been greatly increased, which brings the serious challenging for image registration [10]. For example, the one scene data area of Gaofen-2 satellite is more than 500 square kilometers, while its spatial resolution is better than 1m. The spatial resolution of world-view-3 even reaches 0.3m, and their image size basically exceeds  $10000 \times 10000$ . For the very large scale remote sensing images, the above typical feature-based and area-based methods hardly obtain excellent registration results. In addition, it is not feasible to use traditional algorithms to directly perform image registration on the entire large-scale image, due to the limitation of both computational memory and time consuming. An intuitive way is to divide the large-scale images into many regions of fixed sizes. However, the multi-temporal remote sensing images are all taken by satellites in various time periods, and thus the obvious local features in the acquired images

have changed, such as houses, mountain forest, and lakes (see Fig.1). The large scene changes in the fixed sub-blocks of different temporal images will lead to insufficient image features and feature mismatch. In addition, the difference of imaging conditions in different temporal images also will seriously affect the performance of registration for the local regions [37]. Therefore, it is difficult to improve the accuracy of image registration only by the fixed sub-blocks for the large-scale high resolution remote sensing images.

To address the above issues, in this paper, we propose a novel method called adaptive regional multiple features (ARMF) for large-scale high resolution multi-temporal remote sensing images registration. The method mainly consists of the following steps. First, we use multiple features for feature detection of local image regions, and use pyramid magnification technology to adaptively find suitable regions according to the number of matching features. Then, the proposed method can adaptively eliminate outlier features from the adaptive searched regions by feature matching errors. After that, we downsample the amplified local region image to the original image size and repeat the first step, and then remap the matched feature coordinates back to the coordinate system before sampling. Finally, we obtain the sufficient number of matching features to model calculations for large-scale remote sensing images to achieve high precision image registration.

The rest of this paper is organized as follows. Related works are reviewed in the Section II. Then, the proposed ARMF framework is presented in Section III. The registration performances of the proposed method are evaluated in Section IV. Finally, we conclude this paper and suggest future works in Section V.

## II. RELATED WORKS

Image registration is widely used in various fields, such as computer vision, pattern recognition, etc., especially in the fields of photogrammetry and remote sensing. For more comprehensive review of image registration methods, please refer to the work in [38]. In this section, we will mainly review some registration works related to our proposed method.

### A. Local Features

- 1) Local gradient feature of SIFT [16]: local feature extraction and description is a key step in image registration, and the stable and robust features determine the quality of the registration result. The SIFT describes the image area in the gradient domain and constructs a 128-D histogram, which is robust to scale and direction changes. The main algorithm steps of SIFT include the construction of scale space, feature selection and direction assignment, feature descriptor and matching. First, the algorithm uses the Gaussian function to calculate and search for features, which are invariant to scale and direction in the image at all scales.  $L(x, y, \sigma)$  is defined as the convolution operation of the original image  $I(x, y)$  and a variable scale Gaussian function  $G(x, y, \sigma)$ ,

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (1)$$

where  $*$  is the convolution operation in  $x$  and  $y$ , and the  $G(x, y, \sigma)$  is a variable-scale Gaussian defined as,

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (2)$$

The extremum detection in scale space  $D(x, y, \sigma)$  is defined as the convolution of a difference of Gaussian with image  $I(x, y)$ ,  $k$  is a constant multiplication factor,

$$D(x, y, \sigma) = L(x, v, k\sigma) - L(x, y, \sigma) \quad (3)$$

After constructing the Difference of Gaussians (DoG) pyramid to find the extreme points of the DoG function, each pixel must be compared with all its neighbors. If it is larger or smaller than the neighboring points of its image domain and scale domain, the most suitable extreme point of the edge response is removed. For the key points detected in the DOG pyramid, the gradient and direction distribution characteristics of the pixels in the  $3\sigma$  neighborhood window of the Gaussian pyramid image are collected. SIFT calculates the gradient  $B(x, y)$  and the angular  $\theta(x, y)$  magnitude as,

$$B(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (4)$$

$$\theta(x, y) = \arctan\left(\frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}\right) \quad (5)$$

After completing the gradient calculation of the key points, SIFT uses the histogram to calculate the gradient and direction of the pixels in the neighborhood. The gradient histogram divides the direction range from 0 to 360 degrees into 36 bins. The peak direction of the histogram represents the main direction of the key point, and the maximum value in the histogram is used as the main direction of the key point.

- 2) Line segment feature: The typical line segmentation feature method [39] uses the (Edge Drawing Lines) EDLines [40] detector to extract line segments from the corresponding two images, and performs line verification steps to remove meaningless line segments. First,

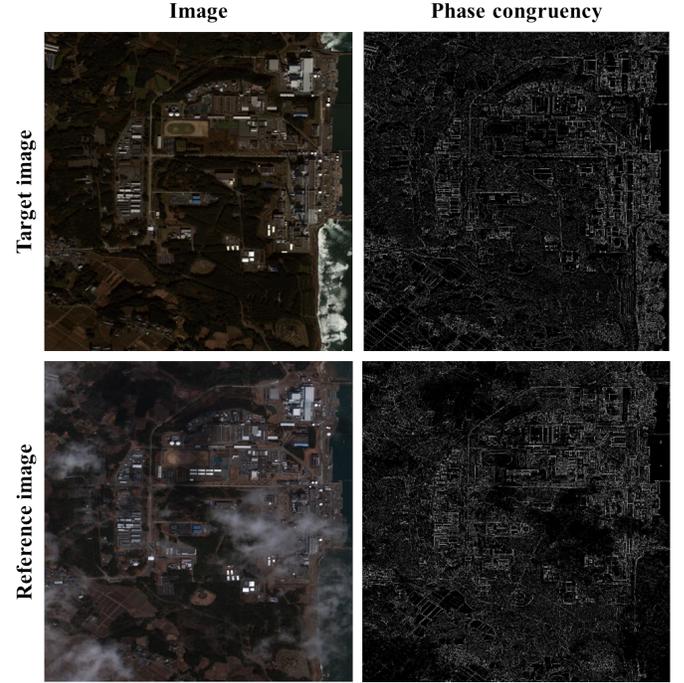


Fig. 2. Reference image and target image of phase congruency.

we compute the gradient magnitude and the level line angle at pixel  $(x, y)$ .

$$g_x(x, y) = \frac{I(x+1, y) - I(x, y) + I(x+1, y+1) - I(x, y+1)}{2} \quad (6)$$

$$g_y(x, y) = \frac{I(x, y+1) - I(x, y) + I(x+1, y+1) - I(x+1, y)}{2} \quad (7)$$

$$g(x, y) = \sqrt{g_x(x, y)^2 + g_y(x, y)^2} \quad (8)$$

$$\text{angle}(x, y) = \arctan\left(\frac{g_x(x, y)}{-g_y(x, y)}\right) \quad (9)$$

where  $I(x, y)$  is the intensity of the input image at pixel  $(x, y)$ ,  $g(x, y)$  is the gradient magnitude, and  $\text{angle}(x, y)$  is the horizontal line angle. Then, based on the position and direction of each line segment, a histogram descriptor that is robust to global geometric distortion is generated.

- 3) Phase feature: Phase feature has excellent feature detection performance compared with image gradient features and structural features. A maximum index image (MIM) is proposed for feature description in Radiation-Variation Insensitive Feature Transform (RIFT) [41]. MIM is constructed by log-Gabor convolution sequence, which is more robust to nonlinear radiation differences (NRD) than traditional gradient maps. The method selects the log-Gabor wavelet transform with direction  $o = 6$  and scale  $n = 4$ , and calculates the phase consistency measure  $PC(\theta_{pc})$ , where  $(x, y)$  represents the position coordinate, and the calculation function expression is:

$$PC(x, y) = \frac{\sum_n \sum_o W_o(x, y) |A_{no}(x, y) \Delta \phi_{no}(x, y) - T|}{\sum_n \sum_o A_{no}(x, y) + \varepsilon} \quad (10)$$

where  $\theta_{pc} = (x, y)$  is the position coordinate in the reference and target image block region,  $W_o(x, y)$  is

the weight of the position coordinates,  $A_{no}(x, y)$  is the direction  $o$  and the position coordinates  $(x, y)$ . The amplitude corresponds to scale  $n$ ,  $\Delta\phi_{no}(x, y)$  is the phase deviation of position coordinates  $(x, y)$ . The  $T$  represents the noise threshold, and the  $\lfloor \cdot \rfloor$  represents if the value is positive, the value is equal to itself, otherwise the value is 0. The  $\mathcal{E}$  is a constant. Fig. 2 shows the reference image and target image of phase congruency. The use of the maximum moment  $M$  and the minimum moment  $m$  is to extract the edge and corner features in the phase information. The maximum moment  $M$  and the minimum moment  $m$  function can be expressed as,

$$\begin{cases} M = \frac{1}{2} \left( f + d + \sqrt{e^2 + (d - f)^2} \right) \\ m = \frac{1}{2} \left( f + d - \sqrt{e^2 + (d - f)^2} \right) \end{cases} \quad (11)$$

$$e = 2 \sum (PC(\theta_{pc}) \cos \theta_{pc}) (PC(\theta_{pc}) \sin \theta_{pc}) \quad (12)$$

$$d = \sum (PC(\theta_{pc}) \cos \theta_{pc})^2 \quad (13)$$

$$f = \sum (PC(\theta_{pc}) \sin \theta_{pc})^2 \quad (14)$$

To get the correct matching of various types of features, eliminating outlier is an important step for improving the accuracy of image matching in the registration. The Random Sample Consensus (RANSAC) [42] is a classic method to remove outliers through geometric verification.

### B. Transformation Model

The critical last step of image registration is the calculation of the relationship with the the model transformation. A new parameterization of deformation fields for image registration method [43] describes a deformation field with its transformation Jacobian and curl of endvelocity field with parameterization. The method called (As-Projective-As-Possible, APAP) [44] accurately align images that differ by more than a pure rotation, which can produce much better results on image stitching. The algorithm assumes that the matching points of the reference image and the target image overlap area are  $z = [x, y]^T$  and  $t = [x', y']^T$ , after homography transformation, the relationship is as follows:

$$\tilde{t} = H\tilde{z} \quad (15)$$

The homogeneous form of the transformation projection formula is expressed as,

$$x' = \frac{x'_1}{x'_3} = \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}} \quad (16)$$

$$y' = \frac{x'_2}{x'_3} = \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}} \quad (17)$$

assuming  $a_i \in R^{2 \times 9}$ , the moving Direct Linear Transformation (DLT) algorithm calculates the 9 elements of the homography matrix  $H$ , and the objective function is calculated as,

$$\tilde{h} = \arg \min_h \sum_{i=1}^N \|a_i h\|^2 = \arg \min_h \|Ah\|^2 \quad (18)$$

where  $\|h\| = 1$  and  $A \in R^{2 \times 9}$ , and then considering the weight  $\{w_i^i\}_{i=1}^N$  to set a homography matrix for each independent position,  $\sigma$  is the scale factor. The weights are calculated as,

$$h_* = \arg \min \sum_{i=1}^N \|w_* a_i h\|^2 \quad (19)$$

$$w_*^i = \exp \left( - \|x_* - x_i\|^2 / \sigma^2 \right) \quad (20)$$

## III. PROPOSED METHOD

For processing multi-temporal remote sensing images with the fixed-blocks, large scene changes make it difficult to extract a sufficient number of image features in the local region, and the large size of the image brings high computational cost. To make the registration algorithm robust in large-scale multi-temporal remote sensing images, the ARMF framework is proposed. In this section, we will detail the proposed ARMF method, including the adaptive multiple features extraction and combination, image region adaptive amplification, image region downsampling and feature remapping. (see Fig.3).

### A. Adaptive Multiple Features Extraction and Combination

One type of feature extraction method may not capture the sufficient number of features and result in the feature matching failure in high resolution remote sensing images. Therefore, given a pair of large-scale multi-temporal reference image  $I_{ref}(x, y)$  and a target image  $I_{tar}(x, y)$ , we propose to use multiple types of features, which can effectively alleviate the problem of changes in local region and feature matching failure. First, we use SIFT [16], RIFT [41] and EDlines [40] to extract multiple features from the local regions of reference and target remote sensing images. The (Line Segment Detection) LSD features are detected on the line structural features of RIFT features. On this basis, the euclidean distance is used for similarity measurement, and the RANSAC [42] algorithm is used to eliminate the wrong matching relationship.

The matched SIFT, RIFT and LSD features detected by the local reference and target image region blocks after the RANSAC algorithm are  $\{P_{ref}^1, \dots, P_{ref}^A\}$ ,  $\{P_{tar}^1, \dots, P_{tar}^A\}$ ,  $\{pf_{ref}^1, \dots, pf_{ref}^C\}$ ,  $\{pf_{tar}^1, \dots, pf_{tar}^C\}$ ,  $\{L_{ref}^1, L_{ref}^2, \dots, L_{ref}^B\}$ ,  $\{L_{tar}^1, L_{tar}^2, \dots, L_{tar}^B\}$ . We calculate the LSD features in the cartesian coordinate system as,

$$\begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} = H_L \begin{pmatrix} u_i \\ v_i \\ 1 \end{pmatrix} \quad (21)$$

where  $x_i$  and  $y_i$ ,  $u_i$  and  $v_i$  represent the horizontal and vertical coordinates of the feature points on the line structure features detected in the local region of the reference image and the target image.  $H_L$  is the transformation model matrix of LSD feature matching,

$$RMSE = \sqrt{\frac{1}{N} \left( \sum_{i=1}^N (x_i - u_i)^2 + (y_i - v_i)^2 \right)} \quad (22)$$

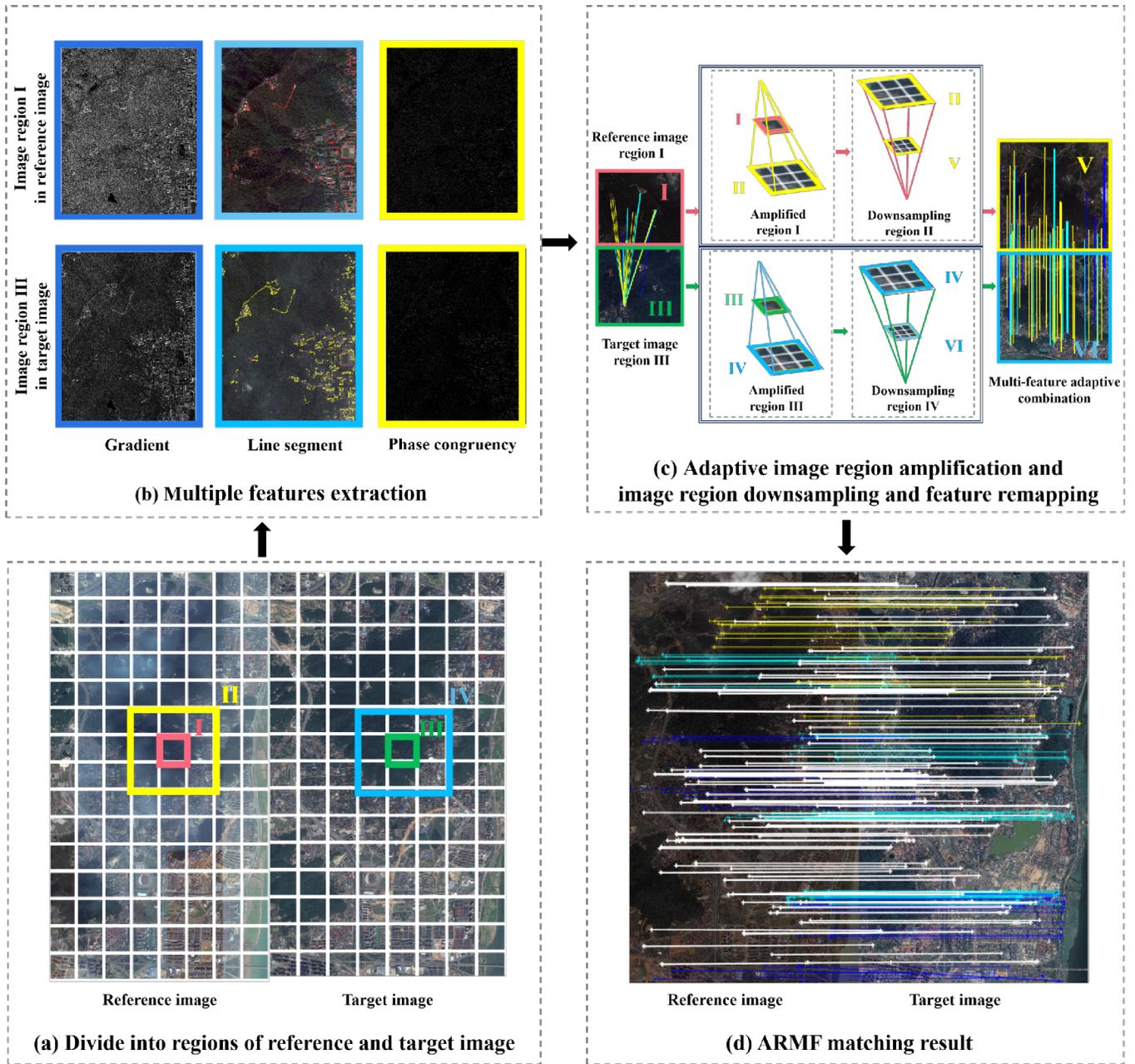


Fig. 3. (1) The dashed block diagram (a) indicates the reference and the target image are divided into the same size image regions; (2) The dashed block diagram (b) indicates that we extract the gradient feature, phase feature, and line structure feature; (3) The dashed block diagram (c) indicates that II and IV represent the image area I and II after the adaptive image region amplification, V and VI represent the image region down-sampling of II and IV, which light blue, yellow and blue are respectively represents three local features of LSD, the phase feature of RIFT, and the gradient feature of SIFT; (4) The dashed block diagram (d) indicates the ARMF matching result in which light white is represents feature after ARMF algorithm.

$$\begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} = H \begin{pmatrix} u_i \\ v_i \\ 1 \end{pmatrix} \quad (23)$$

$$F_{ref} = \left\{ \left\{ P_{ref}^1, \dots, P_{ref}^A \right\}, \left\{ L_{ref}^1, \dots, L_{ref}^B \right\}, \left\{ pf_{ref}^1, \dots, pf_{ref}^C \right\} \right\} \quad (24)$$

$$F_{tar} = \left\{ \left\{ P_{tar}^1, \dots, P_{tar}^A \right\}, \left\{ L_{tar}^1, \dots, L_{tar}^B \right\}, \left\{ pf_{tar}^1, \dots, pf_{tar}^C \right\} \right\} \quad (25)$$

The  $F_{ref}$  and  $F_{tar}$  obtained after SIFT, LSD and RIFT of feature matching in reference and target image blocks. We calculate their feature matching error (FME) by the

formulas of (22) and (23), and we can obtain the matching relationship between each type of the features. The local transformation error  $c_i = \{F_{ref}, F_{tar}\}$  is denoted as  $e(c_i) = \|(u_i, v_i) - H_i(x_i, y_i)\|$  and  $c_i$  represents the multi-feature set of the image region blocks after RANSAC. We remove the features with  $e(c_i) < T_1, T_1 = 3$  (pixels), and combine the features to the final multi-feature sets  $F'_{ref}$  and  $F'_{tar}$ . (see Fig.4),

$$F'_{ref} = \left\{ \left\{ P_{ref}^1, \dots, P_{ref}^a \right\}, \left\{ L_{ref}^1, \dots, L_{ref}^b \right\}, \left\{ pf_{ref}^1, \dots, pf_{ref}^c \right\} \right\} \quad (26)$$

$$F'_{tar} = \left\{ \left\{ P_{tar}^1, \dots, P_{tar}^a \right\}, \left\{ L_{tar}^1, \dots, L_{tar}^b \right\}, \left\{ pf_{tar}^1, \dots, pf_{tar}^c \right\} \right\} \quad (27)$$

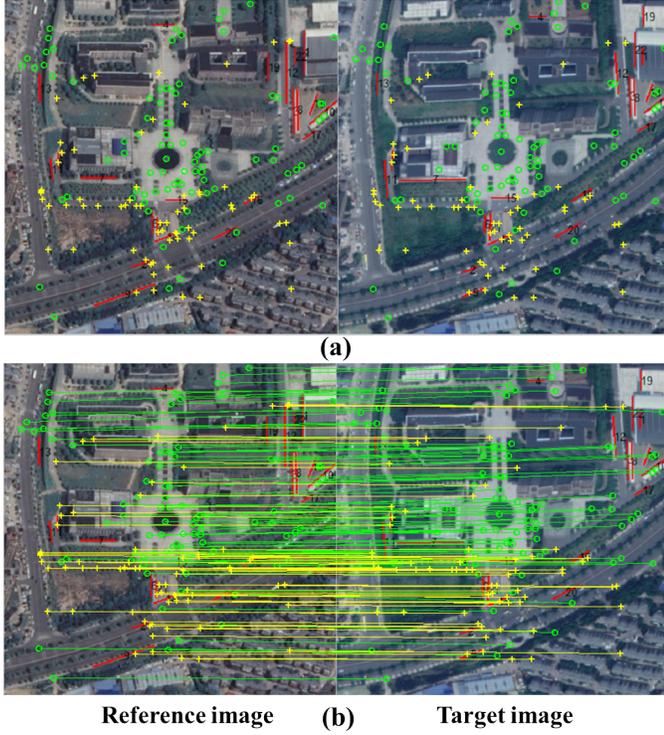


Fig. 4. (a) Multi-features detection in reference image and target image. (b) Matching multi-features in reference image and target image. The red line in the figure represents the extracted line structure feature, the green open circle represents the SIFT feature, and the yellow cross represents the RIFT feature.

### B. Adaptive Image Region Amplification

When the feature matching error (FME) of the multi-features that calculated by formulas (22) and (23) is greater than the threshold  $T_2$  or less than the number of features matching  $T_p = 3$  pairs, we set the complete matching to be failure when the feature error  $c'_i = \{F'_{ref}, F'_{tar}\}$  after the transformation satisfy the formula  $e(c'_i) > T_2, T_2 = 10$ . Then, the current image region block is used as the center to enlarge the image region (see Fig. 5). In this paper, we set the image size of the amplified region to be  $3 \times 3$  side length of the original  $1 \times 1$  image region.

### C. Image Region Downsampling and Feature Remapping

In order to reduce the computational cost of image feature detection and matching in the enlarged image area, we down-sample the enlarged image area. Note that, a sufficient number of feature matching pairs obtained in the original region amplification process. The downsampling process provides a great help to the subsequent calculation of the image transformation model.

When the image transformation model is calculated, we need to map the current down-sampled image feature coordinates back to the original image coordinate system. According to formula (28), the multiple features coordinates of the local area are remapped, and then the final multi-feature matching coordinate pairs are saved (see Fig.6),

$$(x_i, y_i) = (x'_i, y'_i) \times 2^\mu \quad (28)$$

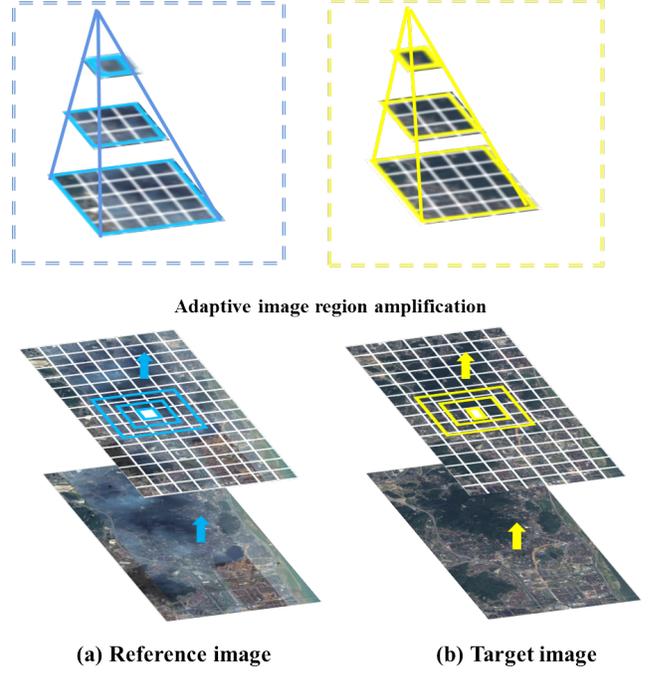


Fig. 5. Main process for adaptive image region amplification. (a) Reference image and (b) Target image.

where  $x_i$  and  $y_i$  represent the row and column coordinates of the feature points in the multi-feature sets local region of the reference image,  $x'_i$  and  $y'_i$  represent the row and column coordinates of  $x_i$  and  $y_i$  after downsampling, and  $\mu$  represents the number of times that the amplified region is downsampling. The features extracted after downsampling are re-matched and then mapped back to the original image coordinate system. The new feature sets that we obtain during the downsampling process are  $F_{ref}^{new}$  and  $F_{tar}^{new}$ , and then we merge all the features obtained during the downsampling process to obtain the final feature sets  $F_{ref}^{total}$  and  $F_{tar}^{total}$ .

$$F_{ref}^{total} = \{F'_{ref}, F_{ref}^{new}\} \quad (29)$$

$$F_{tar}^{total} = \{F'_{tar}, F_{tar}^{new}\} \quad (30)$$

Finally, we calculate the image global transformation model  $H_g$  of the multi-features sets  $F_{ref}^{total}$  and  $F_{tar}^{total}$  matching obtained from all the reference and target image region blocks. The global transformation model  $H_g$  is expressed as:

$$H_g = \sum_{i=1}^N W_i H_i^i \quad (31)$$

where  $H_g$  represents the global image transformation model,  $N$  is the total number of local image area transformation models,  $H_i^i$  is the local image transformation model, and the function of  $W_i$  is expressed as APAP [44],

$$H_g = \arg \min_h \|W_i A h\| \quad \text{s.t. } \|h\| = 1 \quad (32)$$

where  $A \in R^{2 \times 9}$ , The method divides the source image into a grid, takes the center point of the grid, and calculates the Euclidean distance and weight between each center point and

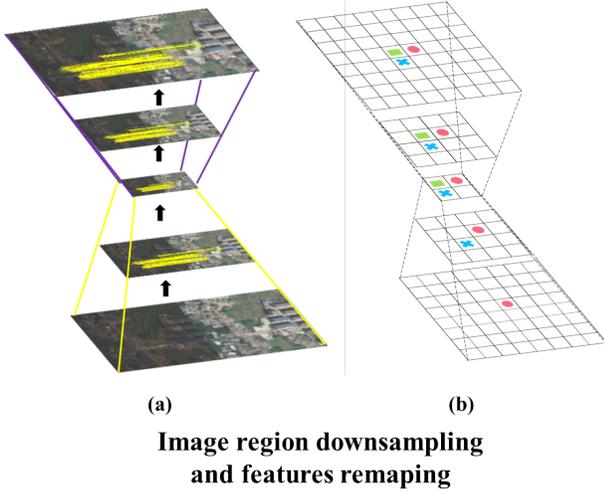


Fig. 6. (a) The processing diagram of feature coordinate remapping. The yellow line represents the matched feature point pairs. (b) Feature coordinate mapping details diagram. The red solid dot represents a feature in the original image region. The light blue solid cross and light green solid rectangle represents the new matched feature after downsampling the original image region.

the inner point on the source image. Finally, each grid is traversed, and the local homography matrix is used to map to the full image. The closer the weight of the pixel value is to the neighborhood, the calculated value of the Gaussian function is used for the weight value, and the smaller the weight of the pixel value will be.

#### IV. PROPOSED METHOD EVALUATION AND EXPERIMENTAL RESULTS

##### A. Test Data Sets

The performance of the proposed ARMF method is evaluated with a series of large-scale multi-temporal remote sensing images from Google Earth (see Fig.7). The multi-temporal remote sensing images are separately acquired from the he University Science and Technology City of Changsha (see Fig.7 (a)), Huanghua Airport area in Changsha (see Fig.7 (b)), and urban area of Changsha (see Fig.7 (c)). The purpose of Fig.7 (a) is to test the effectiveness of image registration in the multi-temporal change images of mountains, urban areas, lakes, etc. Fig.7 (b) contains a multi-temporal change area of farmland and mountain areas. Fig.7 (c) includes multi-temporal changes of high-rise buildings and urban roads. The details of these images are reported in Table.I.

##### B. Evaluation Criteria and Implementation Deta

This paper uses root mean square error (RMSE) [45], correct correspondence number (NOCC), mutual information (MI) [34], rate of correct correspondences (ROCC) to evaluate the registration methods. The metrics to measure the accuracy of the algorithms are RMSE and MI. NOCC and ROCC are used to measure the robustness of the algorithms. For all image pairs, we take the average of five to ten times image regions to

#### Algorithm 1 Proposed Algorithm.

- Input:**  $I_{ref}$  and  $I_{tar}$  : The reference image and target image;  
 $L_{ref}$  and  $L_{tar}$  : The LSD feature in  $I_{ref}$  and  $I_{tar}$  ;  
 $P_{ref}$  and  $P_{tar}$  : The SIFT feature in  $I_{ref}$  and  $I_{tar}$  ;  
 $pf_{ref}$  and  $pf_{tar}$  : The RIFT feature in  $I_{ref}$  and  $I_{tar}$  .
- Output:** The transformation model  $T$  and the correct multi-features matching pairs  $F_{ref}^{total}$  and  $F_{tar}^{total}$  .
- 1: Divide the  $I_{ref}$  and  $I_{tar}$  image into region blocks  $B_{ref-i}$  and  $B_{tar-i}$ .
  - 2: Calculate the matched LSD feature, the SIFT feature and the RIFT feature detected in  $B_{ref-i}$  and  $B_{tar-i}$  after RANSAC as  $F_{ref}$  and  $F_{tar}$ ;
  - 3: Calculate the local transformation model  $T_l^i$  by  $\{L_{ref}^1, \dots, L_{ref}^a\}$  and  $\{L_{tar}^1, \dots, L_{tar}^a\}$ , the local transformation model  $T_p^i$  by  $\{P_{ref}^1, \dots, P_{ref}^b\}$  and  $\{P_{tar}^1, \dots, P_{tar}^b\}$ , and the local transformation model  $T_{pf}^i$  by  $\{pf_{ref}^1, \dots, pf_{ref}^c\}$  and  $T_{pf}^i$  by  $\{pf_{tar}^1, \dots, pf_{tar}^c\}$  from  $B_{ref-i}$  and  $B_{tar-i}$ .
  - 4: Calculated the correct multi-features matching error and eliminated mismatched feature pairs. Add  $\{L_{ref}^1, \dots, L_{ref}^d\}$   $\{L_{tar}^1, \dots, L_{tar}^d\}$ ,  $\{P_{ref}^1, \dots, P_{ref}^e\}$   $\{P_{tar}^1, \dots, P_{tar}^e\}$ ,  $\{pf_{ref}^1, \dots, pf_{ref}^f\}$   $\{pf_{tar}^1, \dots, pf_{tar}^f\}$  to  $F_{ref}'$  and  $F_{tar}'$ .
  - 5: Calculate the local transformation model  $T_i'$ , and eliminated the local transformation error value of  $\{F_{ref,i}', F_{tar,i}'\}$  over than the transformation error of the matched feature pairs less than  $T_p = 3$  .
  - 6: Pyramid-shape image region adaptive amplification, and add the new feature  $F_{ref}^{new}$  and  $F_{tar}^{new}$  to the  $F_{ref}^{total}$  and  $F_{tar}^{total}$  .
  - 7: Repeat the step 1, until the end of the feature search for all image region blocks.
  - 8: Obtain final features  $F_{ref}^{total}$  and  $F_{tar}^{total}$  of ARMF, and then, calculate the transformation model  $T$ .

obtain the final RMSE results. We use the RMSE to evaluate the registration accuracy.

$$RMSE = \sqrt{\frac{1}{N} \left( \sum_{i=1}^N (x_i - u_i)^2 + (y_i - v_i)^2 \right)} \quad (33)$$

where  $N$  is the total feature point pairs,  $x_i$  and  $y_i$  represent the row and column coordinates of the feature points in the reference image,  $u_i$  and  $v_i$  represent the row coordinates and column coordinates of the corresponding feature points in the target and the reference image after transformation. The robustness of the algorithm is evaluated by the correct number of matches. The more the NOCC, the higher the accuracy of the result. MI is used as a measure of image pair similarity by counting the correlation between two data sets. In addition, this method is suitable for remote sensing image registration. We evaluate the accuracy of registration for large scale remote

TABLE I  
DETAILS OF REAL MULTI-TEMPORAL REMOTE SENSING IMAGE PAIRS.

Image No	Size	Date	Spatial resolution(m)	Location	
1	12000 × 12000	09/2012	1	University Science and Technology City of Changsha	
	12000 × 12000	06/2015	1		
2	12000 × 12000	09/2012	1		
	12000 × 12000	02/2017	1		
3	12000 × 12000	09/2012	1		
	12000 × 12000	08/2019	1		
4	13056 × 11008	12/2013	1.5		Changsha Huanghua Airport Area
	13056 × 11008	12/2015	1.5		
5	13056 × 11008	12/2013	1.5		
	13056 × 11008	12/2017	1.5		
6	13056 × 11008	12/2013	1.5		
	13056 × 11008	12/2019	1.5		
7	11520 × 10752	12/2008	1.5	Urban Area of Changsha	
	11520 × 10752	12/2017	1.5		
8	11520 × 10752	12/2008	1.5		
	11520 × 10752	12/2018	1.5		
9	11520 × 10752	12/2008	1.5		
	11520 × 10752	12/2019	1.5		

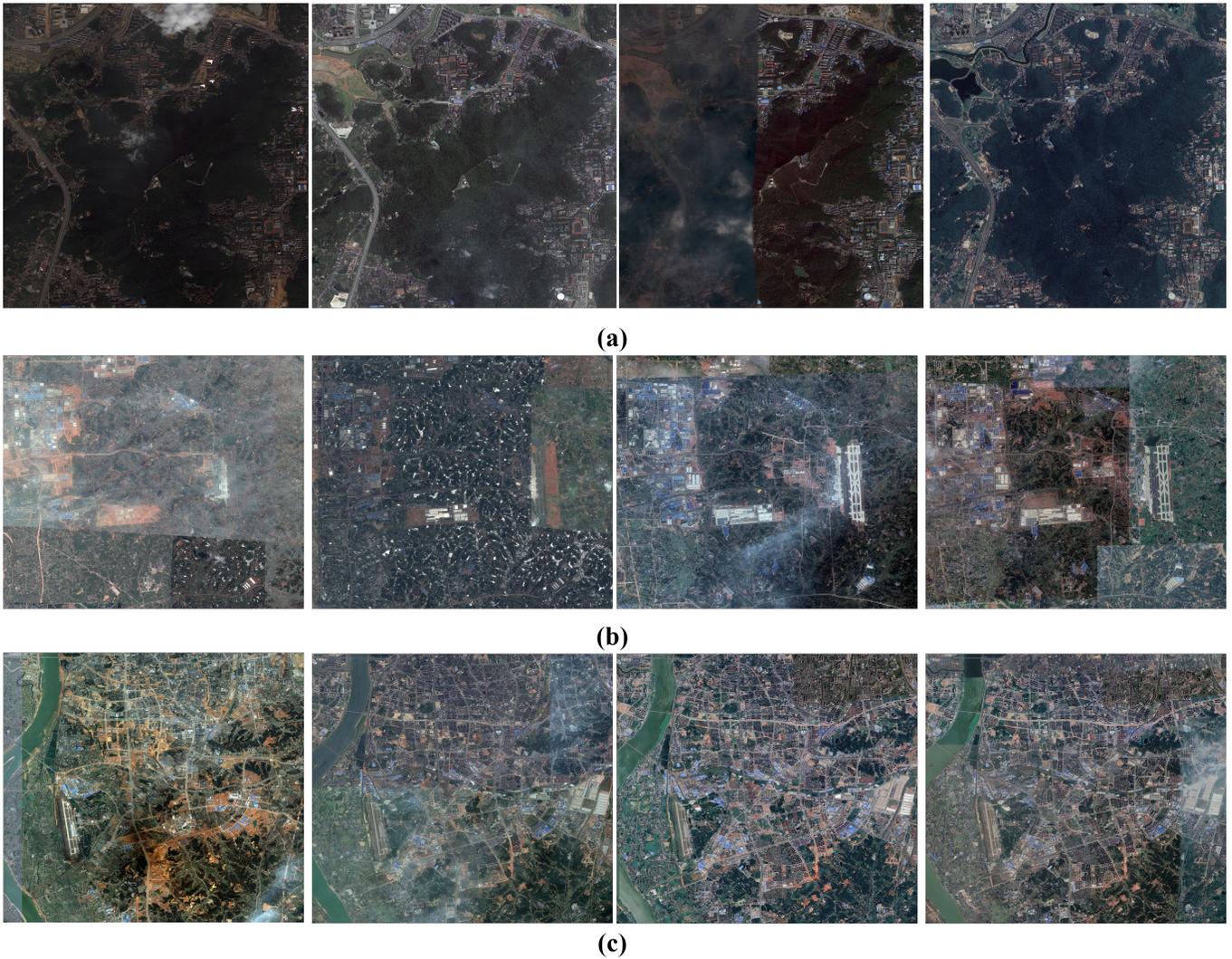


Fig. 7. Multi-temporal large scale high-resolution remote sensing images. (a) The University Student Science and Technology City area in Changsha. (b) The Huanghua Airport area in Changsha. (c) The urban area of Changsha.

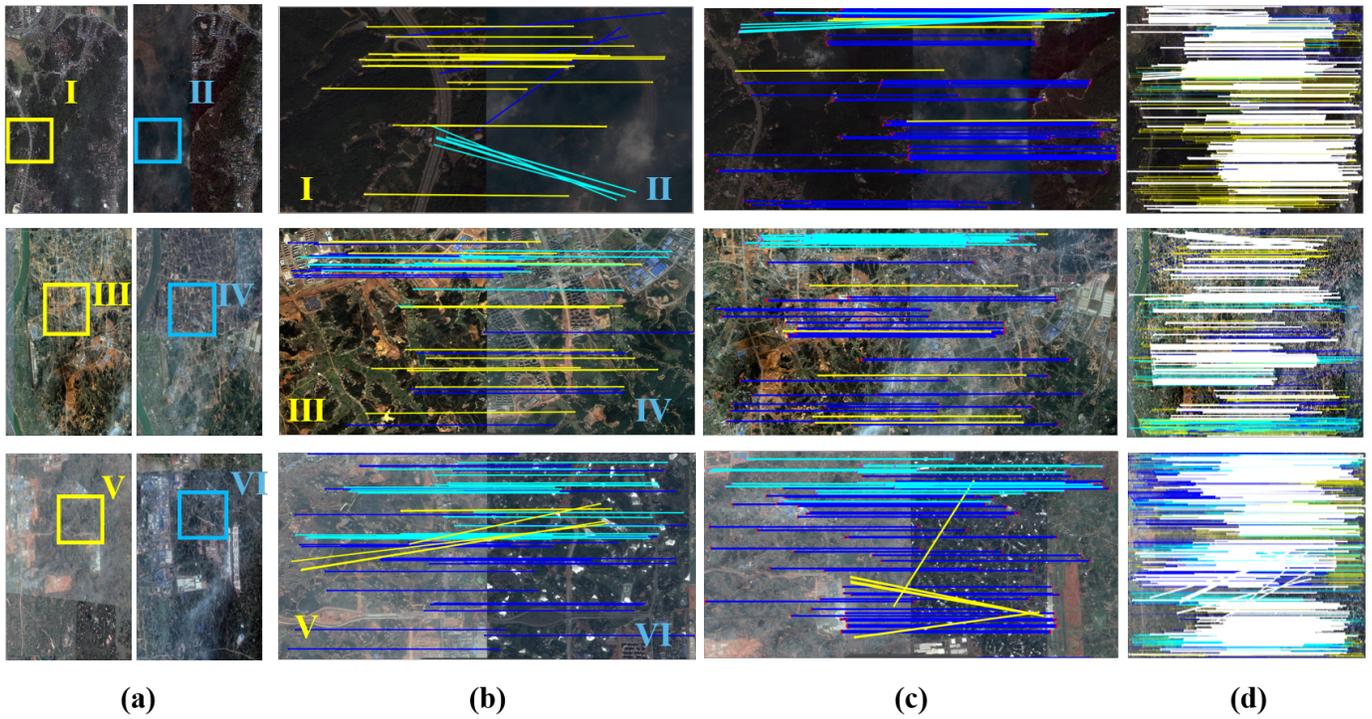


Fig. 8. The columns (a) are the test image pair 2, 4, and 7 from top to bottom. Column (b) is the feature error matching map of the partial image blocks in the yellow box and blue box of columns (a). Column (c) is the ARMF algorithm feature matching of the amplified image. The column (d) is the final global image feature matching map obtained by the ARMF algorithm. Note that, The light green line represents the matching feature points in the line feature we used, the blue line represents the SIFT matching point feature, the yellow line represents the RIFT phase feature, and the white line represents the new matching feature obtained by the ARMF algorithm.

TABLE II  
EXPERIMENTAL RESULTS FOR IMAGE NO 2. THE BEST RESULTS IN THIS TABLE HAVE BEEN MARKED.

Methods	RMSE	NOCC	MI	ROCC
Manual	*	10	0.1007	*
SIFT	35.8814	246	0.1331	0.1667
SURF	*	*	*	*
RIFT	27.4317	486	0.1418	0.0843
<b>Proposed+SIFT</b>	31.9110	576	0.1371	<b>0.3191</b>
<b>Proposed+RIFT</b>	26.7091	534	0.1452	0.0918
<b>Proposed+SIFT+APAP</b>	20.9653	576	0.1655	<b>0.3191</b>
<b>Proposed+RIFT+APAP</b>	8.6876	534	0.1869	0.0918
<b>Proposed+APAP</b>	<b>5.9181</b>	<b>1198</b>	<b>0.1930</b>	0.1073

TABLE III  
EXPERIMENTAL RESULTS FOR IMAGE NO 4. THE BEST RESULTS IN THIS TABLE HAVE BEEN MARKED.

Methods	RMSE	NOCC	MI	ROCC
Manual	12.0994	10	0.0257	*
SIFT	7.5104	1248	0.0581	0.2323
SURF	*	*	0.0558	*
RIFT	5.0992	619	0.0651	0.0542
<b>Proposed+SIFT</b>	6.9672	1625	0.0598	<b>0.2827</b>
<b>Proposed+RIFT</b>	3.3518	716	0.0749	0.0621
<b>Proposed+SIFT+APAP</b>	2.8563	1625	0.0842	<b>0.2827</b>
<b>Proposed+RIFT+APAP</b>	*	716	*	0.0621
<b>Proposed+APAP</b>	<b>1.7651</b>	<b>3025</b>	<b>0.0853</b>	0.1101

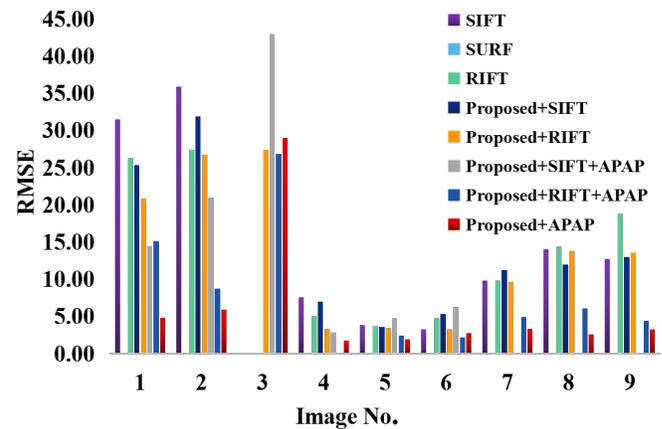


Fig. 9. This histogram represents the RMSE test value results of different feature algorithms on our data set.

sensing image with MI. The MI are calculated as follows:

$$MI(R, T) = H(R) + H(T) - H(R, T) \quad (34)$$

where  $H(R)$  and  $H(T)$  are the information entropy of the reference image and the target image. The joint entropy in the reference image and the target image is  $H(R, T)$ .

$$H(R) = - \sum_{i=0}^{255} PX(i) \log(PX(i)) \quad (35)$$

TABLE IV  
EXPERIMENTAL RESULTS FOR IMAGE NO 7. THE BEST RESULTS IN THIS TABLE HAVE BEEN MARKED.

Methods	RMSE	NOCC	MI	ROCC
Manual	*	10	0.0191	*
SIFT	9.8022	170	0.0331	0.1847
SURF	*	*	0.0176	*
RIFT	9.7730	320	0.0345	0.0582
<b>Proposed+SIFT</b>	11.2113	284	0.0304	<b>0.2746</b>
<b>Proposed+RIFT</b>	9.6149	327	0.0346	0.0594
<b>Proposed+SIFT+APAP</b>	*	284	*	<b>0.2746</b>
<b>Proposed+RIFT+APAP</b>	4.8741	327	0.0485	0.0594
<b>Proposed+APAP</b>	<b>3.2885</b>	<b>791</b>	<b>0.0503</b>	0.0744

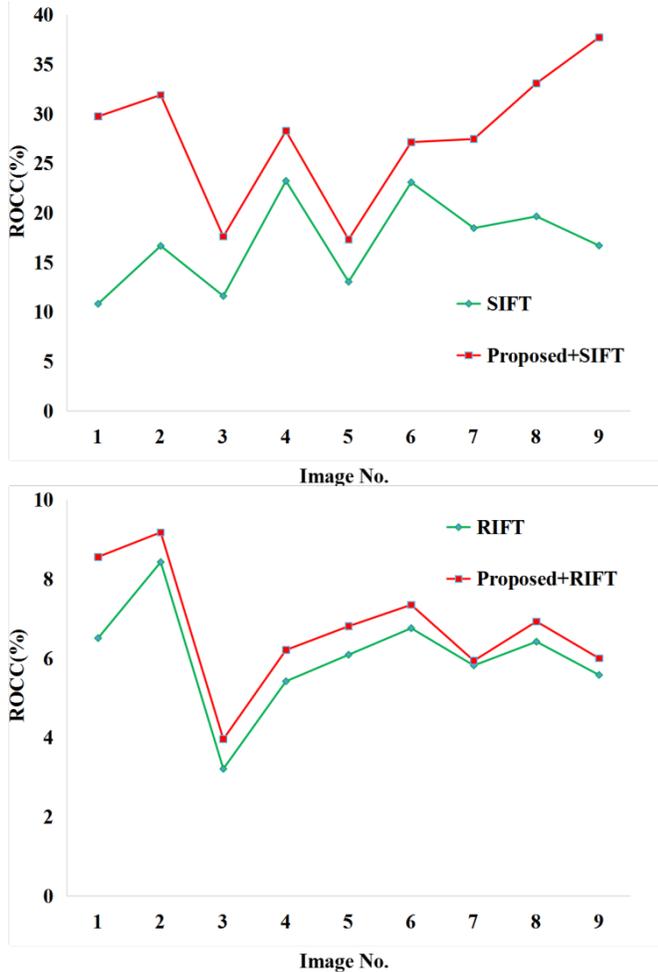


Fig. 10. The line chart is the result of ROCC comparison between our ARMF algorithm and RIFT and SIFT respectively.

$$H(R, T) = - \sum_{i=0}^{255} PXY(i, j) \log(PXY(i, j)) \quad (36)$$

where  $PX(i)$  represents the gray-level probability distribution of  $X$ , and the joint probability distribution of  $X$  and  $Y$  is represented by  $PXY(i, j)$ . Among them, the  $XY$  joint probability distribution and marginal probability distribution of  $X$  and  $Y$  are obtained through statistical histograms.

Usually, there are many incorrect matching point pairs in the image matching result. The number of false correspon-

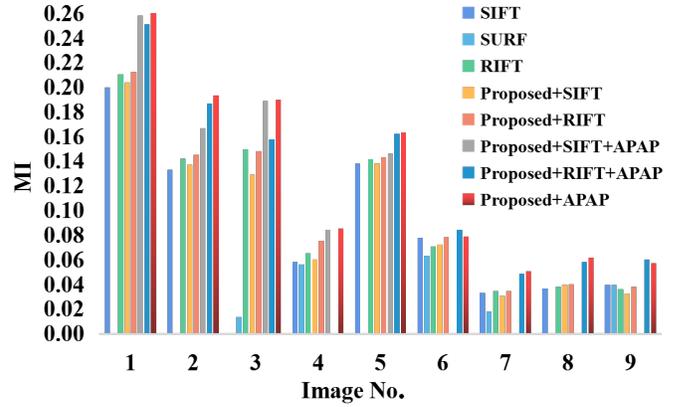


Fig. 11. This histogram represents the MI test value results of different feature algorithms on our data set.

dences (NOFC) has a negative impact on the accuracy of the registration result. Therefore, we use the ROCC to evaluate the robustness of the registration method. All methods and experiments are implemented on Intel Core i7-7700 4.2-GHz processor, NVIDIA GeForce GTX 1080ti with 16 GB of physical memory.

$$ROCC = \frac{NOCC}{NOCC + NOFC} \quad (37)$$

### C. Results and Discussion

In summary, we want to verify the feature robustness, accuracy and advantages of multiple feature combinations in adaptive region amplification, we have selected three remote sensing image configurations: SIFT [16], SURF [19], and RIFT [41]. These methods are compared with the proposed method. In addition, these methods are widely used in remote sensing image processing. We tested on multi-temporal data sets in different area. Figs. 9-11 depict the comparative results of the RMSE, ROCC, and MI. In addition, we use RANSAC [42] to remove outliers. The  $d_{ratio}$  of all the compared methods is set to 0.6, and the line segment edges detection thresholds value is set to 0.2.

From the results, our algorithm is better than the results of SIFT, SURF and RIFT when it is applied to large-scale remote sensing image pairs. The accuracy of MI and RMSE, and the robustness of multi-temporal remote sensing image feature matching also are better. In order to ensure the fairness of the data results, we selected 5-10 different image local regions when testing the registration accuracy of RMSE. The results of multiple image regions are averaged to reduce the error caused by the feature matching error elimination algorithm. All the results use the same set of error removal and calculation methods. Due to the changes in the local background features of the large-scale multi-temporal images, most of the image local region have mismatch features or the few matching features. In Fig. 8, we select three images (pairs in Fig.7) from different area to show the specific image matching result and comparison. In Tables II-IV, where “\*” represents the feature matching failure. Compared with the algorithms, the results obtained by our method with the lower RMSE.

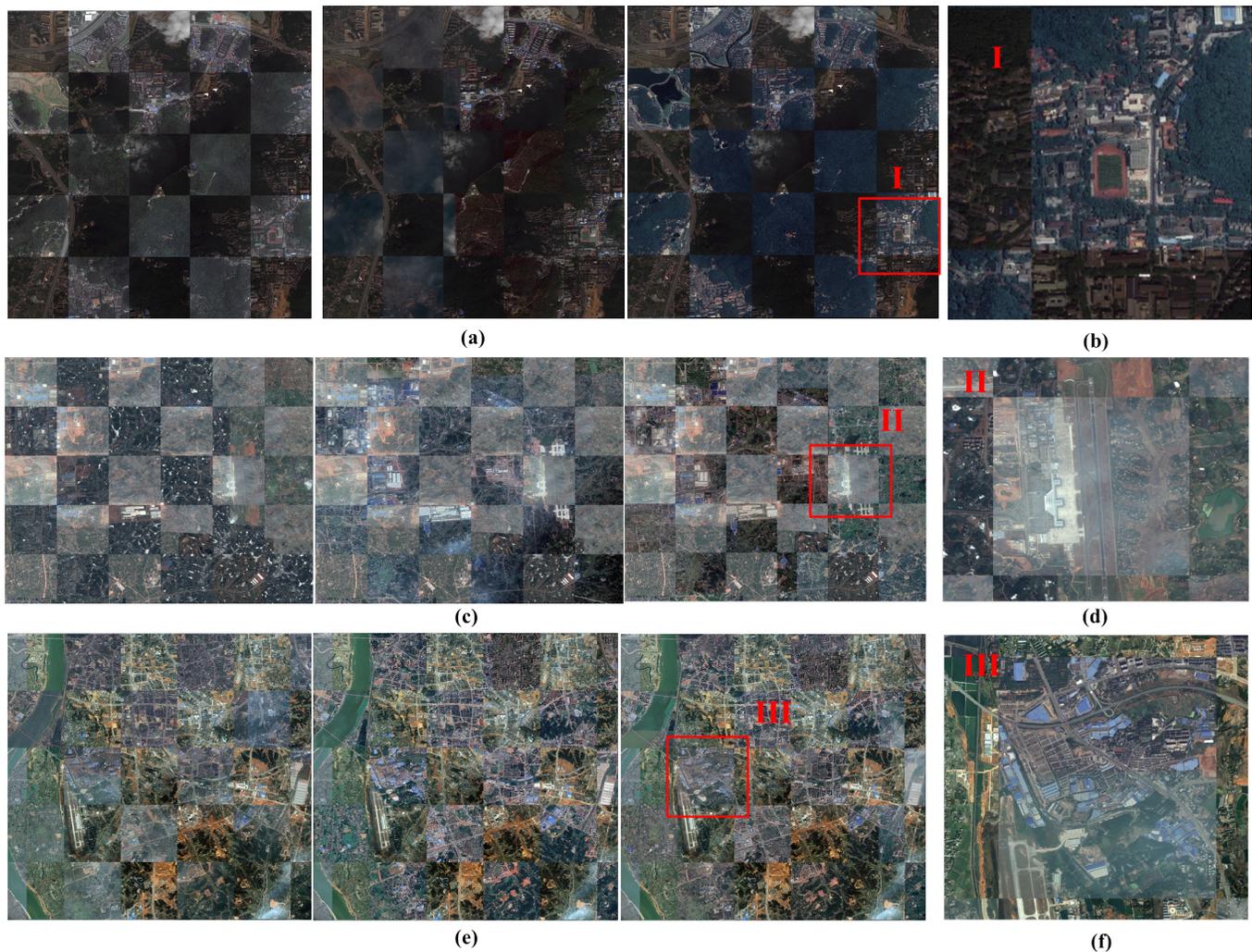


Fig. 12. (a), (c) and (e) the results of checkerboard mosaic images from multi-temporal image pairs; (b) The local image region I, (d) The local image region II, (f) The local image region III.

We made an experimental comparison of ROCC between SIFT and RIFT. In the experiment, the NOCC may be repeated, but does not affect the calculation of the final transformation model. The main results are shown in Fig. 10. It can be seen from the data comparison that our adaptive region amplification and feature search strategy improves the accuracy of the original features and the number of matches. Finally, we can find a sufficient number of correct feature matches.

The histogram shown in Fig. 11 is the MI value of the image after testing nine pairs of multi-temporal image data with our method. It can be seen from the figure that our method also improves the registration accuracy of the image correspondingly, while, the similarity ratio is higher.

The ARMF algorithm extracts a sufficient number of correct feature matching relationship pairs and we obtain a more uniform matching feature distribution, which increases the features robustness of process in multi-temporal large scale high-resolution remote sensing image. The registration result is shown in Fig. 12, and we use the checkerboard stitching image

to represent the final registration result, which can intuitively see the degree of deviation between the two images. It can be seen that our method has good robustness and applicability on multi-temporal large scale high-resolution remote sensing images from the figure, and provides an excellent solution for the accurate registration of large scale remote sensing images.

## V. CONCLUSIONS

A robust and accurate multi-temporal large-scale high resolution remote sensing image registration strategy called ARMF was proposed in this paper. Our algorithm included: 1) Adaptive multiple features extraction and combination; 2) Adaptive image region amplification; 3) Image region downsampling and feature remapping. The proposed ARMF algorithm can not only adaptively eliminate the good correspondences matching pairs among multiple features, but also adaptively selected the number of features through pyramid image regions amplification. Then, we downsample the amplified image region and feature remapping, thus, reducing the cost consumption of the overall algorithm. Finally, we calculate the feature distribution

after all the ARMF algorithm to achieve the registration of the image. Experimental results on large-scale image data demonstrated the effectiveness of the proposed method. In the future, we will research how to efficiently and quickly improve the processing speed of large scale image registration, so that it can be efficiently applied to image processing research fields, such as image stitching, change detection, etc.

#### ACKNOWLEDGMENT

The authors would like to thank the handling editor and anonymous reviewers for their valuable comments and suggestions, which are of great help in improving the quality of this paper.

#### REFERENCES

- [1] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, Nov. 2003.
- [2] S. Shi, Y. Zhong, J. Zhao, P. Lv, Y. Liu, and L. Zhang, "Land-use/land-cover change detection based on class-prior object-oriented conditional random field framework for high spatial resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, pp. 1–16, 2020.
- [3] D. Peng, L. Bruzzone, Y. Zhang, H. Guan, H. Ding, and X. Huang, "Semicdnet: A semisupervised convolutional neural network for change detection in high resolution remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5891–5906, Jul. 2021.
- [4] Y. Guo, C. Huang, Y. Zhang, Y. Li, and W. Chen, "A novel multitemporal image-fusion algorithm: Method and application to goci and himawari images for inland water remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 4018–4032, Jun. 2020.
- [5] Y. Zhou, A. Rangarajan, and P. D. Gader, "An integrated approach to registration and fusion of hyperspectral and multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3020–3033, May. 2020.
- [6] R. Liu, L. Mi, and Z. Chen, "Afnet: Adaptive fusion network for remote sensing image semantic segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7871–7886, Sep. 2021.
- [7] H. Gonçalves, J. Gonçalves, L. Corte-Real, and A. Teodoro, "Chair: Automatic image registration based on correlation and hough transform," *Int. J. Remote Sens.*, vol. 33, no. 24, pp. 7936–7968, Jul. 2012.
- [8] Y. Han and J. Oh, "Automated geo/co-registration of multi-temporal very-high-resolution imagery," *Sensors*, vol. 18, no. 5, p. 1599, May. 2018.
- [9] S. Skakun, J.-C. Roger, E. Vermote, C. Justice, and J. Masek, "Automatic co-registration of multi-temporal landsat-8/oli and sentinel-2a/msi images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 5272–5274.
- [10] Y. Han, F. Bovolo, and L. Bruzzone, "Segmentation-based fine registration of very high resolution multitemporal images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2884–2897, May. 2017.
- [11] Y. Li, Q. Liu, L. Jing, S. Liu, and F. Miao, "A genetic-optimized multi-angle normalized cross correlation sift for automatic remote sensing registration," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 2586–2589.
- [12] X. Xie, Y. Zhang, X. Ling, and X. Wang, "A new registration algorithm for multimodal remote sensing images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 7011–7014.
- [13] Z. Yang, T. Dan, and Y. Yang, "Multi-temporal remote sensing image registration using deep convolutional features," *IEEE Access*, vol. 6, pp. 38 544–38 555, 2018.
- [14] J. Ma, H. Zhou, A. Fan, J. Zhao, Y. Gao, J. Jiang, and J. Tian, "Robust feature matching for remote sensing image registration via locally linear transforming," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6469–6481, Dec. 2015.
- [15] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, "Image matching from handcrafted to deep features: A survey," *Int. J. Comput. Vis.*, vol. 129, no. 1, pp. 23–79, Jan. 2021.
- [16] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Feb. 2004.
- [17] A. Sedaghat, M. Mokhtarzade, and H. Ebadi, "Uniform robust scale-invariant feature matching for optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4516–4527, Oct. 2011.
- [18] A. Sedaghat and H. Ebadi, "Remote sensing image matching based on adaptive binning sift descriptor," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5283–5293, Oct. 2015.
- [19] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2006, pp. 404–417.
- [20] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2010, pp. 778–792.
- [21] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2006, pp. 430–443.
- [22] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 2564–2571.
- [23] R. G. Von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall, "Lsd: a line segment detector," *Image Process. On Line*, vol. 2, pp. 35–55, 2012.
- [24] W. Rong, Z. Li, W. Zhang, and L. Sun, "An improved canny edge detection algorithm," in *IEEE Int. Conf. Adv. Intell. Mechatron.* IEEE, 2014, pp. 577–582.
- [25] W. Gao, X. Zhang, L. Yang, and H. Liu, "An improved sobel edge detection," in *Int. Conf. Comput. Sci. Inf. Technol.*, vol. 5. IEEE, 2010, pp. 67–71.
- [26] Y. Chen, X. Zhang, Y. Zhang, S. J. Maybank, and Z. Fu, "Visible and infrared image registration based on region features and edginess," *Mach. Vis. Appl.*, vol. 29, no. 1, pp. 113–123, Jan. 2018.
- [27] X. Huang, Y. Sun, D. Metaxas, F. Sauer, and C. Xu, "Hybrid image registration based on configurational matching of scale-invariant salient region features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. Workshops*, 2004, pp. 167–167.
- [28] S. Wei, L. Pu, X. Tang, X. Zhang, and J. Shi, "Efficient registration for insar large-scale image using quadtree segmentation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 4435–4438.
- [29] B. Ayhan, M. Dao, C. Kwan, H.-M. Chen, J. F. Bell, and R. Kidd, "A novel utilization of image registration techniques to process mastcam images in mars rover with applications to image fusion, pixel clustering, and anomaly detection," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 10, no. 10, pp. 4553–4564, Oct. 2017.
- [30] C. Ying, L. Guoqing, and C. Hengshi, "Multi-temporal remote sensing image registration based on multi-layer feature fusion of deep residual network," in *Int. Conf. Intel. Inf. Biomed. Sci.*, 2019, pp. 363–367.
- [31] W. Ma, J. Zhang, Y. Wu, L. Jiao, H. Zhu, and W. Zhao, "A novel two-step registration method for remote sensing images based on deep and local features," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4834–4843, Jul. 2019.
- [32] Y. Ye, J. Shan, L. Bruzzone, and L. Shen, "Robust registration of multimodal remote sensing images based on structural similarity," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2941–2958, May. 2017.
- [33] Y. Hel-Or, H. Hel-Or, and E. David, "Matching by tone mapping: Photometric invariant template matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 317–330, Feb. 2014.
- [34] M. Gong, S. Zhao, L. Jiao, D. Tian, and S. Wang, "A novel coarse-to-fine scheme for automatic image registration based on sift and mutual information," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4328–4338, Jul. 2014.
- [35] X. Fan, H. Rhody, and E. Saber, "A spatial-feature-enhanced mmi algorithm for multimodal airborne image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 6, pp. 2580–2589, Jun. 2010.
- [36] H. Foroosh, J. B. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," *IEEE Trans. Image Process.*, vol. 11, no. 3, pp. 188–200, Mar. 2002.
- [37] S. Cui, Y. Zhong, A. Ma, and L. Zhang, "A novel robust feature descriptor for multi-source remote sensing image registration," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 919–922.
- [38] X. Jiang, J. Ma, G. Xiao, Z. Shao, and X. Guo, "A review of multimodal image matching: Methods and applications," *Inf. Fusion*, 2021.
- [39] 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, p. 426, May. 2016.
- [40] C. Akinlar and C. Topal, "Edlines: A real-time line segment detector with a false detection control," *Pattern. Recognit. Lett.*, vol. 32, no. 13, pp. 1633–1642, 2011.
- [41] 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, 2020.
- [42] R. Raguram, J.-M. Frahm, and M. Pollefeys, "A comparative analysis of ransac techniques leading to adaptive real-time random sample

consensus,” in *Proc. Eur. Conf. Comput. Vis.* Springer, 2008, pp. 500–513.

- [43] H.-m. Chen, A. Goela, G. J. Garvin, and S. Li, “A parameterization of deformation fields for diffeomorphic image registration and its application to myocardial delineation,” in *Med. Image Comput. Comput.* Springer, 2010, pp. 340–348.
- [44] J. Zaragoza, T.-J. Chin, Q.-H. Tran, M. S. Brown, and D. Suter, “As-projective-as-possible image stitching with moving dlt,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1285–1298, Jul. 2014.
- [45] H. Goncalves, J. A. Goncalves, and L. Corte-Real, “Measures for an objective evaluation of the geometric correction process quality,” *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 2, pp. 292–296, Feb. 2009.



**Zezhou Li** received the B.S. degree from Northeast Forestry University, Harbin, China, in 2019, where he is currently pursuing the M.S. degree with the College of Electrical and Information Engineering, Hunan University, Changsha, China.

His research interests include remote sensing image processing, registration, etc.



**Jun Yue** received the B.Eng. degree in geodesy from Wuhan University, Wuhan, China, in 2013 and the Ph.D. degree in GIS from Peking University, Beijing, China, in 2018.

He is currently an Assistant Professor with the Department of Geomatics Engineering, Changsha University of Science and Technology. His research interests include satellite image understanding, pattern recognition, and few-shot learning. Dr. Yue serves as a reviewer for IEEE Transactions on Geoscience and Remote Sensing, ISPRS Journal of Photogrammetry

and Remote Sensing, IEEE Geoscience and Remote Sensing Letters, IEEE Transactions on Biomedical Engineering, IEEE Journal of Biomedical and health Informatics, Information Fusion, Information Sciences, International Journal of Remote Sensing, Remote Sensing Letters, etc.



**Leyuan Fang** (S'10-M'14-SM'17) received the Ph.D. degree from the College of Electrical and Information Engineering, Hunan University, Changsha, China, in 2015.

From September 2011 to September 2012, he was a visiting Ph.D. student with the Department of Ophthalmology, Duke University, Durham, NC, USA, supported by the China Scholarship Council. From August 2016 to September 2017, he was a Postdoc Researcher with the Department of Biomedical Engineering, Duke University, Durham, NC, USA. He

is currently a Professor with the College of Electrical and Information Engineering, Hunan University. His research interests include sparse representation and multi-resolution analysis in remote sensing and medical image processing. He is the associate editors of IEEE Transactions on Image Processing, IEEE Transactions on Geoscience and Remote Sensing, IEEE Transactions on Neural Networks and Learning Systems, and Neurocomputing. He was a recipient of one 2nd-Grade National Award at the Nature and Science Progress of China in 2019.