ISPRS Journal of Photogrammetry and Remote Sensing 196 (2023) 1–15

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Histogram of the orientation of the weighted phase descriptor for multi-modal remote sensing image matching

Yongjun Zhang [a,*], Yongxiang Yao [a], Yi Wan [a,*], Weiyu Liu [a], Wupeng Yang [a], Zhi Zheng [a], Rang Xiao [b]

[a] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China
[b] Guizhou Tuzhi Information Technology Co., Ltd, Guiyang 550081, China

ARTICLE INFO

Keywords:
Multi-modal remote sensing image
Aggregation feature
Weighted phase orientation feature
Log-polar of regularized
Bidirectional matching

ABSTRACT

Multi-modal remote sensing image (MRSI) has nonlinear radiation distortion (NRD) and significant contrast differences to which image gradient features are usually sensitive. Although image phase features are more robust against NRD, they might not be much helpful in resolving the problems of directional inversion or phase extreme value mutations that are common in the phase feature calculation. To address these issues, a new MRSI matching method—"histogram of the orientation of weighted phase" (HOWP)—is proposed in this paper. This method distinguishes itself from other methods in three aspects: (1) a feature aggregation strategy is used to optimize feature points by extracting the corner and blob features separately; (2) a novel weighted phase orientation model is established to replace the traditional image gradient orientation features; and (3) a regularization-based log-polar descriptor is constructed to generate robust feature description vectors. To evaluate the performance of the proposed method, we selected 50 sets of typical MRSIs with translation, scale, and rotation differences for comparison with the other four state-of-the-art methods. The results show that our method is more resistant to radiometric distortion and the contrasting differences in MRSIs. It also performs better in tackling the problems of direction reversal and phase extreme value mutation, as evidenced by more, the number of correct matches (NCM). Since the method has improved the average NCM by 1.6–4.5 times, the average success rate by 35.5%, and the average rate of correct matches by 11.1% with an average root of mean-squared error of 1.93 pixels. Moreover, we have put forward an extended version of the HOWP method (Simplified-HOWP) when there is no image rotation, which manifests in an average 0.75 times improvement in NCM of Simplified-HOWP performance over that of the HOWP method. The executable code and test data are linked in https://skyearth.org/publication/project/HOWP/.

1. Introduction

The rapid development of modern technologies in multi-sensor information processing has assured unhindered access to multi-source/multi-modal image data. Multi-modal remote sensing image (MRSI) matching technology—the process of overlaying two or more images of the same scene taken at different times with different sensors or under different imaging conditions—has been a hot research topic in recent years (Yao et al., 2022). It is widely used in image fusion, change detection, image mosaic, and 3D reconstruction. However, MRSIs have nonlinear radiation distortion (NRD), spectral differences, and great contrast differences due to different imaging mechanisms, and among them, NRD is a serious impediment to MRSI matching. Traditional matching methods do not work well for MRSIs with remarkable NRD, for they inevitably reduce the similarity between corresponding points. Therefore, MRSI matching is still a challenge that is to be solved in current studies.

In the past few years, substantial research on MRSI matching has been carried out, and the methods used can be generally classified into three categories (Xu et al., 2016): area-based matching, feature-based matching, and deep learning-based matching. The area-based method is used to calculate the image similarity mainly by measuring image intensity and mutual information (Viola et al., 1997). With this method, high matching accuracy can be guaranteed. Unfortunately, it is also

* Corresponding authors.
E-mail addresses: zhangyj@whu.edu.cn (Y. Zhang), yaoyongxiang@whu.edu.cn (Y. Yao), yi.wan@whu.edu.cn (Y. Wan), liuwy0225@whu.edu.cn (W. Liu), yangwupeng@whu.edu.cn (W. Yang), zhengzhi@whu.edu.cn (Z. Zheng), xiaorang@ituzhi.com (R. Xiao).

https://doi.org/10.1016/j.isprsjprs.2022.12.018
Received 9 June 2022; Received in revised form 12 November 2022; Accepted 20 December 2022
Available online 30 December 2022

characterized by a heavy computing burden and poor invariance to the scale and rotation of the image, making its scope of application greatly limited. The feature-based matching method concerns image gradient features and often yields satisfactory results when dealing with translation, scale, and rotation differences. However, it is not applicable in MRSI matching due to the sensitivity of image gradient features to the NRD and contrast differences in MRSIs. Although breakthroughs have been made by applying phase features to MRSI matching, such as histogram of phase congruency (PC) direction (Ye et al., 2018) and RIFT (Li et al., 2020) algorithms, these algorithms all constrain the scale and rotation invariance of the algorithm to some extent. Despite attempts to improve them, problems such as phase direction inversion and phase extreme value mutation of MRSIs still exist. Deep learning-based methods, such as LoFTR (Sun et al., 2021), have shown great potential in optical image matching but have not achieved superior performance in MRSI matching. However, it is not an optimal solution either, owing to the difficulty in collecting MRSI sample data and the complex application scenarios of remote sensing images.

As summarized above, the ability to extract significant feature points and robust descriptors become the key to MRSI matching success. In feature extraction, most scholars have focused on finding corner or blob features of images; it is more difficult to describe the advantages of these two feature points simultaneously. In feature description, the phase features extracted using the PC model are usually contaminated with direction reversal and phase extreme value mutation, which would surely affect the shape component of the image represented by the phase, thus destroying the structural integrity of the image. Additionally, the information provided by feature descriptors may not be accurate for the phase orientation features and cannot correctly characterize the directional changes between images.

Traditional image feature description is mainly conducted using gradient amplitude and gradient direction of the image. These gradient features are more sensitive to MRSIs, which cannot characterize the image feature changes robustly. Some scholars adopted the PC model instead of the gradient features of images to better eliminate the NRD, illumination differences, and contrast between MRSI images. However, recent studies have proved that the phase features the model constructs cannot effectively tackle the problems of directional inversion and phase extreme value mutation, which undermine the robustness of the descriptors, especially when scale-change or rotation exists. Against this backdrop, we propose a novel MRSI matching method based on the histogram of the orientation of the weighted phase (HOWP) descriptor, aiming to achieve robust matching of MRSI (Fig. 1). The main contributions of this study are as follows:

- An aggregated feature point optimization strategy is proposed. The strategy uses the advantages of both blob points and corner points to improve the richness of image feature points. These feature points can be better used to describe the feature vector for our proposed descriptor;
- An HOWP feature model and log-polar description method based on regular grid division are proposed. The HOWP features generated by this model are used to replace the traditional image gradient orientation features to address the directional inversion and phase extreme value mutation of phase features, which also improvs the scale and rotation invariance of the method. The model is robust against NRDs, illumination differences, and contrast differences of MRSIs. The log-polar description utilization helps improve the robustness of MRSI feature descriptors.

This paper is structured as follows: Section 1 presents the research purposes and limitations of previous studies, followed by a brief explanation of the significance of the present study in Section 2. The processing procedure of the proposed HOWP method is introduced in Section 3, and Section 4 gives a detailed account of the experimental analysis. The paper is concluded with a discussion of the effects of different parameter settings on HOWP performance in Section 5 and a summary of the study's contributions in Section 6.

2. Related work

MRSI methods can be mainly categorized as follows (Xu et al., 2016): feature-based matching, deep learning-based matching, and area-based matching, which will be elucidated below.

Area-based matching methods simply use the grayscale information of the images to construct a similarity measurement for matching without extracting the common features of images. Frequently used similarity measurements include correlation coefficient (Cole-Rhodes et al., 2003) and mutual information (Chen et al., 2003). The correlation coefficient is invariant with respect to grayscale changes under linear transformations, but it is sensitive to the grayscale differences between images, thus making it difficult to match MRSIs. Mutual information is better for NRD resistance, but the method presents the local optimal solutions, which usually affect MRSI matching performance. The CFOG algorithm (Ye et al., 2019) that came out later also has its limitations in that it relies greatly on geographic coordinate information.

Feature-based matching methods, such as speeded-up robust features (SURF) (Bay et al., 2008), ORB (Rublee et al., 2011), KAZE (Alcantarilla et al., 2012), PSO-SIFT (Ma et al., 2016), and scale-invariant feature transform (SIFT)-like (Öfverstedt et al., 2019) algorithms, have been extensively developed since Lowe (1999) proposed to SIFT matching. Previous studies have examined these methods from different perspectives, such as scale robustness, rotation invariance, binary description optimization, computational cost optimization, and feature matching enhancement. However, they still have certain limitations for MRSI with intensity differences and NRD differences. Scholars then resorted to phase features of the image in the frequency domain and have proposed

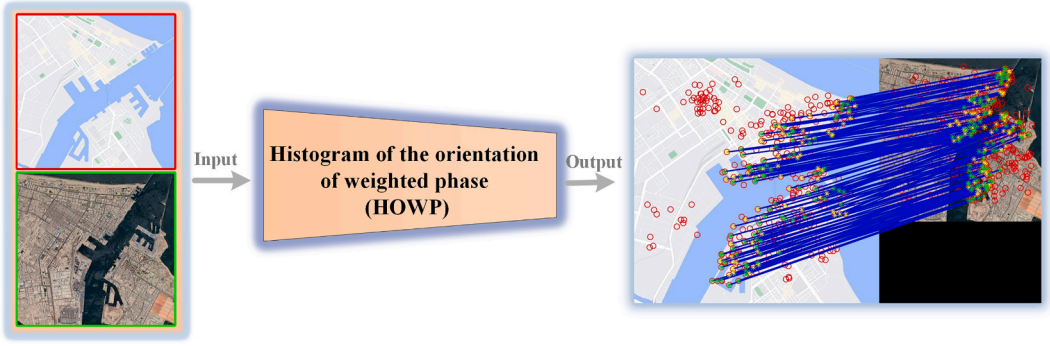


Fig. 1. Matching results of the "histogram of the orientation of weighted phase" method (blue lines represent successful matching point pairs). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Log-Gabor filter optimized matching (LGHD) (Aguilera et al., 2015), robust line segment descriptor matching (Zhao et al., 2016), coarse-to-fine matching for MRSI (3MRS) (Fan et al., 2022), histogram of absolute phase consistency gradients (HAPCG) (Yao et al., 2021), and locally normalized images for the rotation invariant matching method (Li et al., 2022). These algorithms have effectively advanced MRSI matching research, but their applicability is constrained by factors such as geographical location, scale, rotation, and computational complexity.

As deep learning technologies evolve so quickly, they have drawn scholarly attention and been increasingly utilized in MRSI matching, such as the use of the convolutional neural network in the estimation of mapping relationships between image blocks (Wang et al., 2018) and image matching based on spatial attention mechanism (Wiles et al., 2021), the use of D2-Net in multi-source image feature extraction and description (Dusmanu et al., 2019), the feature extraction by VGG network (Efe et al., 2021), and a novel hierarchical extract-and-match transformer (Wang et al., 2022) algorithm based on Transformers converter, matching frameworks based on convolutional gradient features (Zhou et al., 2021) and multiscale frameworks with unsupervised learning (Ye et al., 2022). Deep learning image matching methods present the advantages of being fast and automatically optimizing the parameters and constructing the wanted descriptors. However, due to the large geological differences between MRSIs and the difficulty obtaining training samples, this method's generalization ability and applicability are limited.

3. Method

The proposed HOWP method is implemented in three steps: (i) aggregate feature extraction; (ii) HOWP descriptor construction (the orientation weighted phase feature generation and regular log-polar description feature vector computation are also included); (iii) bidirectional matching. The whole process is shown in Fig. 2.

3.1. Aggregation feature extraction

Feature point extraction is an important part of image matching. The non-linear filtering and the PC model facilitate the extraction of the image's blob and corner point features. Therefore, we adopted the image anisotropy diffusion method optimized by Yao et al. (2021) to establish the nonlinear filtering equation to achieve the diffusion. After diffusion, the PC calculation was performed to obtain the phase map (Kovesi, 1999), i.e., for a better description of the edge features, we computed an independent mapping $PC(\theta_o)$ (Kovesi, 1999) for each direction o, where θ is the angle of direction o. According to the moment analysis method, the axis corresponding to the minimum moment is referred to as the principal axis, which usually represents the directional information of the feature. In this paper, we first construct an image pyramid to achieve

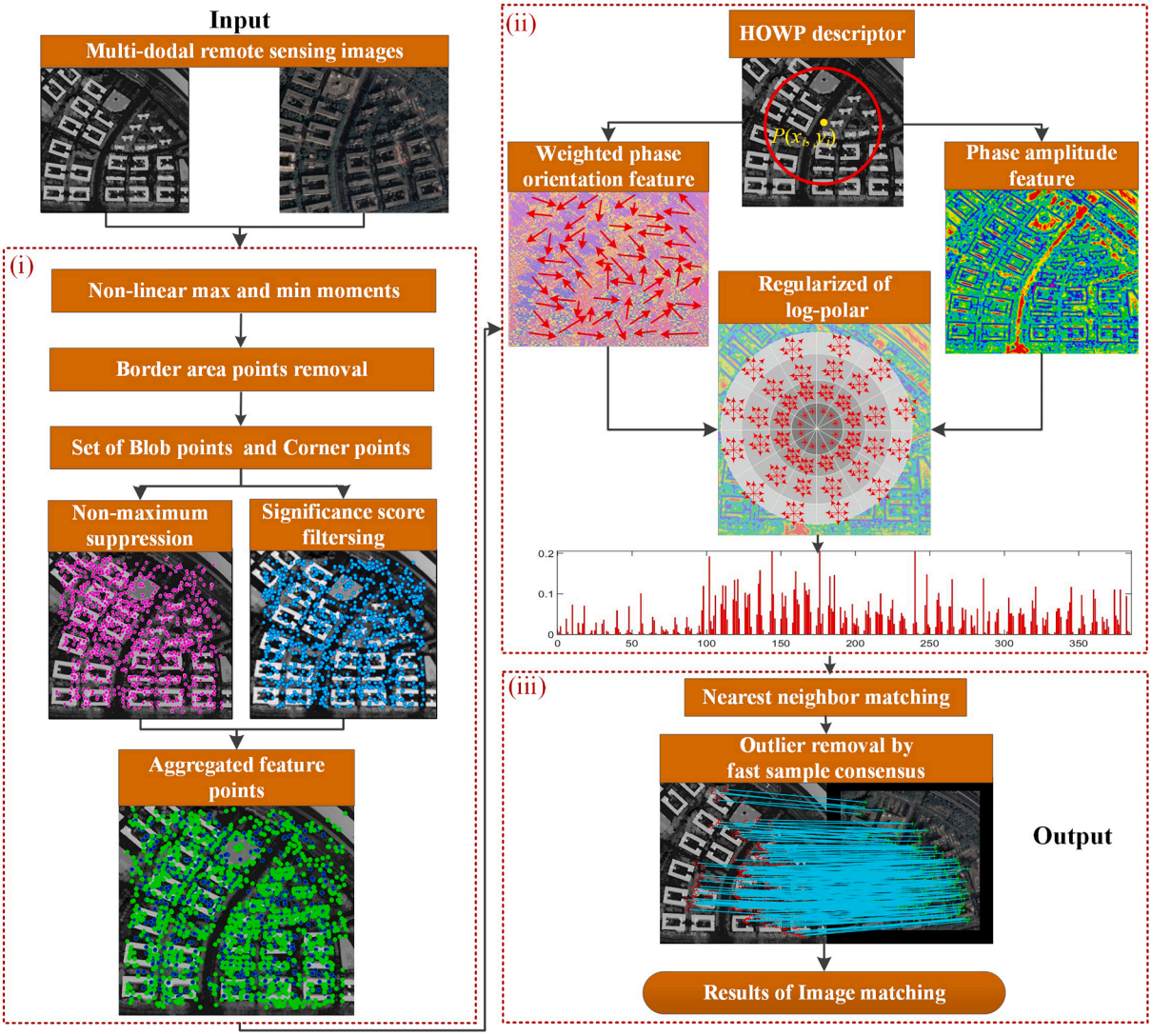


Fig. 2. The flow diagram of the HOWP method.

scale-invariant and then perform PC calculations on all images within the image pyramid. The magnitude of the maximum moment, which is about an axis perpendicular to the principal axis, indicates the salience of the feature. In this paper, the maximum and minimum moments are solved and normalized based on the PC model, and the formula is defined as (1):

$$\begin{cases} \boldsymbol{A} = \sum_{o} (PC'(\theta_o)\cos(\theta_o))^2 \\ \boldsymbol{B} = 2\sum_{o} (PC'(\theta_o)\cos(\theta_o))^{*}(PC'(\theta_o)\sin(\theta_o)) \\ \boldsymbol{C} = \sum_{o} (PC'(\theta_o)sin(\theta_o))^2 \end{cases} \tag{1}$$

$$\begin{cases} \boldsymbol{M}_{\max} = \frac{1}{2}\left(\boldsymbol{C} + \boldsymbol{A} + \sqrt{\boldsymbol{B}^2 + (\boldsymbol{A} - \boldsymbol{C})^2}\right) \\ \boldsymbol{M}_{\min} = \frac{1}{2}\left(\boldsymbol{C} + \boldsymbol{A} - \sqrt{\boldsymbol{B}^2 + (\boldsymbol{A} - \boldsymbol{C})^2}\right) \end{cases} \tag{2}$$

In Eqs. (1) and (2), $\boldsymbol{M}_{\max}$ represents the maximum moment of PC; $\boldsymbol{M}_{\min}$, the minimum moment; $PC(\theta o)$, the mapping of PC' in the o direction. **A**, **B**, and **C** are intermediate quantities of phase moment calculation, and θ denotes the angle in the o direction.

After pre-processing the image, the feature points can be extracted. Feature points are common in two types: Corners and Blobs. Blob points refer to areas that differ in color and grayscale from their surroundings. They have the advantages of being more resistant to noise and more stable. Corner points featuring saliency are usually referred to as areas where converging lines, edges, or sides meet. Traditional detectors only extract one type of the abovementioned key points, which is not profitable for image matching. In this regard, Ye et al. (2020) proposed a PC model that consists of an MMPC-Lap detector and a feature descriptor named local histogram of oriented PC (LHOPC), to increase the richness of feature points. However, as the number of feature points extracted increases, the computational cost and redundancy of the feature points also increase. Fortunately, Sedaghat and Mohammadi (2018) proposed a method of uniform competency to improve the matching performance of feature points that helps reduce computational costs.

Inspired by their research, we used a Hessian determinant (Lowe, 1999) to extract blob features and a FAST feature detector (Rosten and Drummond, 2006) to extract corner features. To reduce unnecessary duplicate points, significant feature points are retained. We also designed an aggregate feature (AF) strategy to optimize feature points, thereby enhancing the richness of the feature points. The aggregation feature optimization strategy has three steps: border area points' removal, non-maximum suppression, and significance score filtering. In the first step, the max-moment and min-moment images need to be masked to remove edge unstable regions, as shown in Eq. (3).

$$S_{\text{points}} = f_{Blob}\left(Mask(\cdot)|_{R=N_W/2} \otimes (\boldsymbol{M}_{\max})\right) + f_{Corner}\left(Mask(\cdot)|_{R=N_W/2} \otimes (\boldsymbol{M}_{\min})\right) \tag{3}$$

In Eqs. (3), S_{points} represent the set of blobs and corners after masking; f_{Blob} represents the blob extraction function; f_{Corner} represents the corner extraction function; $Mask(\cdot)|_{R=N_W/2}$ represents the mask function, R represents the mask radius; N_w represents the neighborhood window.

Next, non-maximum suppression is performed on the feature points obtained by Eq. (3), and the suppression function is represented by $f_{\text{nms}}(\cdot)$. Finally, significance score filtering was proposed. In feature extraction, each pixel strength value of images would be normalized and unified between 0 and 1, followed by a sorting work and construction of a saliency score extraction equation, as shown in (4):

$$\begin{cases} f_{\text{score}}(\cdot) = k\cdot\text{sortrows}\left(S_{\text{points}} | p_i^{score}\right) \\ \text{AF} = f_{\text{score}}\left(f_{nms}\left(S_{points}\right)\right) \end{cases} \tag{4}$$

In Eq. (4), S_{score} represents the set of points filtered by the salience score, k represents the feature points filtering threshold ($k = 0.85$), p_i^{score} represents its PC strength value, $f_{\text{nms}}(\cdot)$ represents the feature points retained after non-maximum suppression, and S_{points} denotes the points removed from the edge.

Those feature points retained after filtering would be defined as the final key points, as shown in Fig. 3.

3.2. Histogram of the orientation of weighted phase

The proposed HOWP method has four steps: (1) characterization of phase amplitude; (2) establishment of weighted phase orientation feature; (3) calculation of principal orientation of feature points; and (4) construction of regularized log-polar descriptor. Among them, (2) and (4) are steps we have mainly been concerned about. Using step (2), we have established the maximum noise and minimum noise response functions by phase noise. This function is combined with the even and odd symmetries of the Log-Gabor filter to establish a weighted phase orientation equation. Finally, the weighted phase orientation feature of the image is calculated using this equation. Step (4) is to generate feature vectors through a regularized grid log-polar coordinate description framework to improve descriptors' robustness. The whole workflow of the HOWP method is shown in Fig. 4.

Considering that the PC model has advantages over the traditional methods in terms of dealing with the sensitivity of the gradient amplitude and gradient direction, we used the phase-amplitude feature calculated by an improved PC model based on Log-Gabor filters proposed by Kovesi (1999), to replace the image gradient amplitude feature.

3.2.1. Weighted phase orientation feature

A PC model is developed as a feature detection method based on local energy, using the maximum Fourier component to calculate the structural features of the images. The algorithm uses a 2D-Log-Gabor filter to extract local phase information to convolve the image. Firstly, the 2D-Log-Gabor filter was constructed, and the 2D log-Gabor filter can generally be obtained by Gaussian spreading of the vertical direction of the log-Gabor filter (Fischer et al., 2007). Therefore, it can be defined as Eq. (5):

$$\boldsymbol{p}(\rho, \theta, s, o) = \exp\left(\frac{-(\rho - \rho_s)^2}{2\sigma_\rho^2}\right)\exp\left(\frac{-(\theta - \theta_{s,o})^2}{2\theta_\rho^2}\right) \tag{5}$$

where (ρ, θ) represents the log-polar coordinates; s and o are the scale and orientation of 2D-log-Gabor, respectively; (ρ_s, θ_s) is the centre frequency of 2D-log-Gabor; σ_ρ and σ_θ are the bandwidths in ρ and θ, respectively. As the 2D-log-Gabor filter is a frequency domain filter, whose corresponding spatial domain filter can be obtained by inverse Fourier transform. In the spatial domain, 2D-log-Gabor can be represented as:

$$\boldsymbol{p}_o(x, y, s) = \boldsymbol{p}_o^{even}(x, y, s) + i\cdot\boldsymbol{p}_o^{odd}(x, y, s) \tag{6}$$

where $\boldsymbol{p}_o^{even}(x, y, s)$ represents the even-symmetric filter of the 2D-Log-Gabor filter; $\boldsymbol{p}_o^{odd}(x, y, s)$ represents the odd-symmetric filter; i represents the imaginary unit of the retest; o is the orientation of 2D-log-Gabor.

The 2D-Log-Gabor filter is decomposed into two parts in the spatial domain: even- and odd-symmetric filters, which are more robust against the noise and the grayscale differences. The equation constructed is shown in (7):

$$\left[\widetilde{\boldsymbol{EO}}_o(x, y), \widetilde{\boldsymbol{OO}}_o(x, y)\right] = \left[\sum_{s=1}^{3} \boldsymbol{I}(x, y) \otimes \boldsymbol{p}_o^{even}(x, y, s), \sum_{s=1}^{3} \boldsymbol{I}(x, y) \otimes \boldsymbol{p}_o^{odd}(x, y, s)\right] \tag{7}$$

where $\boldsymbol{I}(x, y)$ is the image; $\widetilde{\boldsymbol{EO}}_o(x, y)$ represents the response result of

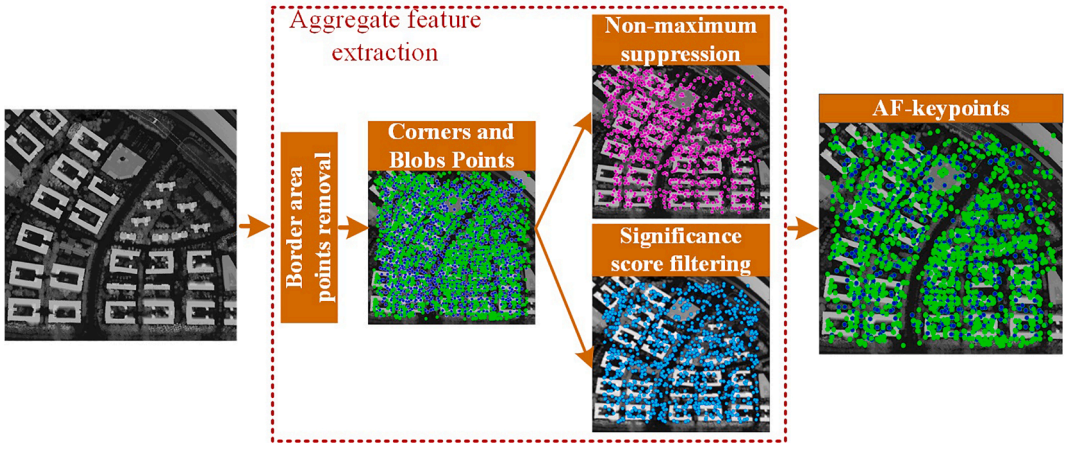


Fig. 3. The flow chart of aggregation feature extraction of key points.

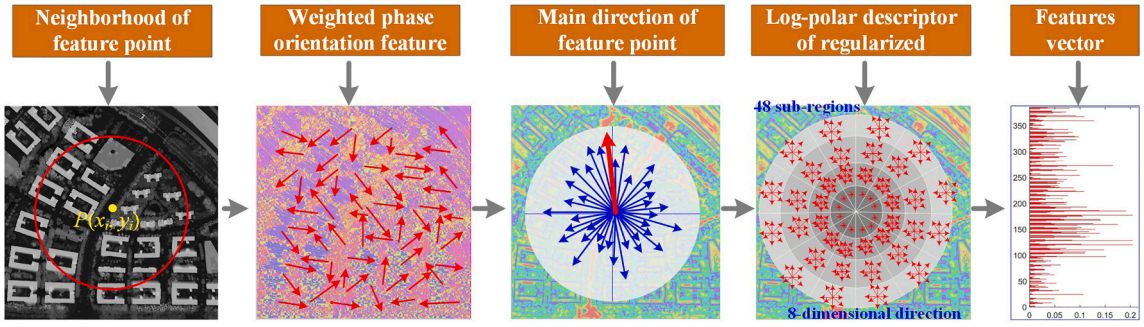


Fig. 4. The flow chart of the histogram of the orientation of robust phase descriptor.

the image on the real part filter; $\widetilde{\boldsymbol{OO}}_o(x,y)$ represents the response result of the image on the imaginary part filter; and $\otimes$ represents the convolution operation. The $\widetilde{\boldsymbol{EO}}_o(x,y)$ and $\widetilde{\boldsymbol{OO}}_o(x,y)$ are normalized to give $\boldsymbol{EO}_o(x,y)$ and $\boldsymbol{OO}_o(x,y)$ respectively.

Problems such as phase direction reversal and phase extreme value mutation usually cause changes in the shape component of the image, which destroys its structural features and increases the difficulty in identifying the descriptors. Log-Gabor's odd symmetric function rather than its even symmetric function filter has often been used to calculate the PC features. For calculating the shape component of the image, the 2D-Log-Gabor filter used a bandwidth function to calculate the shape component ratio between different pixels, which has played a key role in the computation of phase-oriented features. Based on the above-mentioned findings, a weighted bandwidth function model of the 2D-Log-Gabor filter is designed as:

$$\begin{cases} wc' = \max\limits_{o=1}^{6}\left(\dfrac{1}{1+\exp(g\cdot(Cutoff-\boldsymbol{width}_o(x,y)))}\right)-\xi \\ wc'' = \min\limits_{o=1}^{6}\left(\dfrac{1}{1+\exp(g\cdot(Cutoff-\boldsymbol{width}_o(x,y)))}\right)-\xi \end{cases} \tag{8}$$

where wc' and wc'' represent the maximum and minimum weighting coefficients, respectively; exp is an exponential operator;*Cutoff* is the fractional measure of frequency spread; $\boldsymbol{width}_o(x,y)$ is the fractional of image frequency; g controls the sharpness of the transition in the PC model; and ξ is a minimum value excluding zero.

Eq. (8) calculates the maximum and minimum bandwidths of the phase energy components in each direction and scale and then adds the results into 2D-Log-Gabor's odd-symmetry and even-symmetry functions. This operation can effectively overcome the negative effects caused by phase extreme value mutation. Accordingly, the weighted phase orientation feature equation is defined, as shown in Eq. (9):

$$\boldsymbol{WO}=\arctan\left(\frac{\sum_{o=1}^{6}\left(wc'\cdot\sin\left(\frac{\pi}{6}\cdot o\right)\cdot\boldsymbol{OO}_o(x,y)+wc''\cdot\sin\left(\frac{\pi}{6}\cdot o\right)\cdot\boldsymbol{EO}_o(x,y)\right)}{-\sum_{o=1}^{6}\left(wc'\cdot\cos\left(\frac{\pi}{6}\cdot o\right)\cdot\boldsymbol{OO}_o(x,y)+wc''\cdot\cos\left(\frac{\pi}{6}\cdot o\right)\cdot\boldsymbol{EO}_o(x,y)\right)+\Phi}\right) \tag{9}$$

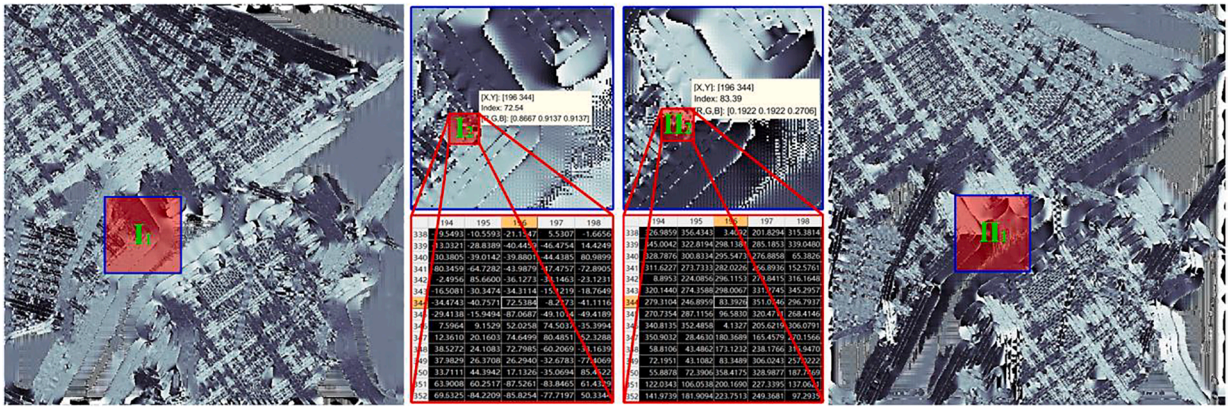
where $\boldsymbol{WO}$ represents the weighted phase orientation feature, $\boldsymbol{OO}_o(x,y)$ represents the odd-symmetric convolution results after normalization at direction o-th layer, $\boldsymbol{EO}_o(x,y)$ represents the even-symmetric convolution results after normalization at direction o-th layer, $\frac{\pi}{6}\cdot o$ represents the angle of rotation, and Φ is a minimal value (Φ=0.0001) excluding zero.

Finally, to eliminate the forward angle reversal caused by convolution values in multiple directions, we transformed the orientation angle between [0°, 360°] and used $\boldsymbol{W}$ to denote the final weighted phase orientation feature. The formula used here is shown in equation (10):

$$\boldsymbol{WP}=\begin{cases}\boldsymbol{WO}+\Phi, & \boldsymbol{WO}>0\\ \boldsymbol{WO}+\pi, & \boldsymbol{WO}<0\end{cases},\quad \boldsymbol{W}=\boldsymbol{WP}\otimes\frac{\omega}{\pi} \tag{10}$$

where $\boldsymbol{W}$ represents the final weighted phase orientation feature, $\boldsymbol{WP}$ represents the radian value of the phase orientation feature, ω represents a non-negative constant term (ω=360), and Φ is a minimum value (Φ=0.0001) excluding zero.

Fig. 5(a) demonstrates the orientation features calculated directly by the PC model; Fig. 5(b) describes our proposed weighted phase orientation features. The global comparison shows that the weighted orientation features are more recognizable than the phase orientation features, for they better preserve the shape components of the image and are far less vulnerable to the phase extreme value mutations. Meanwhile, observation of local areas (I and II) reveals that the problem of the directional inversion of the phase-oriented features is also well resolved



(a) Phase orientation features; (b) weighted phase orientation feature

Fig. 5. Comparison results of phase and weighted phase orientation features.

by the weighted phase orientation features, making the calculation of the principal orientation of the feature more accurate.

The principal orientation of the feature point needs to be calculated to guarantee the rotation invariance of HOWP. Thus, we have selected a fixed circular neighborhood centered on the feature points and identified the principal direction of each feature point based on the phase-amplitude feature and the weighted phase orientation feature. First, the histogram was evenly divided into 36 equal parts at intervals of 10°. Then, the phase-amplitude features and weighted phase orientation features of each equal part are counted, and the peaks with amplitudes higher than 80% of the highest peak were assigned as the principal orientation of the feature point.

The current remote sensing image matching study found that the rational polynomial coefficient or POS data can be directly used for rough geographical registration (Noh and Howat, 2018) and image rotation elimination. In this respect, we also proposed an improved Simplified-HOWP with no main orientation estimation for generating patches.

3.2.2. *Regularized log-polar descriptor*

To further improve the robustness of our descriptor, a regularized grid-optimized log-polar descriptor framework is designed for computing feature vectors. The log-polar description method using the gradient location and orientation histogram (GLOH) algorithm has the advantage of being quite stable (Mikolajczyk and Schmid, 2005). However, the GLOH algorithm is composed of 17 irregularly divided sub-regions, which, we found, negatively influences the robustness of descriptors in MRSI matching. Therefore, we put forward a framework of the regularized log-polar descriptor that constrains the GLOH for better performance.

The area of each sub-region was approximately the same. The horizontal direction of each grid represented the polar angle of the pixel location in the circular neighborhood, which was divided into 12 fan-shaped intervals by 30°. Therefore, the pixels of each sub-region had a gradient amplitude and direction histogram of eight dimensions. Finally, the number of log-polar sub-region grids (48) and the orientation histogram (eight dimensions) were multiplied to generate a new

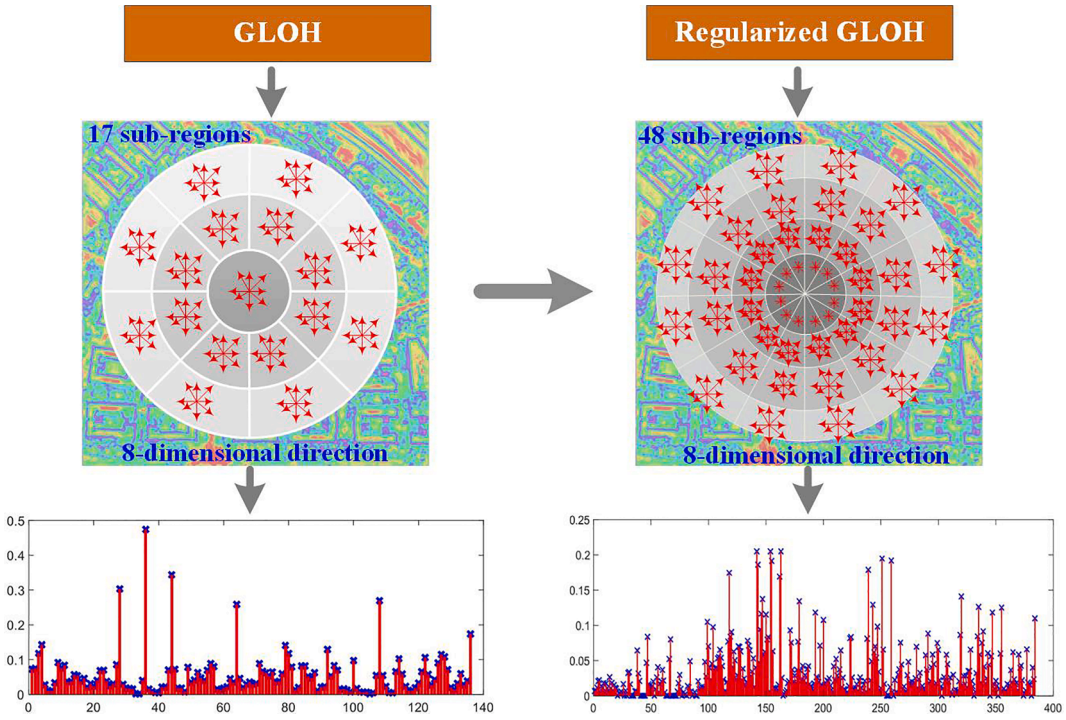


Fig. 6. Extended GLOH descriptor flowchart.

384-dimensional log-polar descriptor vector (Fig. 6). The vector of the descriptor feature is as equation (11):

$$RGLOH = [D_1, D_2, \cdots, D_N]^T \tag{11}$$

where *RGLOH* represents the descriptor set of all feature points, D_i^T the descriptor vector of a feature point, T the matrix transposition character, and N the number of feature points. The components of the feature vector of the 384-dimensional descriptor are expressed as $D_i^T = [V_1, V_2, V_3, \cdots, V_{384}]$, and the dimension of each descriptor can be expressed as $(4*n*d)$, where n represents the number of grids divided by each circular neighborhood, and d represents the orientation dimension of each feature point.

3.3. Bidirectional matching

Feature matching follows the calculation of the HOWP descriptor. Here, we used the similarity measurement of Euclidean distance and the matching strategy of the bidirectional matching to ensure the one-to-one correspondence of the obtained matching points. To remove outliers generated after the bidirectional matching, we used the fast sample consensus (FSC) algorithm (Wu et al., 2014) to cast out wrong matchings. The FSC algorithm can steadily extract the correct matching point pairs from mismatches with fewer iterations.

4. Experimental results

Four state-of-the-art methods, i.e., PSO-SIFT (Ma et al., 2016), LGHD (Aguilera et al., 2015), RIFT (Li et al., 2020), and HAPCG (Yao et al., 2021), were used for comparison in which we set the image scale differences of 1.6. The radius of the neighborhood window (N_W) is set to 38 pixels. The parameters for comparison were adjusted to the optimal stage accordingly and the proposed HOWP method. The PSO-SIFT, LGHD, RIFT, and HAPCG were implemented in Matlab-R2018a, and the number of matched key points was set to 3500. The experiments were performed on a Lenovo-R9000K with an AMD Ryzen 9 5900HX CPU, 3.30 GHz, 32 GB-RAM, and Windows 11 x64 operating system. The image-space affine transformation was used to model the geometric relationships of image pairs. Over 15 well-distributed ground truth points were manually collected for each pair to calculate the affine transformation as the ground truth, which was used to measure the location accuracy of the automatically matched points. The distribution of some ground truth points is shown in Fig. 7.

Four indices, i.e., the number of correct matches (NCM) in the MRSIs. The success rate (SR) was calculated with the following definition of success matching: (1) the NCM should be sufficient to obtain a solution of the selected geometric transformation model and have at least one redundant observation. The definition of SR is as follows:

$$I(p_i) = \begin{cases} 1, & NCM(p_i) \geqslant N_{\min} \\ 0, & else \end{cases}, SR = \frac{1}{M}\sum_i I(p_i)\cdot 100\% \tag{12}$$

In Eq. (10), $I(p_i)$ represents a logical value, 1 represents a successful matching trial and 0 represents a failed matching trial. The $N_{\min}$ represents a minimum number of correct matching points ($N_{\min}$ was set as 4). *SR* represents the matching success rate; M represents the total number of image pairs of multi-modal image sets.

Rate of correct matches (RCM), and root of mean-squared location error (RMSE), were used to examine the matching results. The RCM is defined as Eq. (13):

$$RCM = \frac{N_C}{N_C + N_F} \tag{13}$$

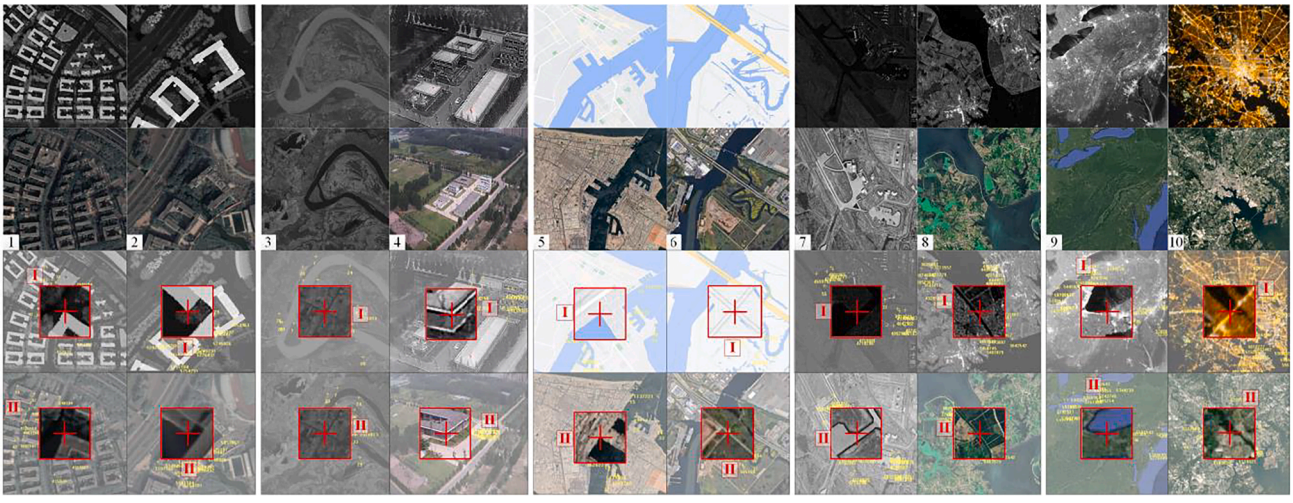
The RMSE of the correct matches is defined as Eq. (14):

$$RMSE = \sqrt{\frac{1}{N}\left(\sum_{i=1}^{N}[(x_i - x_i'') + (y_i - y_i'')]\right)} \tag{14}$$

where N represents the number of ground truth points, and (x''_i, y''_i) is the coordinate of the i-th ground truth point (x'_i, y'_i) converted by correspondence matching.

4.1. Image datasets

In this paper, 50 sets of MRSIs, including depth-optical, infrared-optical, map-optical, synthetic aperture radar (SAR)-optical, and night-day image, were used for demonstration. The image pairs not only have the NRD and the contrast differences but also have geometric transformations such as scale-change, rotation, and displacement (Fig. 7). Specifically, depth images were obtained from airborne LiDAR data, infrared images were from the Landsat TM-5 or the airborne infrared sensors, and maps were from Google Maps. The SAR data were acquired with the GF 3 satellite. The night data were acquired with national aeronautics' Suomi, national polar-orbiting partnership satellite and space administration, and national oceanic and atmospheric administration' satellite. The image sizes range from 381 and 750 pixels.



(a) Depth-optical and ground truth points; (b) infrared-optical and ground truth points; (c) map-optical and ground truth points; (d) SAR-optical and ground truth points; and (e) night-day and ground truth points.

Fig. 7. Part of multi-modal remote sensing images (MRSIs).

4.2. Results and analysis

To verify the robustness and the matching accuracy of the proposed HOWP method, we compared it with the four popular methods PSO-SIFT, LGHD, RIFT, and HAPCG. Fig. 8 shows the results of the quantitative evaluation; the NCMs, the RCMs, and the RMSEs of the proposed method and the four compared methods are presented in this figure. Table 1 shows the average results on four metrics. In Fig. 8(c), the symbol $+\infty$ represents the pairs that failed to match, or when the RMSE was > 9 pixels.

The results of the PSO-SIFT method are represented by orange dotted lines in Fig. 8. The PSO-SIFT method successfully matched 28 out of 50 image pairs and has the worst SR result of only 56.0% (Table 1), with a mean NCM of 37.7 and a mean RCM of 9.86%. Nonetheless, it boasts the advantage that it is favorably invariant to MRSI scaling and rotation invariance, manifest in its average RMSE of 4.53 pixels. Since the core of the PSO-SIFT method is to design a second-order Sobel operator to calculate the image gradient features, it is still sensitive to NRD despite its robustness to the illumination difference.

The light blue dotted lines represent the results of the LGHD method, which uses image frequency domain features to perform matching. In our experiments, it successfully matched 27 out of 50 image pairs, with an SR of 54.0%. However, this method cannot effectively address the scale and rotation differences manifested in its average NCM of 24.54, average RCM of 8.66%, and average RMSE of 4.38 pixels. Meanwhile, it takes more time to calculate frequency domain features. All these factors make it unable to achieve MRSI matching effectively and efficiently.

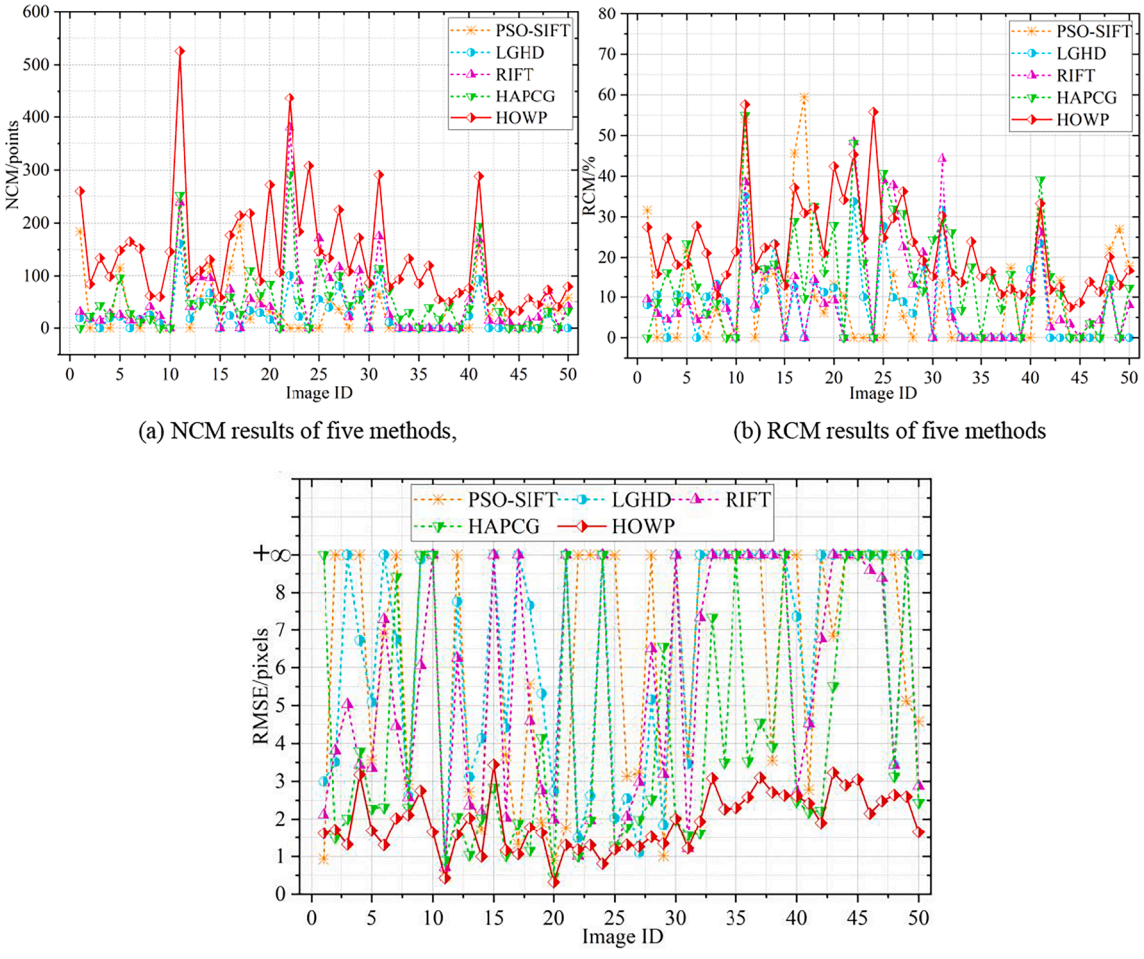
Table 1
Results of five methods with respect to four evaluation indicators.

	PSO-SIFT	LGHD	RIFT	HAPCG	HOWP
SR	56.0%	54.0%	70.0%	78.0%	**100%**
NCM	37.7	24.54	50.62	51.56	**138.72**
RCM	9.86%	8.66%	10.78%	15.53%	**22.32%**
RMSE	4.53	4.38	4.13	2.89	**1.93**

The results of the RIFT method are marked in purple. This method also used image frequency domain features to achieve the matching as LGHD did. The difference is that a robust maximum index map descriptor substantially optimized its matching performance, as its SR increased to 70.0% with an NCM of 50.62, an RCM of 10.78%, and an RMSE of 4.13 pixels.

Both Fig. 8 (b) and (c) show that the RIFT method is characterized by a lower value of RCM with great fluctuations in its RMSE results. The reason might be its incapability to handle scale differences in images. The HAPCG method (presented in green) has better results with the SR of 78.0%, average NCM of 51.56, average RCM of 15.53%, and average RMSE of 2.89 pixels. However, HAPCG can only be utilized to match the images with small rotation differences (Yao et al., 2021).

The proposed HOWP (marked in solid red lines) yielded the most robust matching results. It successfully matched all the 50 MRSI pairs. Its average NCM is 138.72, with it being the only method to obtain >50 correct matches. Its average RCM is 22.32%—the best among the five methods, and the average RMSE is 1.93 pixels. It is the only method with



(a) NCM results of five methods,

(b) RCM results of five methods

(c) RMSE results of five methods

Fig. 8. Results of several indicators.

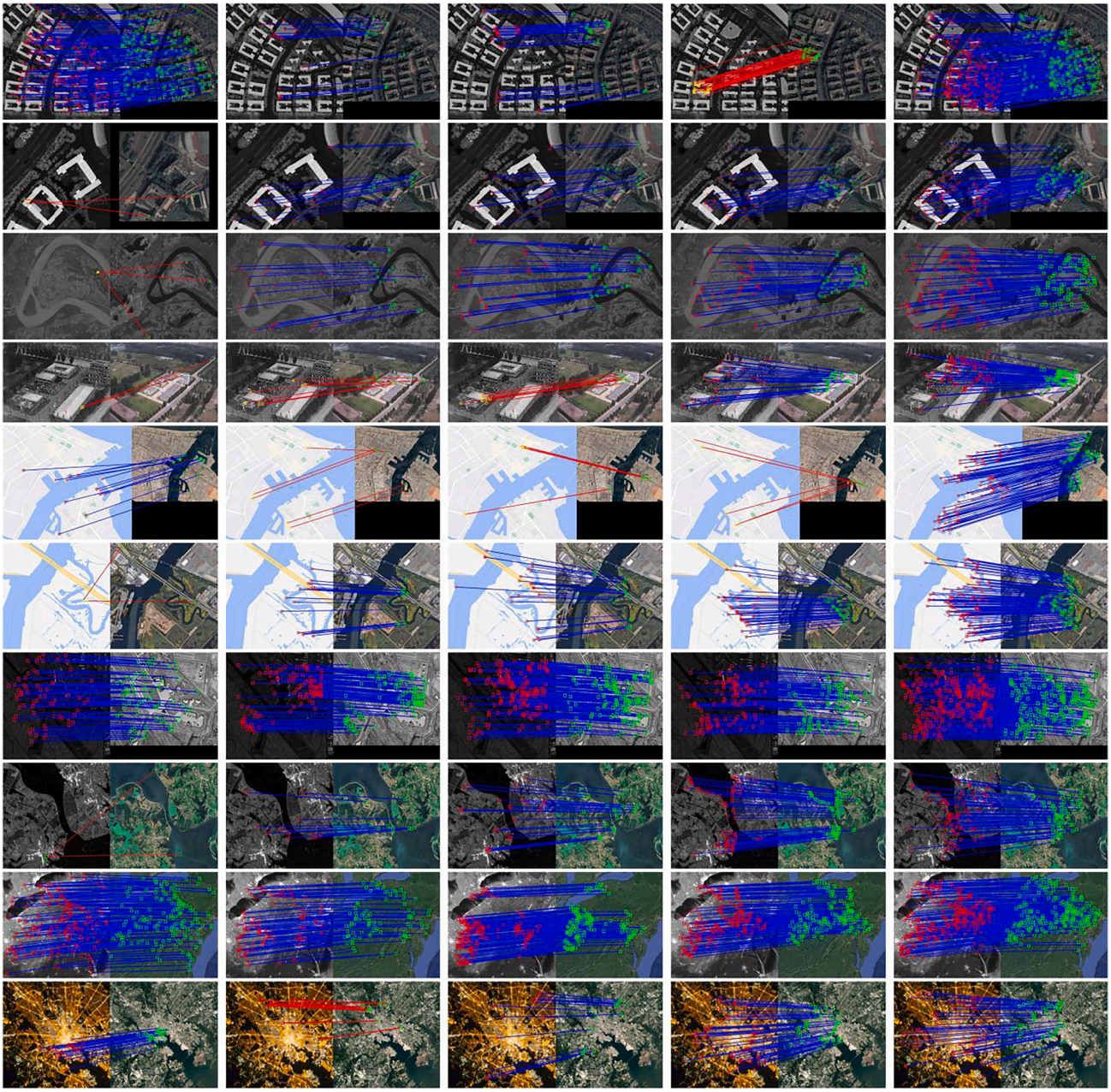
100% success matching rate that also obtained <2-pixel RMSE.

Fig. 9 shows the matched corresponding points of the five methods. Fig. 9 (a) presents the results of the PSO-SIFT method, which has fewer corresponding points in images with significant contrast and illumination differences. Fig. 9 (b) and (c) provide information about LGHD and RIFT, respectively. The latter method features high stability in the matching due to its descriptor of the maximum index map. Besides, they all used the PC model in the feature description, but their ability to deal with scale and rotation differences is strongly limited. The HAPCG method, as shown in Fig. 9 (d), performs better in coping with scale differences and small rotation differences, but it is still not an optimal solution when it comes to large rotation differences. The proposed HOWP achieves the most robust matching of images with NRD, optical differences, and contrast differences, as demonstrated in Fig. 9 (e). It also shows its great potential in tackling scale, rotational, and displacement differences.

The matching results of another 40 images are displayed in Fig. 10, demonstrating the matching performance of the HOWP method. As shown in Fig. 10, the proposed HOWP method shows good stability against MRSI matching and achieves rich NCM in scale, rotation, and translation differences. Therefore, the proposed HOWP method provides a useful reference for further MRSI matching challenges.

The Simplified-HOWP method can improve HOWP's matching performance in MRSI matching when images only have translational differences or rotation angles <5°. Ten sets of MRSI pairs with only translation differences were selected for the experiment, and the results are shown in Fig. 11.

Fig. 11 shows that the proposed Simplified-HOWP method showed outstanding performance when the main orientation of the features were not considered. Fig. 12 shows that its NCM has increased by at least 0.12



(a) PSO-SIFT; (b) LGHD; (c) RIFT; (d) HAPCG; (e) HOWP

Fig. 9. Matching results from the five compared methods.

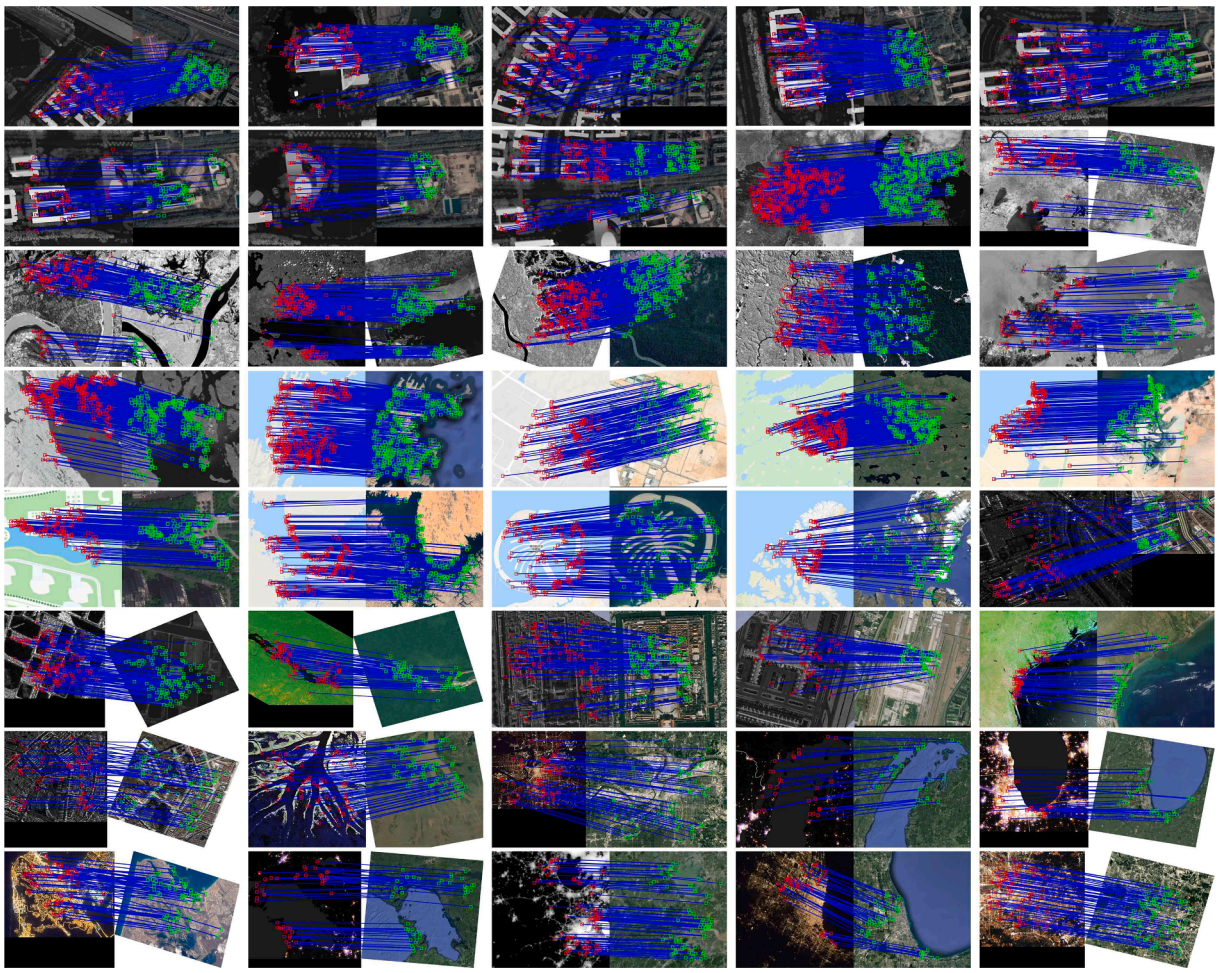


Fig. 10. Matching results of the HOWP method.

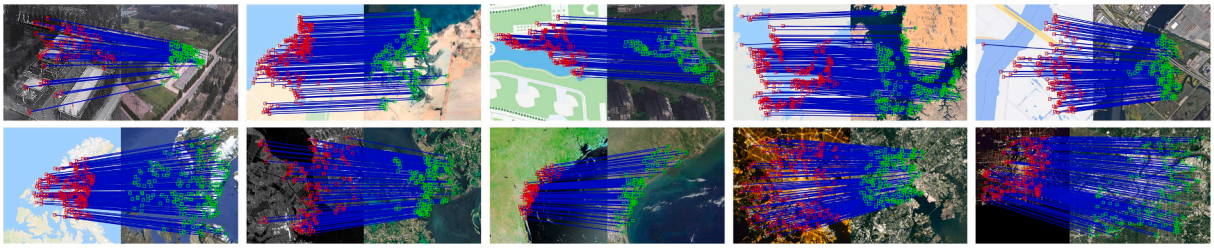


Fig. 11. Matching results of the Simplified-HOWP method.

times, its NCM has increased by at most 1.5 times and the average performance by 0.75 times. At the same time, Simplified-HOWP method' RCM has increased by at least 0.1%, its RCM has increased by at most 19.1% and the average performance by 6.9%. Therefore, the Simplified-HOWP method performs more robustly in matching the MRSIs with only translational differences.

4.2.1. Analysis of scale invariance

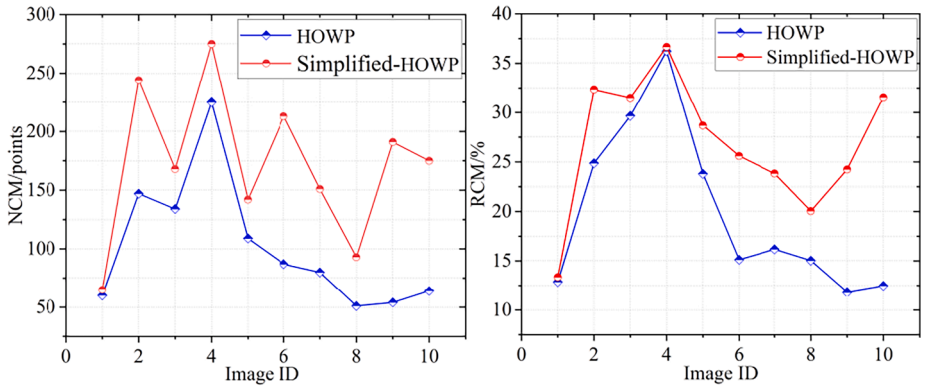
To verify the scale invariance of the HOWP method, a set of map-optical image pairs was used for the matching test. First, the reference images were simulated and computed at 0.2 times image scale intervals to generate 10 sets of simulated images (0.4–2.4 times). The matching results are shown in Fig. 13.

Although the number of corresponding points matched by the HOWP method gradually decreased as the scale differences increased (Fig. 13), the method could still secure the successful matching even when the scales were 0.4–2.4 times different. NCMs we obtained exceeded 20 for the two cases of 0.4 and 2.4 scale difference, which were sufficient for the registration. Therefore, the HOWP method could effectively boost the scale invariance of the MRSI matching.

4.2.2. Analysis of rotation invariance

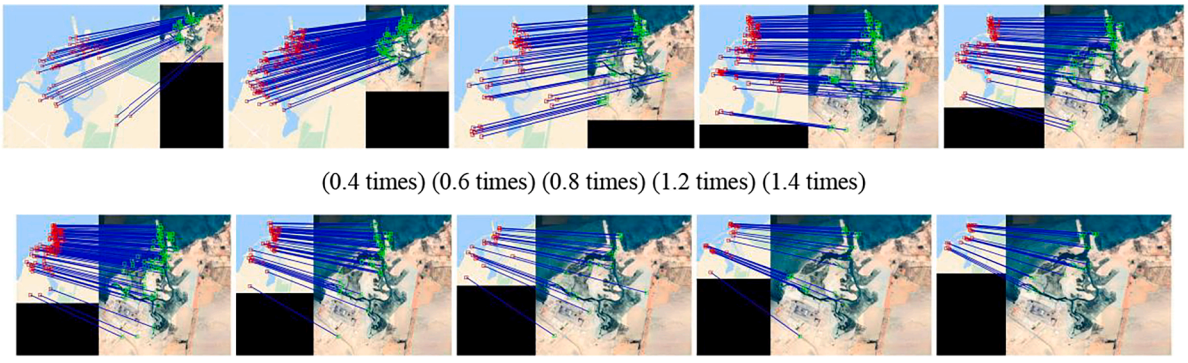
To verify the rotation invariance of the HOWP method, a set of map-optical image pairs was used for the matching tests. First, the reference image was rotated clockwise and counterclockwise at 10-degree intervals to generate five simulated images in both directions. Then, 10 sets of simulated image pairs were obtained for the matching test (Fig. 14).

Although the number of valid corresponding points matched by the HOWP method decreased as the angle increased, the method still secured the successful matching whether the rotation difference was set at −50, 50 between two degrees. The NCMs obtained when the rotation



(a) NCM results of two methods, (b) RCM results of two methods

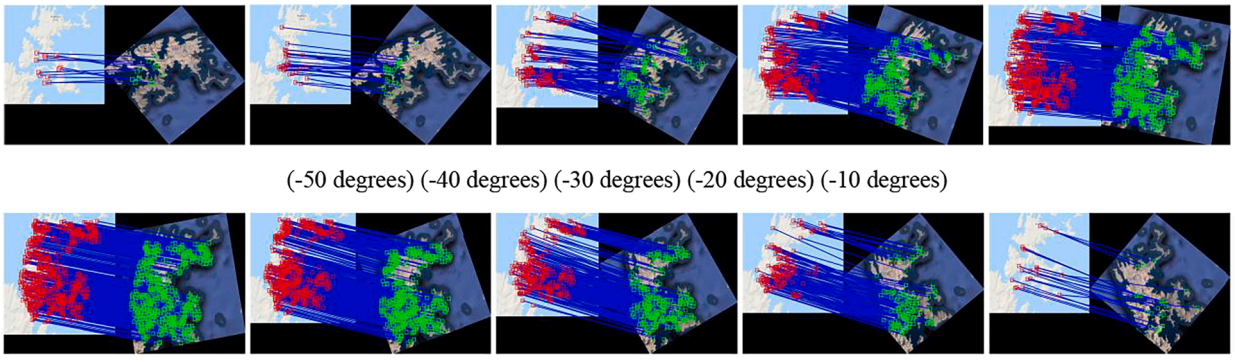
Fig. 12. Quantitative comparison results of HOWP and Simplified-HOWP methods.



(0.4 times) (0.6 times) (0.8 times) (1.2 times) (1.4 times)

(1.6 times) (1.8 times) (2.0 times) (2.2 times) (2.4 times)

Fig. 13. Matching results under different scale differences of the HOWP method.



(-50 degrees) (-40 degrees) (-30 degrees) (-20 degrees) (-10 degrees)

(10 degrees) (20 degrees) (30 degrees) (40 degrees) (50 degrees)

Fig. 14. Matching results under different rotation differences of the HOWP method.

difference was −50 and 50°were all >10 point pairs, which meets the calculation requirements of the transformation model (three-point pairs are required for affine model calculation and four for perspective transformation model). Therefore, the HOWP method could effectively enhance the rotational invariance of the MRSI matching. Compared with traditional methods, the proposed HOWP displays the most robust performance in MRSI matching with reference to its ability to handle the NRD, contrast difference, and scale and rotation invariance.

5. Discussion

To validate the robustness of the HOWP method, we first carried out an in-depth analysis of its performance under different parameter settings and then elaborated on its three main contributions, namely, aggregated feature strategy, weighted phase histogram, and regularized log-polar description, intending to fully demonstrate the innovation accomplished by the proposed HOWP method.

5.1. Parameter settings

The proposed HOWP method has parameters, such as feature neighborhood window (*Nw*) and repeated feature point filtering window (*Fw*), whose different values will affect its matching performance. Therefore, we quantitatively analyzed how the HOWP method functioned under different settings. More details of parameter settings are given in Table 2.

We tested 30 sets of MRSIs according to the parameter settings given in Table 2 to evaluate the impact of different parameters on the HOWP method by observing the mean NCM and mean RCM of images, which are shown in Fig. 14.

Fig. 15 shows how different settings of the two parameters affect the HOWP method. The NCMs progressively decreased when $N_W < 30$ *or* $N_W > 46$ (Fig. 15 (a)). When $N_W = 38$, NCM could obtain the optimal result. The RCM increased as N_W increased; however, when $N_W > 46$, the increase in RCM was no longer sufficient to compensate for the loss brought by the decline of NCM. Therefore, the results would be optimal only when *NW* is set to 38. As shown in Fig. 15 (b), both NCM and RCM decreased as *Fw* increased. When $Fw \leq 3$, small fluctuations were obtained in the NCM and RCM; however, when $Fw > 3$, the indices began to drop. Taken together, the optimal value for *Fw* should be 3.

5.2. Analysis of aggregation features

The effect analysis of the aggregation feature strategy was then conducted where the control variate method was used. In this case, we changed only the feature point extraction method. Two indices (NCM and RCM) were used to evaluate their performance. The results are shown in Figs. 16 and 17.

Fig. 16 shows that the matching results obtained with the aggregation feature strategy are better than those obtained with either the blob or the corner method. Notably, the NCM of the aggregation feature is 107, higher than that of blobs (61) and corners (33), showing the significant improvement of the proposed method after the optimization by the aggregation feature strategy.

Fig. 17 shows that Blob-HOWP method has a poor perform. The Corner-HOWP method performs better than that of the Blob-HOWP method, but the matching results of the Corner-HOWP method are not as significant. Average NCMs and average RCMs of the Blob-HOWP method were 75.38 points and 18.88%, while those of Corner-HOWP were 108.12 points and 21.77% respectively. By comparison, average NCMs of the proposed AF-HOWP, 1.84 times higher than that of the Blob-HOWP method, 1.28 times higher than that of the Corner-HOWP method. Its average RCMs was 3.44% higher than that of the Blob-HOWP method and 0.54% higher than that of the Corner-HOWP method. This is mainly because the proposed aggregation features method could combine the advantages of Blob and Corner features, therefore the acquired feature points are more robust.

5.3. Analysis of different orientation features

The weighted phase orientation histogram is key to solving phase direction inversion and phase extreme value mutation. Nonetheless, we used the control variate method to analyze the significance of the histogram. The proposed HOWP method is used to compare with the initial PC methods and the gradient orientation features methods. Two indices (NCM and RCM) were used to evaluate their performance. The comparison results are shown in Figs. 18 and 19.

The HOWP method has obtained a better NCM result than the initial PC method and gradient orientation feature method. Specifically, the gradient orientation feature method obtained 378 matching pairs, with 8 correct pairs and 370 incorrect pairs, the initial PC method obtained 303 matching pairs, with 43 correct pairs and 260 incorrect pairs, while the HOWP method obtained 321 matching pairs, with 107 correct pairs and 214 incorrect pairs.

Fig. 19 shows that the HOWP (Gradient)'s average NCM was 73 with the average RCM of 13.2%. The HOWP(PC)'s average NCM was 114.18 with the average RCM of 18.47%. Average NCMs of the proposed HOWP (weighted phase orientation, WPO) was 21.5% higher than that of the initial PC method and 90.03% higher than that of the gradient orientation feature method in 50 group pairs. Average RCMs of the HOWP (WPO) was 3.85% higher than that of the initial PC method and 9.12% higher than the gradient orientation feature method. Obviously, the HOWP method has better matching performance. The reasons why HOWP (Gradient) method performed so poorly might be: the gradient of the image was more sensitive to MRSI and the feature information description was insufficient. The HOWP (PC) method is mainly affected by the directional inversion or phase extreme value mutations.

Table 2
Parameter settings of the proposed HOWP Method.

Experiment	Variable	Fixed Parameters
Parameter N_W	N_W = [22,30,32,34,36,38,40,42,44,46,54,62,70,78]	$F_W = 3$
Parameter F_W	F_W = [1,2,3,4,5,6,7,8,9]	$N_W = 38$

5.4. Analysis of regularized log-polar descriptor

We also utilized a regularized log-polar description calculation method characterized by a more finely divided neighborhood division. This method is quite useful in computing descriptors of features that are more abundant. Its performance was tested by comparison with the original log-polar description through the control variate method. Two indices (NCM and RCM) were used for evaluation. The results are shown in Fig. 20.

Evidently, the proposed Regularized-GLOH method significantly outperforms the original GLOH method in both NCM results and RCM results in 50 sets of images. As Fig. 20 (a) and (b) show, the average NCMs and average RCMs of the original GLOH were 95.64 points and 16.03% respectively. The NCM of HOWP (RGLOH) was 0.45 times higher than that of HOWP (GLOH), and its' RCM was 6.3% higher than that of HOWP (GLOH). This is mainly because our method could obtain more effective information, and the statistics of features in the neighborhood of feature points are more uniform and regular.

6. Conclusions

In this paper, we have proposed a novel robust matching method, termed "HOWP," to effectively address the difficulties in MRSI matching, such as the directional inversion of phase features and phase extreme value mutation, which the PC model cannot overcome. The proposed method guarantees the stably robust matching of MRSIs by using the moments of the PC model to extract blob points and corner points separately and enriching the phase congruence orientation features. Using 50 sets of MRSIs, comparisons were also made with PSO-SIFT, LGHD, RIFT, and HAPCG methods with respect to illumination, contrast map, grayscale, and rotation differences. The experimental results have the following indications:

(1) The aggregation feature optimization strategy proposed by the HOWP method can fully extract the corners and blobs features of the image. The SR obtained using the HOWP method would be increased at least by 22.0% compared with that obtained using the four state-of-the-art methods.
(2) The proposed HOWP method has better performance in matching MRSIs with scale, rotational, and translational differences, for its NCM is 4.5 times higher than that of the LGHD method, 2.58

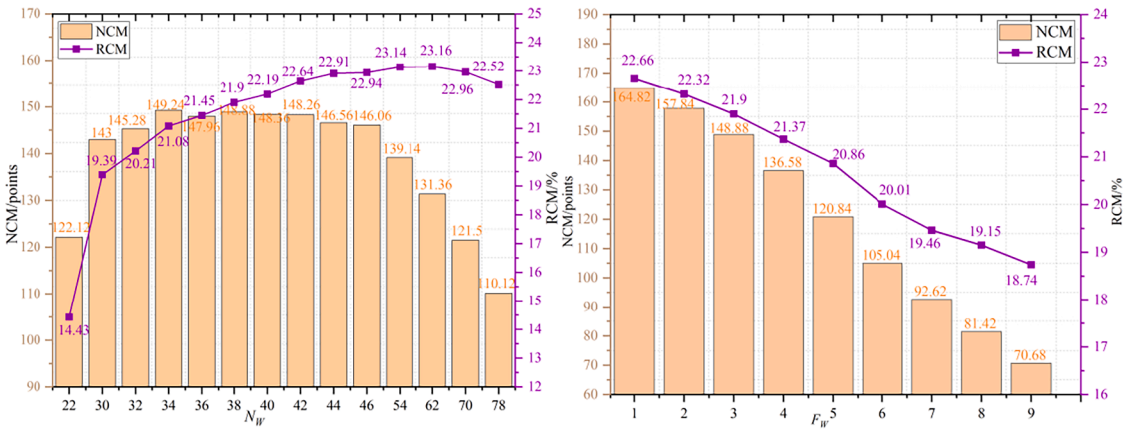


Fig. 15. Evaluation results of two indicators.

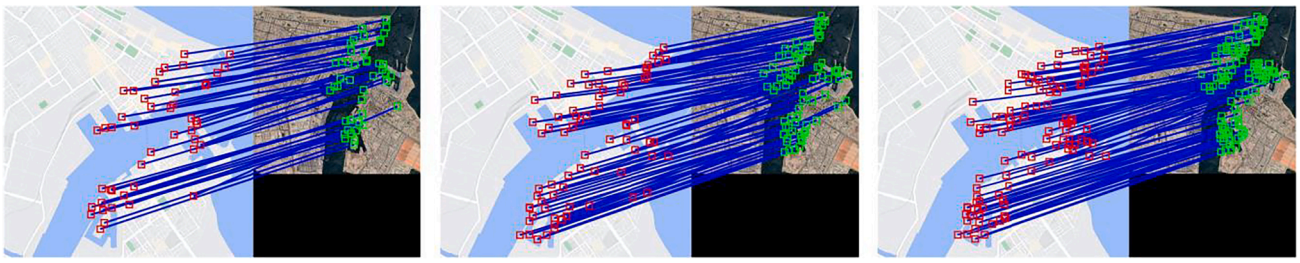


Fig. 16. Matching results of different feature point detection methods.

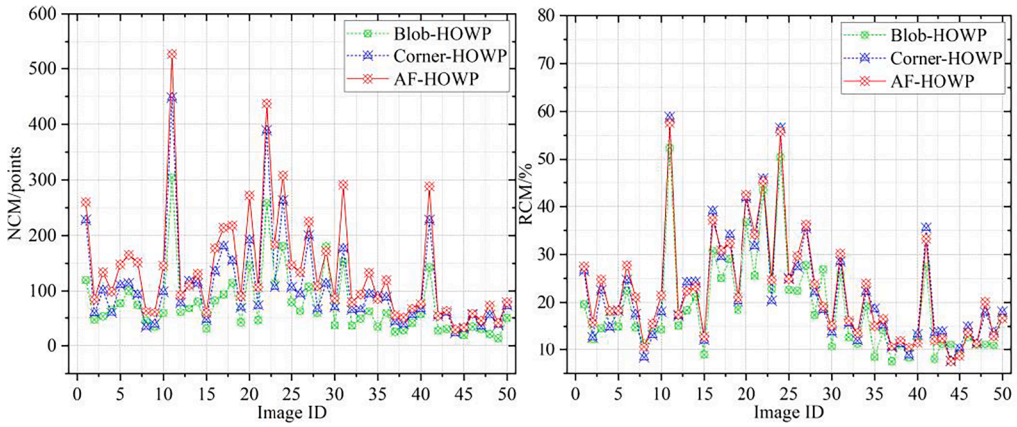


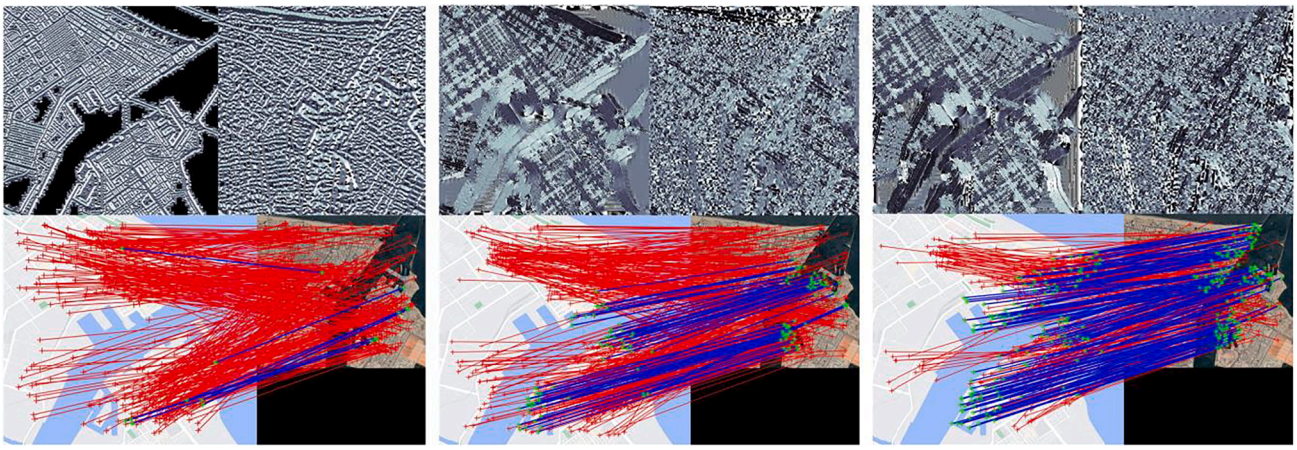
Fig. 17. Comparison results of three feature extraction methods.

times than the PSO-SIFT, 1.66 times than the RIFT method, and 1.62 times than the HAPCG method.

(3) An extended version of the HOWP method, namely, Simplified-HOWP, can perform more robustly in the matching of MRSIs without any rotational differences, for its NCM is 0.75 times higher than the HOWP method, and its RCM is 6.92% higher than that of the HOWP method.

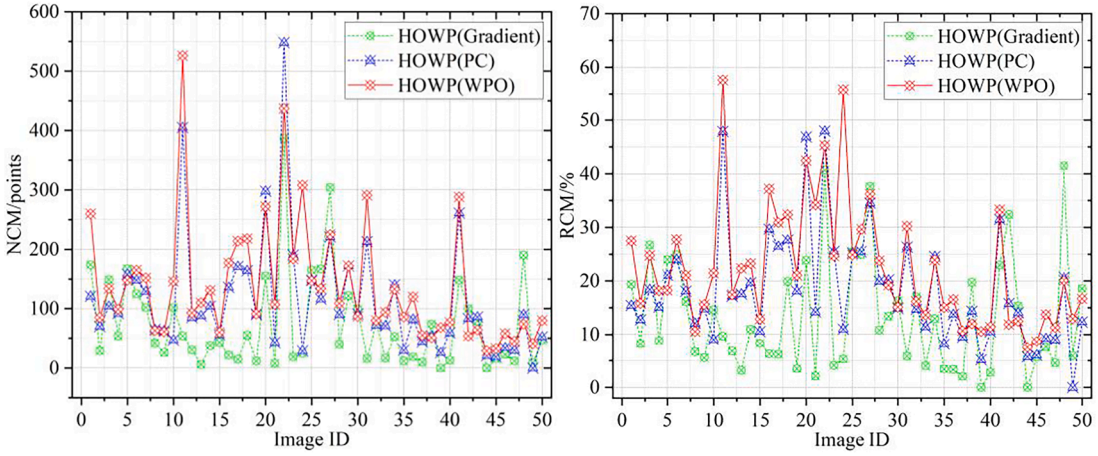
In summary, the proposed HOWP method enables robust and stable MRSI matching, thereby providing regular and reliable data for multi-source image data fusion, multi-target recognition, and SLAM. Additionally, its neighborhood window can be set more flexibly. The size of the neighborhood window can be increased when images have huge modal differences or narrowed when the difference is small. In general, a setting of 38 pixels is sufficient in most cases.

However, the proposed HOWP method still has some limitations. (1) the HOWP method is mainly suitable for the MRSI matching task with rigid transformation. Its performance might be undermined when it comes to the non-rigid image; (2) we executed the HOWP method mainly in a CPU-based environment without migrating it into a GPU computing environment, so it could not achieve real-time matching. Improving the computational efficiency of the method will be the focus of our next stage of research.



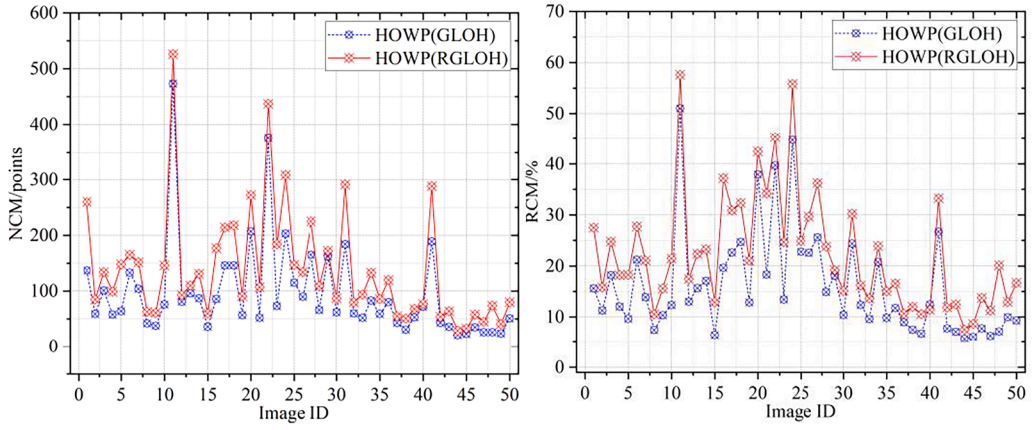
Matching results of (a) gradient orientation features, (b) phase orientation features and (c) our weighted phase orientation feature

Fig. 18. Comparison of matching results of three different orientation features.



(a) NCM of three different orientation features, (b) RCM of three different orientation features

Fig. 19. Matching result statistics of three different orientation features.



(a) NCM results of original GLOH and Regularized-GLOH, (b) RCM results of original GLOH and Regularized-GLOH

Fig. 20. Comparison of matching results original GLOH and Regularized-GLOH.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported in part by the Key Program of the National Natural Science Foundation of China under project No. 42030102, the Basic Research Strengthening Program of China under project No. 2020-JCJQ-ZD-015-00-04, the Fund for Innovative Research Groups of the Hubei Natural Science Foundation under project No. 2020CFA003, and the Major special projects of Guizhou[2022]001.

References

Aguilera, C., Sappa, A., Toledo, R., 2015. LGHD: a feature descriptor for matching across non-linear intensity variations. In: 2015 IEEE International Conference on Image Processing (ICIP). IEEE, pp. 178–181.

Alcantarilla, P., Bartoli, A., Davison, A., 2012. KAZE features. In: European conference on computer vision. Springer, Berlin, Heidelberg, pp. 214–227.

Bay, H., Ess, A., Tuytelaars, T., Van Gool, L., 2008. Speeded-up robust features (SURF). Comput. Vis. Image Underst. 110 (3), 346–359.

Chen, H., Arora, M., Varshney, P., 2003. Mutual information-based image registration for remote sensing data. Int. J. Remote Sens. 24 (18), 3701–3706.

Cole-Rhodes, A., Johnson, K., LeMoigne, J., Zavorin, I., 2003. Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient. IEEE Trans. Image Process. 12, 1495–1511.

Dusmanu, M., Rocco, I., Pajdla, T., et al., 2019. D2-net: A trainable cnn for joint description and detection of local features. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 8092–8101.

Efe, U., Ince, K., Alatan, A., 2021. DFM: a performance baseline for deep feature matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4284–4293.

Fan, Z., Liu, Y., Liu, Y., Zhang, L., Zhang, J., Sun, Y., Ai, H., 2022. 3MRS: an effective coarse-to-fine matching method for multimodal remote sensing imagery. Remote Sens. (Basel) 14 (3), 478–501.

Fischer, S., Šroubek, F., Perrinet, L., Redondo, R., Cristóbal, G., 2007. Self-invertible 2D log-Gabor wavelets. Int. J. Comput. Vis. 75 (2), 231–246.

Kovesi, P., 1999. Image features from phase congruency. Videre: J. Comput. Vis. Res. 1 (3), 1–26.

Li, J., Hu, Q., Ai, M., 2020. RIFT: multi-modal image matching based on radiation-variation insensitive feature transform. IEEE Trans. Image Process. 29, 3296–3310.

Li, J., Xu, W., Shi, P., Zhang, Y., Hu, Q., 2022. LNIFT: locally normalized image for rotation invariant multimodal feature matching. IEEE Trans. Geosci. Remote Sens. 60, 1–14.

Lowe, D., 1999. Object recognition from local scale-invariant features. In: Proceedings of the Seventh IEEE International Conference on Computer Vision. IEEE, no. 2, pp. 1150-1157.

Ma, W., Wen, Z., Wu, Y., Jiao, L., Gong, M., Zheng, Y., Liu, L., 2016. Remote sensing image registration with modified SIFT and enhanced feature matching. IEEE Geosci. Remote Sens. Lett. 14 (1), 3–7.

Mikolajczyk, K., Schmid, C., 2005. A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. 27(10), 1615–1630.

Noh, M., Howat, I., 2018. Automatic relative RPC image model bias compensation through hierarchical image matching for improving DEM quality. ISPRS J. Photogramm. Remote Sens. 136, 120–133.

Öfverstedt, J., Lindblad, J., Sladoje, N., 2019. Fast and robust symmetric image registration based on distances combining intensity and spatial information. IEEE Trans. Image Process. 28 (7), 3584–3597.

Rosten, E., Drummond, T., 2006. Machine learning for high-speed corner detection. In: European Conference on Computer Vision. Springer, Berlin, Heidelberg. pp. 430–443.

Rublee, E., Rabaud, V., Konolige, K., et al., 2011. ORB: an efficient alternative to SIFT or SURF. In: 2011 International conference on computer vision. IEEE, pp. 2564–2571.

Sedaghat, A., Mohammadi, N., 2018. Uniform competency-based local feature extraction for remote sensing images. ISPRS J. Photogramm. Remote Sens. 135, 142–157.

Sun, J., Shen, Z., Wang, Y., Bao, H., Zhou, X., 2021. LoFTR: Detector-free local feature matching with transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8922–8931.

Viola, P., Wells III, William, M., 1997. Alignment by maximization of mutual information. Int. J. Comput. Vis. 24 (2), 137–154.

Wang, Q., Zhang, J., Yang, K., Peng, K., Stiefelhagen, R., 2022. MatchFormer: Interleaving Attention in Transformers for Feature Matching. arXiv preprint arXiv: 2203.09645.

Wang, S., Quan, D., Liang, X., et al., 2018. A deep learning framework for remote sensing image registration. ISPRS J. Photogramm. Remote Sens. 145, 148–164.

Wiles, O., Ehrhardt, S., Zisserman, A., 2021. Co-attention for conditioned image matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15920–15929.

Wu, Y., Ma, W., Gong, M., 2014. A novel point-matching algorithm based on fast sample consensus for image registration. IEEE Geosci. Remote Sens. Lett. 12 (1), 43–47.

Xu, X., Li, X., Liu, X., Shen, H., Shi, Q., 2016. Multimodal registration of remotely sensed images based on Jeffrey's divergence. ISPRS J. Photogramm. Remote Sens. 122, 97–115.

Yao, Y., Zhang, Y., Wan, Y., 2021. Heterologous images matching considering anisotropic weighted moment and absolute phase orientation. Geom. Inform. Sci. Wuhan Univ. 46 (11), 1727–1736.

Yao, Y., Zhang, Y., Wan, Y., Liu, X., Yan, X., Li, J., 2022. Multi-modal remote sensing image matching considering co-occurrence filter. IEEE Trans. Image Process. 31, 2584–2597.

Ye, Y., Shan, J., Hao, S., Bruzzone, L., Qin, Y., 2018. A local phase based invariant feature for remote sensing image matching. ISPRS J. Photogramm. Remote Sens. 142, 205–221.

Ye, Y., Bruzzone, Lorenzo, Shan, J., et al., 2019. Fast and robust matching for multimodal remote sensing image registration. IEEE Trans. Geosci. Rem. Sens. 57 (11), 9059–9070.

Ye, Y., Wang, M., Hao, S., Zhu, Q., 2020. A Novel Keypoint Detector Combining Corners and Blobs for Remote Sensing Image Registration. IEEE Geosci. Remote Sens. Lett. 18 (3), 451–455.

Ye, Y., Tang, T., Zhu, B., Yang, C., Li, B., Hao, S., 2022. A multiscale framework with unsupervised learning for remote sensing image registration. IEEE Trans. Geosci. Remote Sens. 60, 1–15.

Zhao, C., Zhao, H., Lv, J., Sun, S., Li, B., 2016. Multimodal image matching based on multimodality robust line segment descriptor. Neurocomputing 177, 290–303.

Zhou, L., Ye, Y., Tang, T., Nan, K., Qin, Y., 2021. Robust matching for SAR and optical images using multiscale convolutional gradient features. IEEE Geosci. Remote Sens. Lett. 19, 1–5.