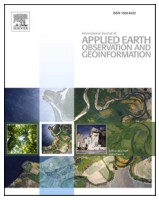
International Journal of Applied Earth Observation and Geoinformation 127 (2024) 103639

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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



Robust registration of multi-modal remote sensing images based on multi-dimensional oriented self-similarity features

Yongjun Zhang [a,b], Wenfei Zhang [a], Yongxiang Yao [a,b], Zhi Zheng [c,*], Yi Wan [a,b,*], Mingtao Xiong [a]

[a] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, HB 430079, China
[b] Technology Innovation Center for Collaborative Applications of Natural Resources Data in GBA, Ministry of Natural Resources, No.468 Huanshi East Road, Yuexiu District, Guangzhou, Guangdong Province 510075, China
[c] Department of Geography and Resource Management, The Chinese University of Hong Kong, Shatin N.T., Hong Kong, China

ARTICLE INFO

Keywords:
Multi-modal remote sensing image (MRSI)
Multi-dimensional oriented self-similarity features
Feature description
Coarse-to-fine strategy
Hybrid registration

ABSTRACT

Registration of multi-modal remote sensing images (MRSI) is crucial for unlocking the full potential of heterogeneous remote sensing imagery. However, achieving accurate registration among MRSI is challenging due to the trade-off between geometric invariance and matching accuracy, caused by differences in signal-to-noise ratio and nonlinear radiometric distortion (NRD) arising from varying imaging mechanisms. To tackle the challenge, this paper proposes a lightweight and hybrid feature-guided registration algorithm for MRSI called the hybrid registration algorithm based on multi-dimensional oriented self-similarity features (MOSS). MOSS leverages the advantages of multi-dimensional oriented self-similarity features to progressively enhance registration performance. In the hybrid feature coarse matching stage, oriented self-similarity features are extracted from MRSI, and their directional information is utilized for feature description to estimate the initial affine transformation. The fine matching under multi-dimensional oriented self-similarity features stage takes the outputs of the coarse matching stage to perform a template-like matching process. To evaluate the performance of MOSS, comprehensive experiments are conducted using six different combinations of MRSI, and seven state-of-the-art registration algorithms are selected for comparison. The experimental results demonstrate that MOSS outperforms the compared methods, with the number of correct matches being at least about 1.6 times higher than the comparison methods. Moreover, MOSS exhibits the lowest root mean square error across all experiments, with an average RMSE of 1.86 pixels, achieving an RMSE within 2 pixels. This highlights its effectiveness in achieving precise alignment and robust registration of MRSI.

1. Introduction

Geometrical registration is an image-processing technique that aligns different images of the same scene acquired at various times and viewing angles and with multiple sensors (Feng et al., 2021). Registration of multi-modal remote sensing images (MRSI) is crucial for achieving geometric alignment and consistent spatial reference across different sensors and modalities. It plays a vital role in applications such as map correction, precise positioning, feature extraction, target recognition, land surface change monitoring, Three-Dimensional (3D) reconstruction, and stereo vision (Zhang et al., 2021a; Zhang et al., 2021b). However, MRSI registration still faces challenges due to variations in imaging conditions, including sensor differences, modality differences, and geometric disparities. One of the challenges is the limited accuracy of feature matching due to significant differences in signal-to-noise ratios in MRSI. Feature extraction may fail for pixels with high signal-to-noise ratios, resulting in matching failures or misalignment. Another challenge lies in the template matching strategy commonly used for feature points correspondence, which requires pixel-level pre-registration of the images. This approach necessitates precise initial localization and often results in high computational complexity and poor resistance to geometric transformations.

To address these challenges, researchers have proposed many innovative methods to address signal-to-noise interference, nonlinear radiometric distortion (NRD), and geometric differences in MRSI. For example, SIFT (Lowe, 1999) and RIFT (Li et al., 2020) proposed feature-

* Corresponding authors at: School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, HB 430079, China (Y. Wan).
E-mail addresses: zhizheng@cuhk.edu.hk (Z. Zheng), yi.wan@whu.edu.cn (Y. Wan).

https://doi.org/10.1016/j.jag.2023.103639
Received 19 August 2023; Received in revised form 22 November 2023; Accepted 23 December 2023
Available online 5 January 2024

based methods focusing on geometric variations between MRSI. By employing flexible feature descriptors, these methods handle geometric transformations effectively. LoFTR (Sun et al., 2021) and based on Convolutional Neural Networks (CNN) proposed learning-based methods. While significant progresses have been made in enhancing the performance of MRSI registration, there is still room for improvement, particularly in balancing geometric invariance and achieving prompt registration.

This paper introduces a lightweight and hybrid feature-guided registration algorithm called the hybrid registration algorithm based on multi-dimensional oriented self-similarity features (MOSS), which leverages the advantages of multi-dimensional oriented self-similarity features. MOSS adopts a coarse-to-fine matching strategy to achieve high-precision and prompt feature point correspondence. The key contributions of this research are as follows:

(1) A lightweight framework for the registration of MRSI with hybrid feature guidance was proposed. The framework adopts a coarse-to-fine matching strategy, maintaining geometric invariance in matching while achieving high-precision identification of feature point correspondence.
(2) A representation method based on multi-dimensional oriented self-similarity features was introduced. It fully utilizes the multi-channel self-similarity maps constructed in the hybrid feature coarse matching stage to build multi-dimensional oriented self-similarity template features, improving feature utilization and promoting registration efficiency.

This article is structured as follows. Section 1 outlines the background and challenges of MRSI registration. Section 2 offers a concise overview of relevant studies. Section 3 details the proposed MOSS algorithm. Section 4 establishes the optimal parameters for the MOSS algorithm and presents the experimental results. Section 5 analyzes the performance of MOSS in handling rotation and scaling transformations.. Finally, Section 6 summarizes the paper.

2. Related works

Currently, the registration methods for MRSI can be broadly categorized into area-based, feature-based, deep learning-based, and joint multi-feature-based registration methods.

The key to area-based methods lies in similarity measurement, including techniques such as normalized cross-correlation (NCC) (Yoo and Han, 2009), mutual information (Suri and Reinartz, 2010) and phase correlation (Foroosh et al., 2002). The dense local self-similarity descriptor utilized NCC for similarity measurement (Ye et al., 2017a). The histogram of orientated phase congruency facilitated rapid matching (Ye et al., 2017b). The PCSD (Fan et al., 2018) combined nonlinear diffusion and phase consistency for image registration. Channel features of orientated gradient extended HOG using FFT for similarity measurement (Ye et al., 2019). MoTIF employed a diffusion tensor model and polar coordinates for feature vector descriptors (Yao et al., 2022a). Area-based methods were sensitive to geometric distortions, necessitating georeferencing to enhance registration accuracy and reliability.

Feature-based methods extract local features from images to achieve faster registration. Since the introduction of SIFT (Lowe, 1999), several variations of SIFT were developed (Dellinger et al., 2015; Ma et al., 2017; Sedaghat et al., 2011; Sedaghat and Ebadi, 2015a; Xiang et al., 2018; Yi et al., 2008). Ji et al. (2013) proposed a nonlinear intensity difference correlation matching method. LGHD (Aguilera et al., 2015) utilized phase information in the frequency domain. The histogram of oriented self-similarity (Sedaghat and Ebadi, 2015b) incorporated orientation information. RIFT (Li et al., 2020) adopted phase consistency and maximum index map descriptor. Xiang et al. (2020) improved the PC model. Yao et al. (2022) proposed a method using co-occurrence filter space and feature displacement optimization. Fan et al., (2022a) designed a multi-scale PC descriptor. Zhang et al. (2023) introduced the histogram of the orientation of weighted phase (HOWP) for feature aggregation. R2FD2 (Zhu et al., 2023) consisted of repeatable feature detectors and descriptors. Huang et al. (2023) constructed the descriptor HOSC with spectrum congruency features. Chen et al. (2023) proposed a method based on adaptive line segment features.

Compared to gradient information and phase consistency, local self-similarity (LSS) (Suetake et al., 2008) features capture the structural information of an image. Xiong et al. (2021) introduced a descriptor based on oriented self-similarity (OSS) features. This approach enhances the computational efficiency of the original LSS features while suppressing image noise. However, the utilization of OSS features is relatively limited, resulting in lower registration accuracy.

Deep learning models can automatically learn image feature representations. Therefore, many researchers have proposed deep learning-based approaches to address challenges in MRSI registration. Zhang et al. (2019) introduced a framework for MRSI image registration based on convolutional Siamese networks. D2-net (Dusmanu et al., 2019) combined feature description and detection using convolutional neural networks. Ji et al. (2022) proposed a feature fusion-based registration method using deep convolutional neural networks. R2D2 (Revaud et al., 2019) performed keypoint detection and description through learning feature detectors, descriptors, and local descriptor predictors. LoFTR (Sun et al., 2021) employed self-attention and cross-attention layers to obtain feature descriptors of two images. SIFNet (Liu et al., 2023) was an algorithm that utilized a self-attention interactive fusion network. Zhang et al. (2022) proposed a self-supervised training network for detecting optimal keypoints. Li et al. (2022) introduced the cross-modal Matching Network (CM-Net). Meng et al. (2021) introduced an end-to-end registration network named DSIM. Li et al., (2023a) presented an end-to-end framework with self-attention and dual-supervision loss. Xu et al. (2023) proposed the local descriptor SODescNet. Ji et al. (2023) evaluated traditional and deep learning-based methods, finding room for improvement in deep learning approaches. Deep learning methods face challenges in obtaining large-scale annotated data, particularly in specific scenarios or with sparse data. Moreover, these methods require substantial computational resources, and their transferability and generalization capabilities still need improvement.

Researchers have been exploring methods that combine multiple features for image registration to leverage the advantages of both area-based and feature-based approaches. For instance, Gong et al. (2014) combined SIFT with mutual information. Xiong et al. (2016) integrated line features and mutual information. Zhang et al. (2020) combined SAR-SIFT with the area-based ROEWA-HOG method. The 3MRS algorithm (Fan et al., 2022b) employed phase consistency and Log-Gabor filters to achieve a combination of feature matching and template matching.

However, these methods still have limitations. They lack robustness in multi-modal registration scenarios. Furthermore, some feature descriptors fail to fully utilize shape or structural information in images, leading to performance degradation. Moreover, many methods struggle with handling significant geometric transformations, limiting their robustness. Therefore, a lightweight MRSI hybrid registration algorithm based on MOSS was proposed. The MOSS algorithm is designed to be robust, adaptable to multi-modal matching scenarios, handle different image types, and accommodate large geometric transformations.

3. Method

In this study, a coarse-to-fine registration strategy was proposed to progressively enhance the matching performance of MRSI. The hybrid feature coarse matching stage extracts self-similarity features and matching them in a unified feature space. The OSS algorithm (Xiong et al., 2021) is employed in this stage for the description of feature points, leveraging the advantages of the directional information of self-similarity features. This stage eliminates the geometric differences

between MRSI and then estimates the initial affine transformation model. The fine matching under multi-dimensional oriented self-similarity features stage takes multi-channel self-similarity maps as input, constructs multi-dimensional oriented self-similarity features, and develops a template-like matching method to refine the results from the coarse matching stage. In this stage, for more efficient template matching, a feature sparsification strategy was designed, and 3D Gaussian kernel convolution was applied to enhance the feature representation capability. Subsequently, a 3D phase correlation matching strategy was used to establish more accurate registration relationships. Lastly, the Fast Sample Consensus (FSC) algorithm(Wu et al., 2015) was utilized to remove misalignments. Fig. 1 illustrates the workflow of the MOSS algorithm.

3.1. Hybrid feature coarse matching

In hybrid feature registration, the first step is to perform hybrid feature coarse matching, primarily using the OSS algorithm. This algorithm utilizes the directional information of self-similarity features to describe the feature points and employs the nearest neighbor distance ratio matching strategy to determine the initial correspondences. Additionally, the FSC is applied to effectively remove incorrect matches. In this section, the process of extracting and describing the OSS features is elaborated upon.

3.1.1. Extraction of oriented self-similarity features

In the hybrid feature coarse matching stage, the self-similarity features of the image are first rapidly extracted using the offset mean filtering method to generate a multi-channel self-similarity feature map. This part will be highlighted in the fine matching under multi-dimensional oriented self-similarity features stage. The oriented self-similarity feature detector considers pixels that are highly dissimilar to their surrounding pixels as keypoints. Hence, for each pixel in the image, its self-similarity feature values are extracted, and the n smallest values among them are computed. The feature response λ for a point q is calculated as shown in Equation (1). By assembling the feature responses λ of all pixels, a feature response map is obtained, which is then subjected to local non-maximum suppression to obtain the feature points.

$$\lambda(q) = \frac{1}{n}\sum_{i=1}^{n} S_i^q \tag{1}$$

The above equation $\lambda(q)$ represents the feature response of point q and $\{S_i^q | i = 1, 2, \ldots, n\}$ represents the n smallest self-similarity values, where n is set to 4 in this paper.

To address the potential scale differences in MRSI registration, this paper adopts a multi-scale Gaussian pyramid approach to feature extraction. Features are extracted from images at different scales by detecting feature points layer by layer. Additionally, an image blocking strategy is employed to obtain uniformly distributed feature points.

3.1.2. Description of oriented self-similarity features

The oriented self-similarity feature description consists of two key steps: calculation of the primary orientation of the keypoint and statistics of the descriptor feature vectors.

(1) Calculation of the primary orientation of the keypoint: This step aims to ensure the rotation invariance of the descriptor. A fixed circular neighborhood is selected around the keypoint, and the main direction is determined based on the orientation histogram generated from the self-similarity feature values. First, the histogram is divided into 36 equal bins, each representing a 10° interval. Then, for the feature point P whose main direction is to be determined, 36 points are uniformly sampled on the boundary of a circular neighborhood with a radius of r around P. For these 36 points, their self-similarity feature sequences are computed and denoted as $S_1, S_2 \ldots S_{36}$. Next, the self-similarity feature sequences are normalized, and the peak directions that account for more than 80 % of the histogram are selected as the primary orientation of the keypoints.

(2) Statistics of the descriptor feature vectors: To improve the efficiency of feature matching and reduce the dimensionality of feature descriptors, this study utilizes a logarithmic polar grid to statistically analyze the feature vectors. The descriptor neighborhood of a feature point is a circular region extracted from the multi-channel self-similarity map. At each pixel within the

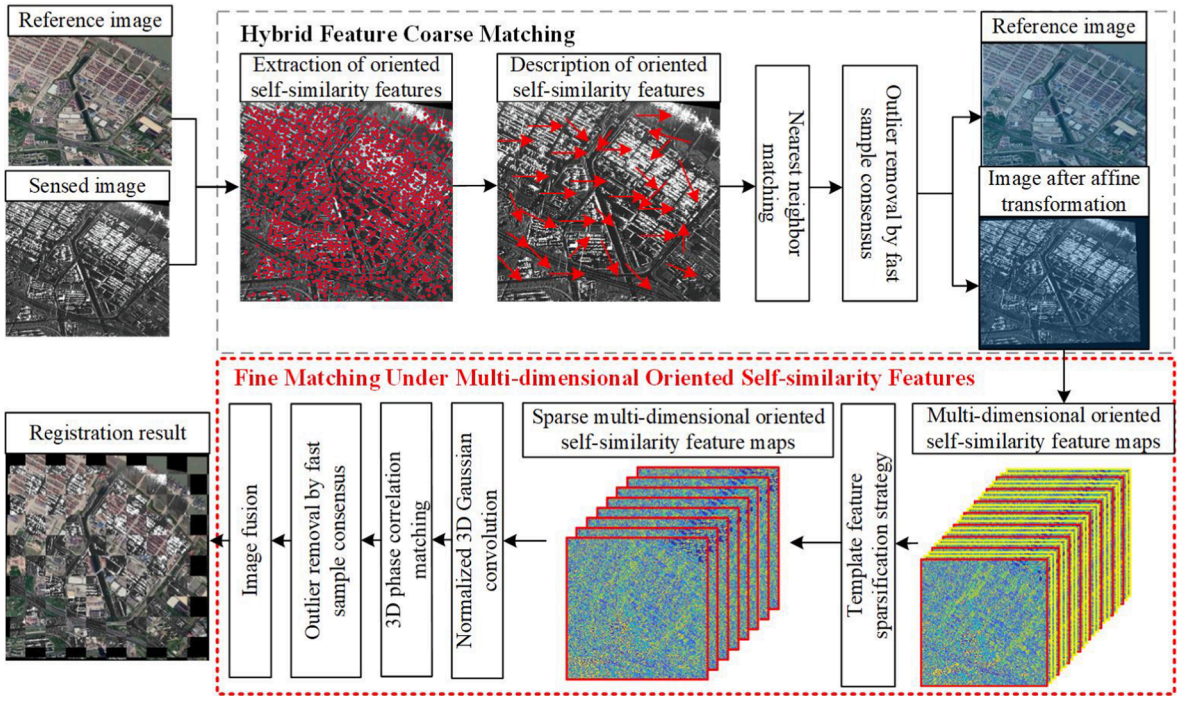


Fig. 1. MOSS algorithm registration workflow.

descriptor neighborhood, the index value of the minimum self-similarity direction is calculated, resulting in an orientation index map. Finally, the orientation index map is transformed into a logarithmic polar grid, and a histogram is computed within each grid interval to generate the feature descriptor vector.

3.2. Fine matching under multi-dimensional oriented self-similarity features

In the hybrid feature coarse matching stage, the geometric differences between multi-modal images are addressed by the oriented self-similarity feature description. However, due to the limitations of feature point detection accuracy, the registration accuracy is not high. To improve registration accuracy, this paper combines the advantages of template matching and proposes a registration method based on multi-dimensional oriented self-similarity template features. In the fine matching under multi-dimensional oriented self-similarity features stage, the following steps are performed. First, a multi-channel self-similarity map is constructed. Then, the multi-dimensional oriented self-similarity template features are computed, and a sparsity strategy is employed. Next, the template feature channels are enhanced using normalized 3D Gaussian convolution kernels. Simultaneously, an adaptive non-maximal suppression (ANMS) feature detector (Brown et al., 2005) is used to detect keypoints in MRSI. Considering that the template features are three-dimensional, the Fourier transform is applied to transform the feature template from the spatial domain to the frequency domain, and 3D phase correlation is used as the similarity measure to accelerate the matching of template-like features. Finally, the erroneous matches are eliminated through the employment of the FSC algorithm.

In this section, the focus is on describing the construction process of multi-dimensional oriented self-similarity features, which includes three steps: construction of the multi-channel self-similarity map, computation of the multi-dimensional oriented self-similarity features, and normalized 3D Gaussian convolution. The entire process is illustrated in Fig. 2.

3.2.1. Construction of multi-channel self-similarity maps

The multi-channel self-similarity maps are obtained using the offset mean filtering technique, which involves: image cropping and mean filtering. Firstly, the original image is cropped to create the center image block I^c_{crop}, representing the region around the point $q(\rho,\theta)$(ρ represents the distance from a pixel point to its neighboring pixels in the feature neighborhood, and θ represents the angle, $\theta \in [0,\pi]$). Then, the offset image block I^o_{crop} is obtained by shifting the center block in different directions. Once the image blocks are obtained, the self-similarity map S^q in the direction of the point $q(\rho,\theta)$ can be computed using the following equation:

$$S^q = meanFilter\left(\left|I^c_{crop} - I^o_{crop}\right|\right) \tag{2}$$

Where $meanFilter(\bullet)$ represents the process of applying mean filtering to image blocks. The filtering window is set as a circular shape with a radius of 2 pixels, as it enhances the rotational invariance of the self-similarity features.

A multi-channel self-similarity map named $\{S^q\}_C^1$ can be obtained by $C = N/2$ offset mean filtering operations, where C represents the number of channels and N represents the number of pixels in the feature neighborhood. Due to the symmetry of self-similarity features, the multi-channel self-similarity map $\{S^q\}_C^1$ can capture the self-similarity characteristics of the entire image.

3.2.2. Solving multi-dimensional oriented self-similarity features

This paper proposes an angle-weighting strategy. Firstly, for the obtained multi-channel self-similarity map $\{S^q\}_C^1$, g_x and g_y are calculated for each channel using the Sobel operator $h = [1,2,1;000;-1,-2,-1]$, as shown in equation (3):

$$\begin{cases} g_x(x,y) = S^q \otimes h \\ g_y(x,y) = S^q \otimes h^T \end{cases} \tag{3}$$

Where S^q represents the input single-channel self-similarity map, and $\otimes$ represents the convolution operation.

Then, the gradient direction is calculated using equation (4). For each channel, an angle-weighting strategy is applied to the multi-channel self-similarity map, resulting in a single-channel multi-dimensional oriented self-similarity feature S^q_o. This strategy effectively enhances the rotational invariance of the self-similarity template feature. It is worth noting that if the gradient direction is less than 0,π should be added to ensure that the gradient direction value is always greater than 0.

$$S^q_o = \tan^{-1}\left(\frac{-g_y(x,y)}{g_x(x,y)}\right) \tag{4}$$

$$S^q_o = \begin{cases} S^q_o, & S^q_o \geqslant 0 \\ S^q_o + \pi, & S^q_o < 0 \end{cases} \tag{5}$$

Where S^q_o represents the single-channel multi-dimensional oriented self-similarity feature.

Finally, an angle-weighting strategy is applied to the multi-channel self-similarity features, channel by channel, to construct the multi-channel multi-dimensional oriented self-similarity features $\{S^q_o\}_C^1$. In the construction of the multi-channel self-similarity map, a circular feature neighborhood with a radius of R and N pixels, and $C = N/2$ channels are used. Therefore, the dimension of the multi-channel multi-dimensional oriented self-similarity features $\{S^q_o\}_C^1$ is C. To enhance the computational efficiency of template feature matching, a sparsification strategy is applied to the feature channels by setting the sparsification interval as S_t. Equation (6) is used for sparsification, resulting in a w-dimensional template feature map $\{S^q_o\}_w^1$.

$$w = \lceil C/S_t \rceil \tag{6}$$

where w represents the dimension of the sparsified multi-

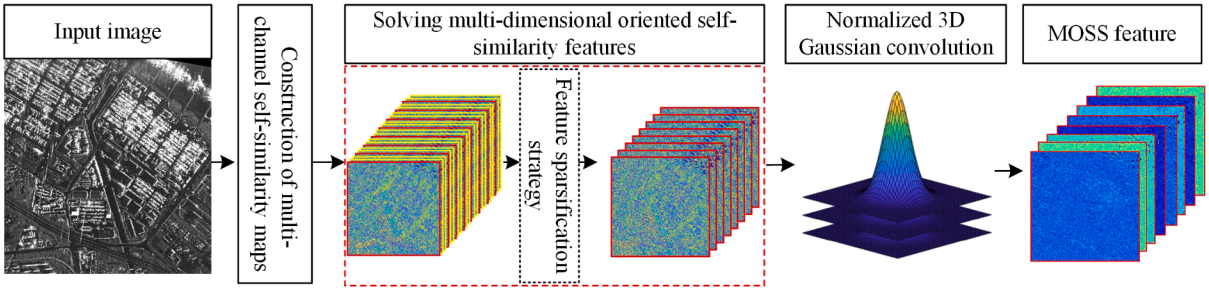


Fig. 2. Construction process of multi-dimensional oriented self-similarity features.

dimensional oriented self-similarity features, C represents the dimension of the original multi-dimensional oriented self-similarity features, S_t is the sparsification interval, and the operator indicates rounding up to the nearest integer in the direction of increasing absolute value.

3.2.3. Normalized 3D Gaussian convolution

After forming the multi-dimensional oriented self-similarity feature channels, a three-dimensional Gaussian convolution is applied to enhance the feature channels and mitigate the impact of local distortions resulting from geometric deformations and variations in intensity in the three-dimensional images. This Gaussian convolution kernel comprises a two-dimensional Gaussian kernel on the plane (with a standard deviation of 0.8) and $d_z = [1, 3, 1]^T$. The 3D Gaussian convolution is represented by the equation:

$$S_o^\sigma(x, y) = g_{xy}^\sigma \otimes S_o^q \tag{7}$$

$$M^\sigma(x, y) = d_z \otimes S_o^\sigma(x, y) \tag{8}$$

Where $S_o^\sigma(x, y)$ represents the feature after the two-dimensional Gaussian filtering, g_{xy}^σ represents the Gaussian kernel in the xy plane,S_o^q represents the single-channel multi-dimensional oriented self-similarity feature, $M^\sigma(x, y)$ represents the template feature after the three-dimensional Gaussian filtering,d_z represents the Gaussian kernel in the z-direction,$\otimes$ represents the convolution operation, and σ is the standard deviation of the Gaussian convolution kernel.

Next, normalization is applied to scale the feature vector $M^\sigma(x, y)$ to the same scale, enhancing its robustness. It can be represented by equation (9):

$$M_i^\sigma(x, y) = \frac{M_i^\sigma(x, y)}{\sqrt{\sum_{i=1}^{w} \left|M_i^\sigma(x, y)\right|^2 + \varepsilon}} \tag{9}$$

To prevent division by zero, the value of ε is set to a non-zero extremely small constant.

Finally, the features of $M_i^\sigma(x, y)$ are combined to form the multi-dimensional oriented self-similarity feature map $MOSS(x, y)$. It can be expressed using the equation (10):

$$MOSS(x, y) = \left\{M_i^\sigma(x, y)\right\}, \quad i = [1, 2, 3\cdots, w] \tag{10}$$

where w represents the dimension of the sparsified multi-dimensional oriented self-similarity features.

4. Experimental results

To evaluate the performance of the MOSS algorithm, this study compared it with seven advanced algorithms, including SIFT (Lowe, 1999), PSO-SIFT (Ma et al., 2017), RIFT (Li et al., 2020), OSS (Xiong et al., 2021), MS-HLMO (Gao et al., 2022), HOWP (Zhang et al., 2023), and SRIF (Li et al., 2023b). To ensure a fair comparison, the provided codes from the respective authors were utilized, and optimal parameters were set. The image scale difference was set to 1.2, the number of pyramid layers was set to 6, and the threshold for coarse outlier rejection was set to 3 pixels, with a minimum of 4 matched point pairs required (Yao et al., 2022). The parameters were fine-tuned for optimal performance, with a neighborhood radius (R) of 4 in the hybrid feature coarse matching stage, a template window size of 96 in the fine matching under multi-dimensional oriented self-similarity features stage, and a sparsification interval (S_t) of 2 for generating the multi-dimensional oriented self-similarity template features. In the MOSS algorithm, the calculation of the number of extracted feature points is given by:

$$PT_{nums} = round(M{*}N{*}K) \tag{11}$$

PT_{nums} represents the number of extracted feature points, where M and N represent the length and width of the image, and K is a proportionality factor for determining the number of extracted feature points. A parameter analysis was conducted for K, with its values ranging from 0.001 to 0.009. In the text, it was set to 0.003. Further details are discussed in Section 4.3.

The experiments were conducted using Matlab R2018a and performed on the following experimental platform: AMD Ryzen 9 5900HX with Radeon Graphics processor running at a frequency of 3.30 GHz, 64 GB of memory, and Windows 10 x64 operating system.

4.1. Experimental dataset

This study utilized an experimental dataset comprising six categories of MRSI pairs, including multi-temporal optical images, infrared and optical images, day and night images, SAR and optical images, map and optical images, and optical and depth images. The dataset consists of a total of 48 pairs of MRSI, with 8 pairs of images for each data type. For demonstration purposes, one image pair from each data type was selected as shown in Fig. 3. This dataset of MRSI encompasses nearly all application scenarios of MRSI matching. Due to differences in time, lighting conditions, and sensors, significant challenges exist in terms of noise interference, NRD, and geometric transformation differences between image pairs. Over 10 well-distributed ground control points were manually collected for each image pair to evaluate the accuracy of matched correspondences using the MOSS algorithm. Therefore, this dataset can be used to test the performance of the MRSI registration algorithms.

4.2. Evaluation metrics

The MOSS algorithm in this study was qualitatively and quantitatively evaluated to comprehensively validate its performance. Five metrics: success rate (SR), number of correct matches (NCM), matching time (MT), standard deviation (SD) and root mean square error (RMSE).

SR denotes the rate of successful matches. This measure indicates the resilience of the matching technique when employed on particular categories of MRSI pairs. NCM refers to the number of image pairs with more than 20 matched corresponding points while excluding image pairs with RMSE greater than 7 pixels. The count of matched pairs with RMSE within 7 pixels is considered the NCM. MT refers to the execution time of the matching algorithm, used to measure the efficiency of the matching algorithm. SD is an indicator to evaluate the uniformity of feature point extraction (Yao et al., 2022). It measures the distribution of feature points within a defined standard deviation box, which is a fixed area used to determine the neighborhood of feature points. For each standard deviation box, the standard deviation of feature points within the box is calculated. The standard deviation measures the dispersion of the data, and smaller the standard deviation, the more evenly distributed the points are within the box. RMSE indicates the accuracy of the match. The mathematical representation of RMSE is presented in Equation (12).

$$RMSE = \sqrt{\frac{1}{N}\left(\sum_{i=1}^{N}\left[(x_i^{'} - x_i^{''})^2 + (y_i^{'} - y_i^{''})^2\right]\right)} \tag{12}$$

where N is 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.

4.3. Parameter sensitivity analysis

To comprehensively evaluate the performance of the MOSS algorithm, an analysis of its four core parameters was conducted. The parameters are the neighborhood radius (R), the template window size (T_w), the sparsification interval (S_t) and the proportionality factor (K) for extracting feature points. The RMSE, NCM, and MT were used as evaluation metrics to measure the impact of these parameters. The specific

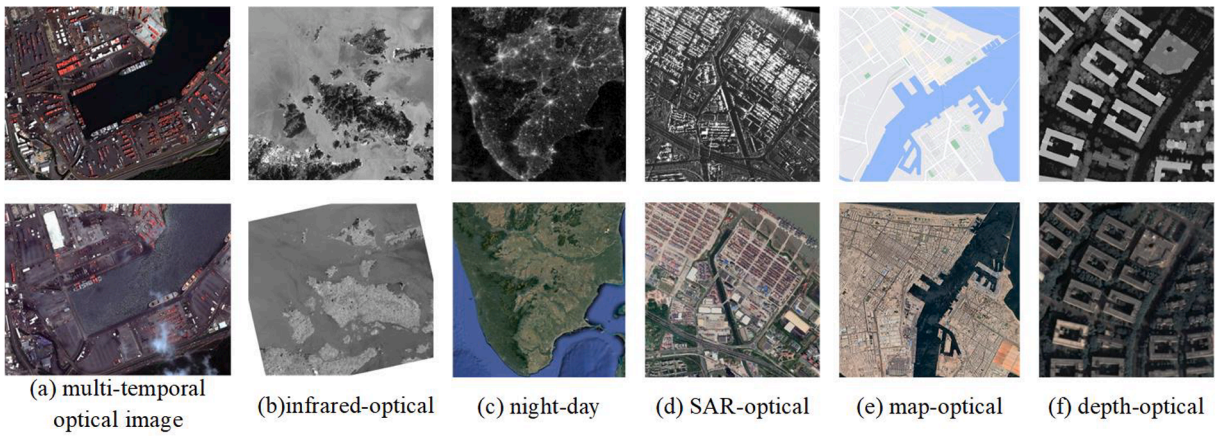


Fig. 3. Partial display of the MRSI dataset.

parameter settings are shown in Table 1. In Fig. 4(b), R and S_t jointly determine the dimensions of the multi-dimensional oriented self-similarity features. Therefore, a comprehensive analysis of these two factors can reveal their impact on the matching results. By adjusting and evaluating these parameters, the optimal parameter configuration was obtained to achieve the best matching performance. This parameter selection not only achieved desirable results in terms of RMSE and NCM but also completed the matching task within an acceptable time range. The units for RMSE are pixels, for NCM are the number of points, and for MT are seconds.

Then, we tested the performance of MOSS with different parameter settings using 48 sets of MRSI data. Fig. 4(a) illustrates the matching results at different values of T_w. As T_w increases from 56 to 96, the NCM results show a steady upward trend, reaching 420. However, when T_w exceeds 96, NCM starts to slightly decrease, then slightly increases again until reaching $T_w = 136$, where NCM stabilizes at around 430. For RMSE, the lowest value is achieved at $T_w = 96$, which is 1.86 pixels. At this point, the MT is maintained at a relatively low level. Considering both NCM, MT and RMSE, setting T_w to 96 can achieve the highest accuracy and abundant matching correspondences, thus it is recommended to set T_w as 96.

Fig. 4(b) displays the results at different values of R and S_t. From the graph, it can be observed that as R increases, the NCM count also increases, but at the same time, the time consumption increases, resulting in decreased matching efficiency. Under the same R value, when S_t varies between 1 and 6, NCM shows a decreasing trend. Considering the RMSE factor, the optimal parameter settings are $R = 4$ and either $S_t = 1$ or $S_t = 2$. However, it was found that only when $R = 4$ and $S_t = 2$ ensured a good accuracy-efficiency trade-off, that is, with higher RMSE and NCM values and the highest matching efficiency. As R increases, although more matching points can be obtained, the matching efficiency decreases as the increase in the number of matching points cannot compensate for the decrease in matching accuracy and the increase in time consumption. Therefore, this study selects $R = 4$ and $S_t = 2$ as the parameter settings with high matching efficiency and optimal accuracy.

Fig. 4(c) illustrates the results under different settings of the feature point extraction proportionality factor K. From the graph, it can be observed that as K increases, the number of NCMs gradually rises. However, simultaneously, there is an increase in time consumption, leading to a decline in matching efficiency. Taking into account the RMSE factor, for K values greater than 0.003, some images experience matching failures due to unsuccessful feature point extraction, resulting in an increase in RMSE but still maintaining around 2 pixels. Therefore, considering $K = 0.003$ achieves a good balance between RMSE and NCM values while minimizing time consumption, achieving the highest matching efficiency.

4.4. Experimental results

4.4.1. Quantitative evaluation

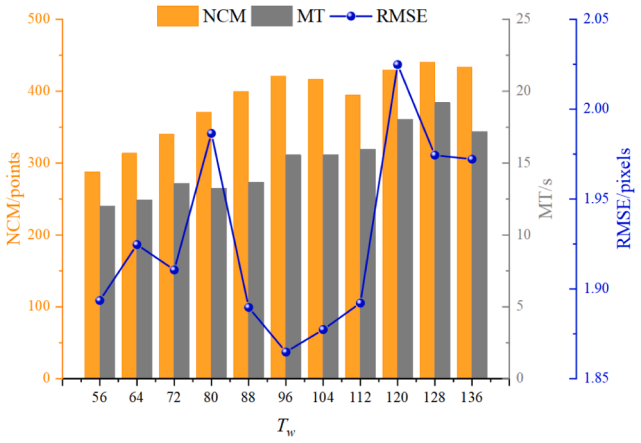
Table 2 shows the average results for the five indicators. Since the SIFT, PSO-SIFT and MS-HLMO algorithms performed poorly on this MRSI dataset and extracted very few feature points, the SD metric is not evaluated for these three algorithms. In Fig. 5(a), "Image Type 1″ represents multi-temporal optical–optical, "Image Type 2″ represents infrared-optical, "Image Type 3″ represents night-day, "Image Type 4″ represents SAR-optical, "Image Type 5″ represents map-optical, and "Image Type 6″ represents depth-optical. In Fig. 5(b), the symbol $+\infty$ indicates matching failure or RMSE > 7 pixels. SR is in %, NCM is in points, RMSE is in pixels and the unit of MT is seconds.

Fig. 5 presents a comparison of matching results between the MOSS algorithm and seven other methods using two metrics, NCM and RMSE. As shown in Table 1, the SR of the SIFT method is only 41.7 %, with an average NCM of 27.90, which is the lowest among the six algorithms. The RMSE is 4.88 pixels, indicating the lowest matching accuracy. The PSO-SIFT algorithm's matching performance is slightly better than the SIFT algorithm, with an SR of 62.5 % and an average NCM of 75.69. The RMSE is 4.57 pixels. However, the PSO-SIFT algorithm is more sensitive to signal-to-noise interference. Thus, the applicability of both the SIFT and PSO-SIFT algorithms is limited in MRSI matching scenarios.

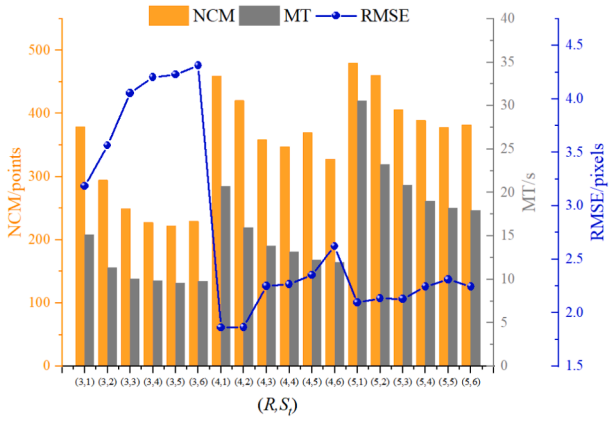
The results of the RIFT algorithm are represented by a green dashed line. This algorithm uses image frequency domain features for matching and significantly improves matching performance through the maximum index map descriptor. The SR is improved to 75.0 %, and the average NCM is 136.13. With an SD of 1.02, it indicates a relatively uniform distribution of extracted feature points. The average RMSE is 3.82 pixels, and the RMSE results fluctuate due to the algorithm's lack of support for scale differences.

Table 1
Recommended parameter settings for the MOSS algorithm.

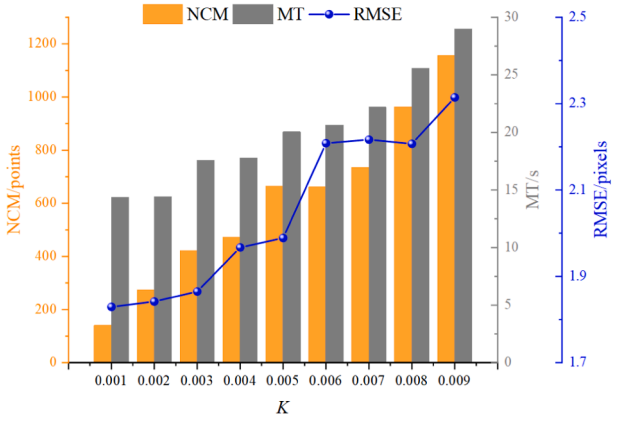
Parameter	Value	Fixed Parameter
T_w	T_w = [56,64,72,80,88,96,104,112,120,128,136]	$R = 4$, $S_t = 2$, $K = 0.003$
S_t	S_t = [1,2,3,4,5,6]	$T_w = 96$, $K = 0.003$
R	R = [3,4,5]	$T_w = 96$, $K = 0.003$
K	K = [0.001,0.002,0.003,0.004,0.005,0.006,0.007,0.008,0.009]	$T_w = 96$, $R = 4$, $S_t = 2$



(a) Matching results of different T_w.



(b) Matching results of different R and S_t.



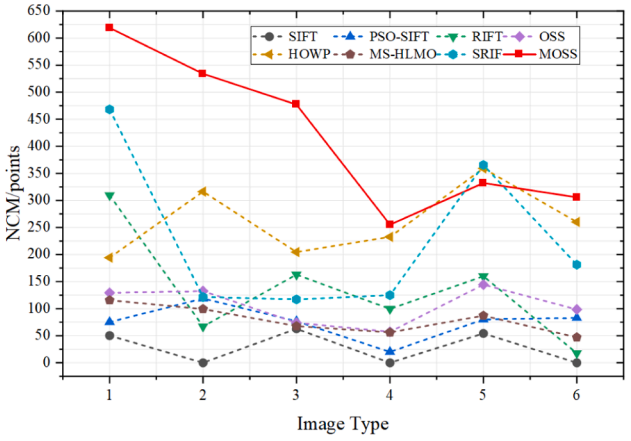
(c) Matching results of different K.

Fig. 4. Evaluation results of the four parameters.

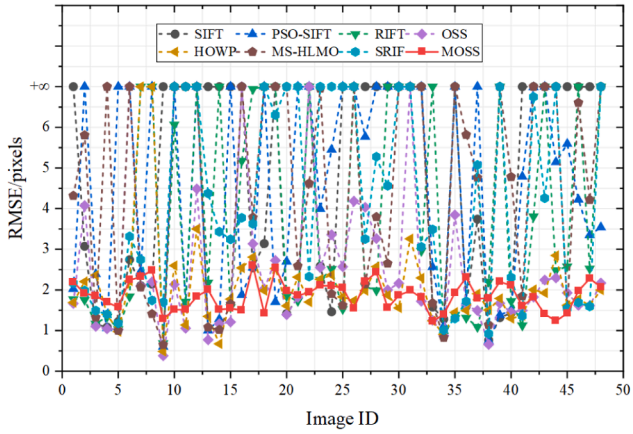
The results of the OSS algorithm are represented by a purple dashed line. This algorithm utilizes offset mean filtering for fast computation of self-similarity features. It shows good SR results, reaching 91.7 %, with an SD of 1.12, indicating a moderate level. However, the NCM is relatively low, with an average NCM of 106.04, lower than RIFT, HOWP, SRIF and MOSS algorithms. This is because the LSS descriptor has low discriminative power and cannot capture rich NCM. The RMSE of this algorithm is relatively low among the eight methods, with an average RMSE of 2.45 pixels.

The results of the HOWP algorithm are represented by a yellow dashed line. This algorithm achieves successful matches in most of the data, with an SR of 95.8 %, an average NCM of 261.19, and an average RMSE of 2.09 pixels, second only to the MOSS algorithm. However, the SD is 1.48, indicating uneven distribution of extracted feature points, and the algorithm lacks robust rotation invariance for MRSI matching with large rotation angles.

The results of the MS-HLMO algorithm are represented by the brown dashed line. This algorithm utilizes Generalized Gradient Location and



(a) NCM results for eight methods



(b) RMSE results for the eight methods

Fig. 5. Test results of the eight methods.

Table 2
Evaluation results of the eight methods on five evaluation indicators.

	SIFT	PSO-SIFT	RIFT	OSS	HOWP	MS-HLMO	SRIF	MOSS
SR	41.7 %	62.5 %	75 %	91.7 %	95.8 %	60.42 %	75 %	**100 %**
NCM	27.90	75.69	136.13	106.04	261.19	78.83	229.75	**420.52**
RMSE	4.88	4.57	3.82	2.45	2.09	4.51	3.94	**1.86**
MT	4.23	15.62	6.23	18.65	9.32	53.20	5.89	**15.56**
SD	/	/	1.02	1.12	1.48	/	1.47	**0.87**

Orientation Histogram (GGLOH) to construct feature descriptors, providing rotation and scale invariance. The algorithm exhibits poor applicability for SAR image matching, with an overall SR result of only 60.42 %. This is attributed to the limited number of successfully matched points, making it challenging to assess the uniformity of the extracted feature points. Consequently, standard deviation is not statistically analyzed. The NCM is relatively low, with an average NCM of 78.83, only slightly higher than the SIFT and PSO-SIFT algorithms. The RMSE of this algorithm is relatively high among the eight methods, with an average RMSE of 4.51 pixels.

The results of the SRIF algorithm are represented by the blue dashed line. The algorithm introduces a Local Intensity Binary Transformation (LIBT) for feature description, achieving successful matches in most data with an SR of 75 %. However, it is susceptible to MRSI signal-to-noise interference, resulting in failures in matching certain SAR and day-night images. The average NCM is 229.75, and the average RMSE is 3.94 pixels, with an SD of 1.47. The algorithm exhibits a relatively uniform distribution of extracted feature points. The algorithm demonstrates fast processing speed but lacks high matching accuracy, occasionally failing to register certain images, indicating a lack of robust registration performance.

The results of the MOSS algorithm are represented by a red solid line, showing the most robust matching results. The MOSS algorithm successfully matches all 48 pairs of MRSI, with an SR of 100 % and the highest average NCM of 420.52. The SD is the lowest at 0.87, indicating a uniform distribution of extracted feature points. The lowest average RMSE is 1.86 pixels. The MOSS algorithm can handle geometric transformation differences, nonlinear radiometric differences, illumination differences, and contrast differences in MRSI, exhibiting scale invariance and rotation invariance. It achieves high-precision identification of corresponding points while maintaining geometric invariance in multi-modal matching.

After fair testing, as shown in Table 2, the MOSS algorithm exhibits an average matching time of 15.56 s, performing better than OSS, PSO-SIFT and MS-HLMO algorithms but slightly inferior to SIFT, RIFT, HOWP, and SRIF algorithms. It is noteworthy that SIFT performs poorly in the majority of matches, making it less suitable for multi-modal matching. In comparison, RIFT, HOWP, and SRIF algorithms belong to the field of feature matching, showing slightly superior efficiency to the algorithm presented in this paper. However, RIFT does not support scale differences in images, HOWP is not applicable to large rotation differences, and the SRIF algorithm is significantly affected by MRSI signal-to-

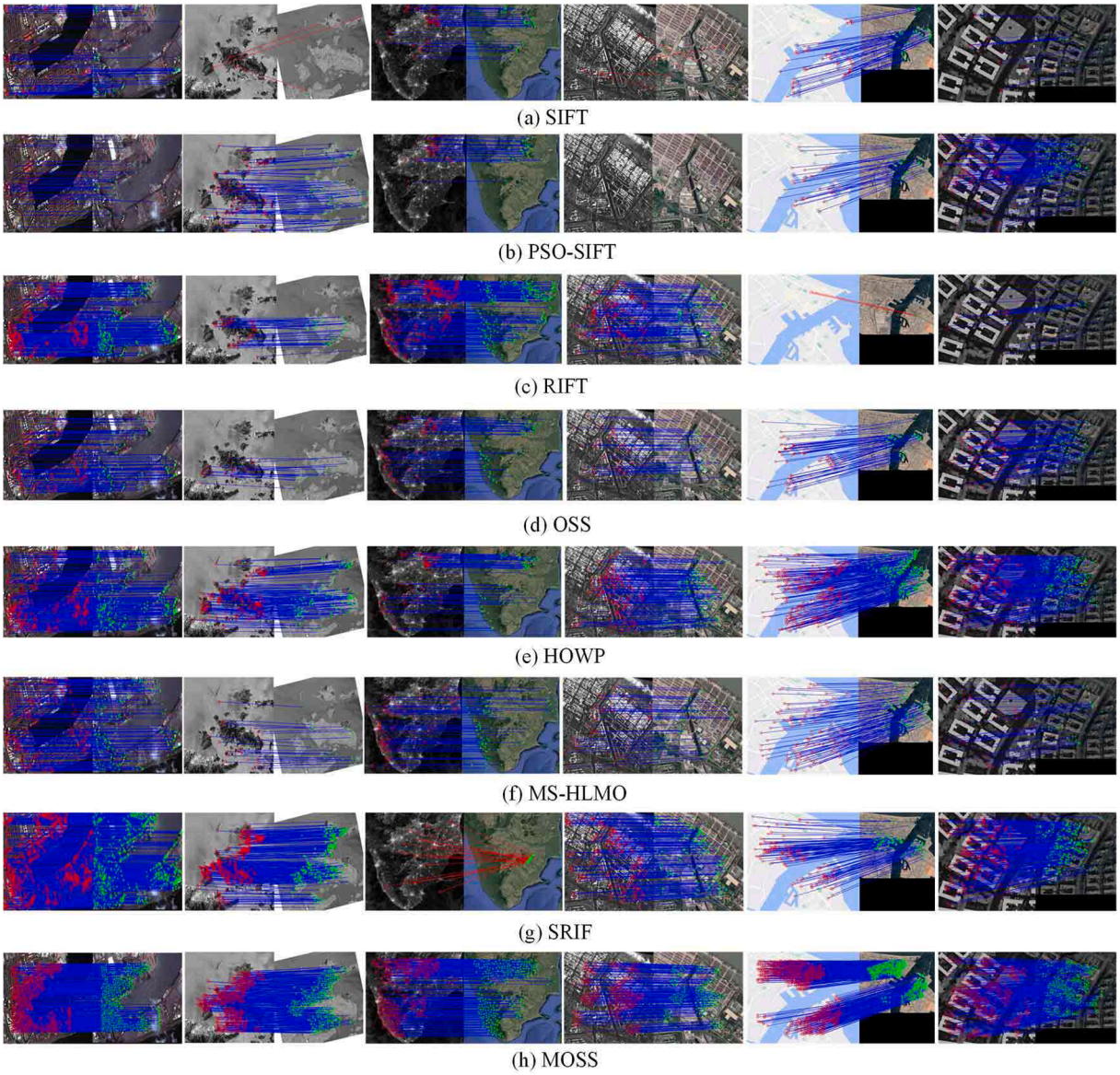


Fig. 6. Matching results of the MOSS algorithm compared with seven other algorithms.

noise ratio interference, with some shortcomings in modality adaptability. Simultaneously, our proposed MOSS algorithm employs a hybrid matching framework, encompassing both feature matching and class template matching. We further calculated the average matching times for its feature matching and template-like matching components, which are 9.3 s and 6.26 s, respectively, demonstrating a notably significant computational efficiency. Therefore, despite the slightly higher time consumption of the MOSS algorithm compared to the latest feature matching algorithms, it excels overall in terms of matching the quantity of corresponding points and matching accuracy compared to other algorithms.

In conclusion, the MOSS algorithm generates more robust matching results in MRSI matching tasks and demonstrates superior overall performance compared to the other seven algorithms.

4.4.2. Qualitative evaluation

In order to further demonstrate the performance of the MOSS algorithm, this study also conducted a qualitative evaluation based on the matching results. The matching outcomes of the MOSS algorithm and the seven comparative methods are depicted in Fig. 6.

Based on Fig. 6, it can be observed that the SIFT algorithm performs the worst in terms of matching performance. The PSO-SIFT algorithm has fewer corresponding points or even matching failures in optical and SAR images with significant nonlinear radiometric differences. The RIFT method, which extracts features based on the phase-consistent maximum and minimum moment images, balances the number and repeatability of features. Its proposed maximum index map descriptor has more robust matching performance but lacks scale invariance. As shown in the figure, its performance is poor in datasets with scale differences. The OSS algorithm, which utilizes offset mean filtering to rapidly compute self-similarity features, exhibits geometric invariance but has fewer and unevenly distributed corresponding points. The HOWP method shows excellent performance in handling small-angle rotational differences, but cannot handle large-angle rotational images. The MS-HLMO algorithm matches a relatively small number of corresponding points, leading to a large registration error and poor applicability in matching SAR images. Although the SRIF algorithm is efficient, its matching results lack robustness. Matches failure may occur in certain modalities. The MOSS algorithm achieves robust matching of images with geometric differences and NRD while achieving high-precision identification of corresponding points. The experimental results further validate the effectiveness and robustness of the MOSS algorithm.

Fig. 7 showcases the matching results of the MOSS algorithm for the remaining 42 images, revealing its excellent performance in MRSI matching. As depicted in Fig. 7, the proposed MOSS algorithm exhibits robustness in MRSI matching. It effectively mitigates the noise interference in multi-modal images and overcomes geometric transformations between different modalities, achieving scale and rotation invariance and yielding abundant NCM. Consequently, the MOSS algorithm holds great potential for applications in MRSI matching.

Fig. 8 demonstrates the registration results of the MOSS algorithm using the chessboard pattern visualization method. In the displayed image chessboard patterns, the well-aligned overlapping regions between the images can be clearly observed. Through quantitative calculations based on ground truth points, registration errors within two pixels are achieved, enabling high-precision identification of matching points in MRSI registration. In general, the evaluation and analysis results confirm the robustness of the proposed MOSS algorithm in dealing with noise interference, NRD, and geometric transformation differences between multi-modal images. The exceptional matching performance of the MOSS algorithm can be attributed to two primary factors. Firstly, the MOSS algorithm employs a hybrid feature matching framework that combines hybrid feature coarse matching with fine matching under multi-dimensional oriented self-similarity features. This approach effectively leverages the advantages of both feature matching and

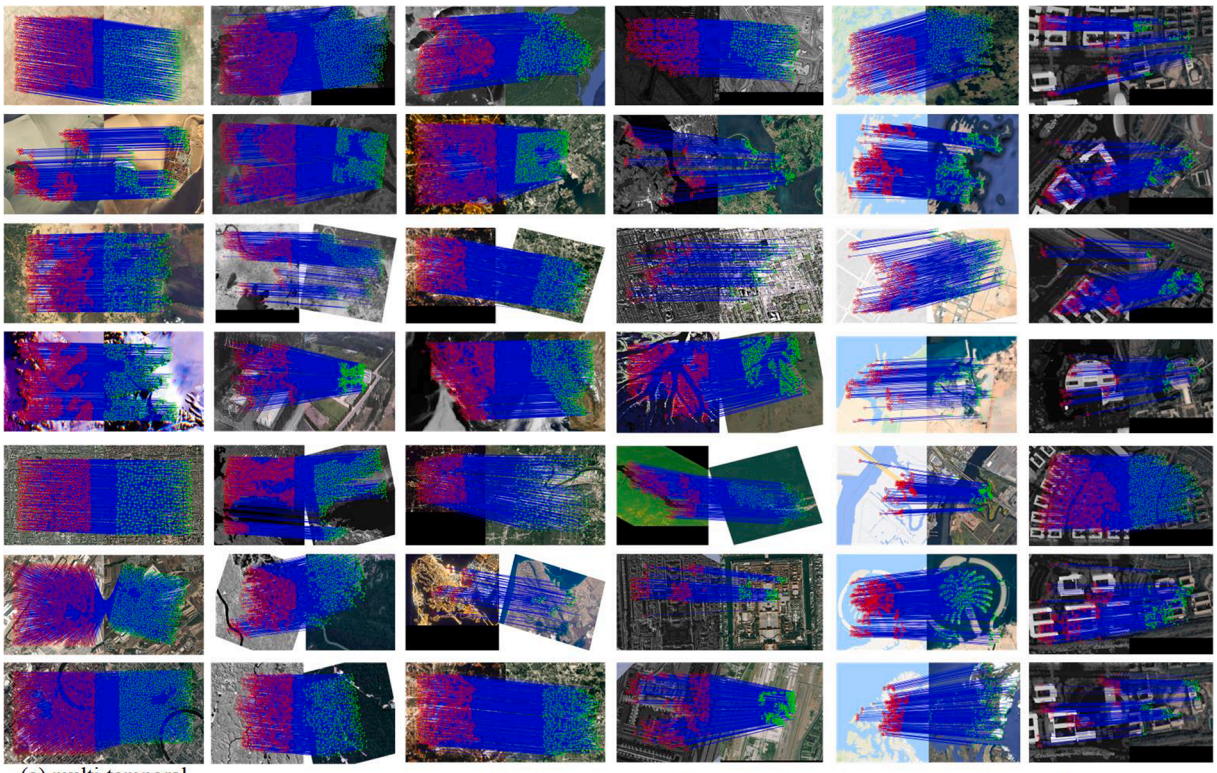


Fig. 7. Matching results of the MOSS algorithm.

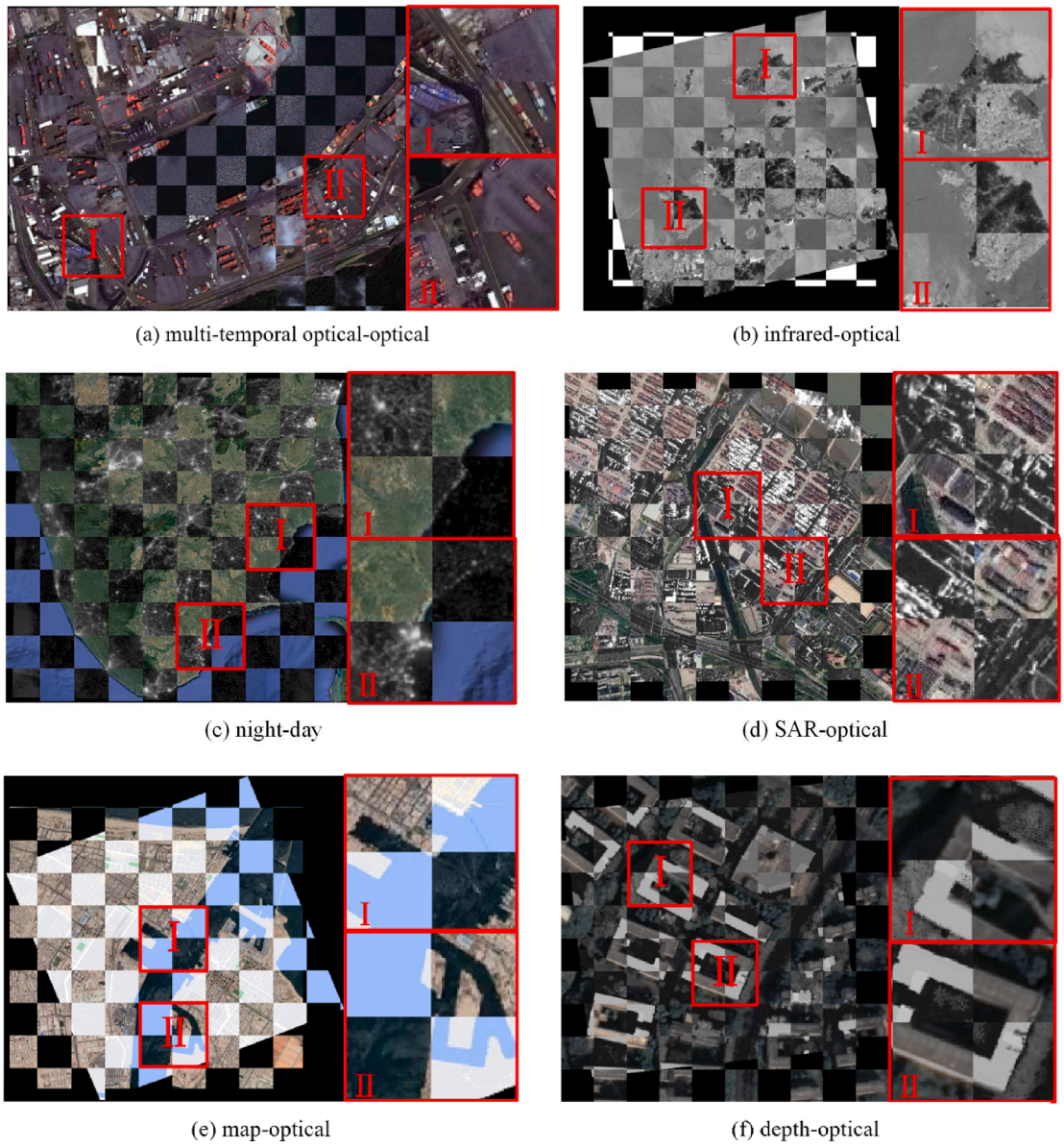


Fig. 8. Partial registration results of the MOSS algorithm.

template matching, enhancing the geometric invariance of features and the accuracy of matching. Secondly, the multi-dimensional oriented self-similarity template features constructed in the fine matching stage possess excellent feature representation capabilities, significantly improving matching accuracy.

5. Discussion and analysis

5.1. Scale invariance analysis

Using a set of map-optical image pairs to assess the scale invariance performance of the MOSS algorithm. Firstly, the reference image was processed to generate 12 simulated images with scale intervals of 0.2, ranging from 0.6 to 3 times the original scale. Fig. 9 shows the matching results. It can be observed that as the scale difference increases, the number of NCM decreases gradually, but it is still sufficient for the registration requirements. In the hybrid feature coarse matching stage, a strategy of constructing a multi-scale Gaussian pyramid is adopted to ensure the uniform extraction of feature points from different scales. Therefore, the MOSS algorithm exhibits scale invariance.

5.2. Rotation invariance analysis

To validate the performance of MOSS in terms of rotation invariance, we further conducted matching tests using multiple sets of MRSI. Firstly, for each modality, a set of images was chosen, and the reference image was rotated in both clockwise and counterclockwise directions at intervals of 30 degrees, resulting in the generation of 6 directions (30°, 60°, 90°, 120°, 150°, and 180°), resulting in 12 simulated images for each modality. Then, these 12 simulated images were used for matching tests, and the matching results were visualized. As shown in Fig. 10 ("-" denotes counterclockwise rotation), the MOSS algorithm consistently achieved successful matches and obtained abundant NCM even with rotation differences ranging from −180° to 180°. This is because, in the hybrid feature coarse matching stage, the algorithm incorporates oriented self-similarity descriptors that exhibit rotation invariance. These descriptors assign a dominant orientation to each feature point and generate orientation histograms based on self-similarity feature values to specify the dominant orientation. In the fine matching under multi-dimensional oriented self-similarity features stage, an angle weighting strategy is applied to the multi-channel self-similarity maps to construct

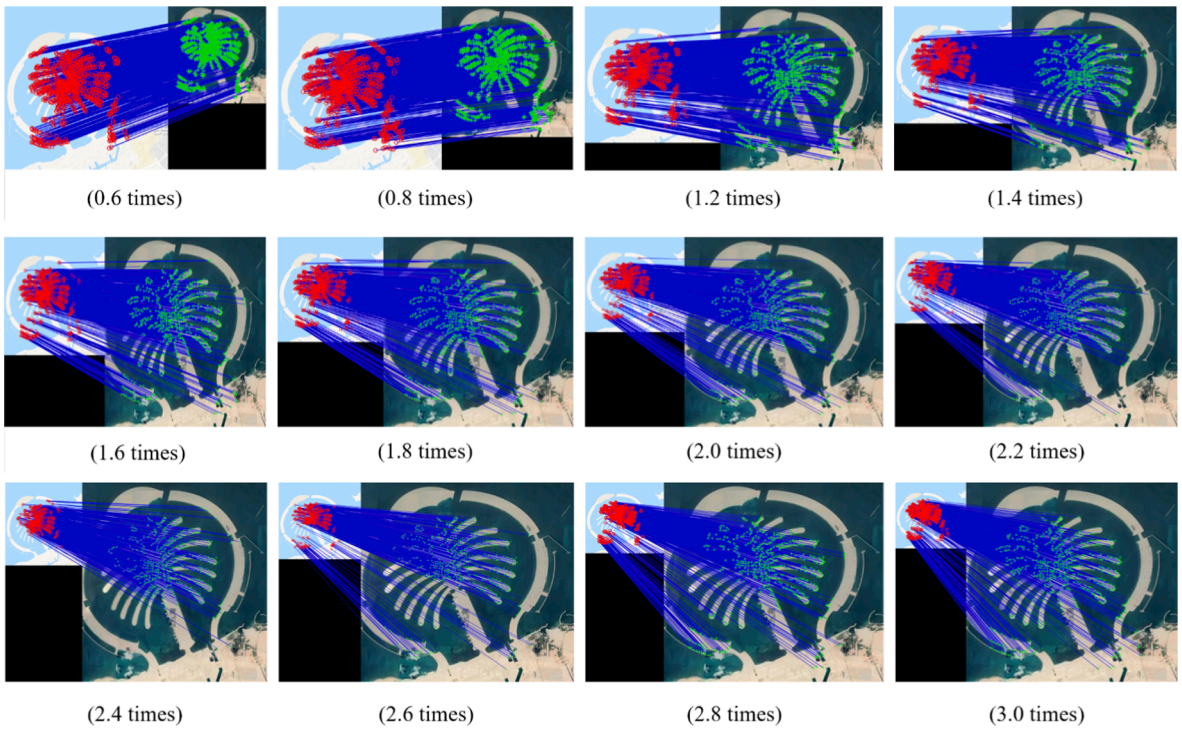


Fig. 9. Matching results of the MOSS algorithm under different scale differences.

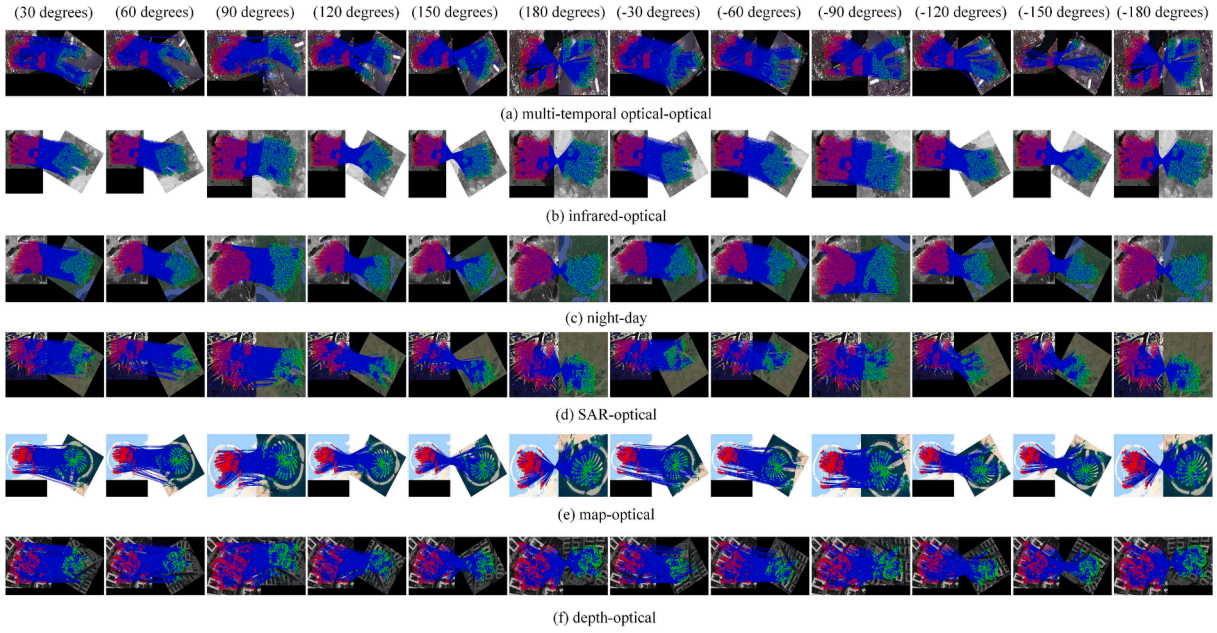


Fig. 10. Matching results of the MOSS algorithm under different rotation differences.

multi-dimensional oriented self-similarity features, which greatly enhances the rotation invariance of self-similarity template features.

In conclusion, the MOSS algorithm demonstrates rotation invariance. Compared to other methods, the proposed MOSS algorithm exhibits the most robust MRSI matching performance in handling challenges such as noise interference, NRD, scale and rotation invariance, and matching accuracy.

6. Conclusion

This study proposes a lightweight and hybrid feature-guided registration algorithm for MRSI called MOSS, which addresses the trade-off between geometric invariance and matching accuracy in MRSI registration effectively. In the hybrid feature coarse matching stage, the geometric transformation differences in multi-modal images were overcome by using oriented self-similarity descriptors. In the fine matching under multi-dimensional oriented self-similarity features stage, a template-like matching algorithm based on multi-dimensional oriented self-similarity features was proposed, with a focus on the construction process of multi-dimensional oriented self-similarity features. Experimental results on various multi-modal image pairs with differences in illumination, contrast, scale, and rotation were compared

with several registration algorithms (SIFT, PSO-SIFT, RIFT, OSS, HOWP, MS-HLMO and SRIF). The quantitative analysis shows that the MOSS algorithm outperforms the SIFT algorithm by 15.1 times, the PSO-SIFT algorithm by 5.6 times, the MS-HLMO algorithm by 5.3 times, the OSS algorithm by 4 times, the RIFT algorithm by 3.09 times, the SRIF algorithm by 1.8 times and the HOWP algorithm by 1.6 times in terms of the NCM. The RMSE achieved the lowest value among all the registration methods, within 2 pixels. The MOSS algorithm exhibits better performance in matching scale, rotation, and translation differences, demonstrating scale invariance and rotation invariance. In the chessboard pattern visualization, it can be observed that the object edges are consistently connected without any crossing patterns, providing qualitative evidence for the effectiveness of the proposed method. To fully evaluate the performance of this method, the four core parameters were also analyzed, and the optimal parameter configuration was determined.

Through MOSS shows high-precision and robust registration performance for MRSI registration, it still has certain limitations. MOSS relies on texture structure features in the images, which may result in suboptimal registration performance in mountainous or complex scenes where distinct texture structures are lacking. Future work will optimize MOSS's matching ability to images with weak texture, and its generalizability to other modalities will also be enhanced.

CRediT authorship contribution statement

Yongjun Zhang: Methodology, Writing – original draft, Writing – review & editing. **Wenfei Zhang:** Data curation, Methodology, Software, Writing – review & editing. **Yongxiang Yao:** Methodology, Writing – review & editing. **Zhi Zheng:** . **Yi Wan:** . **Mingtao Xiong:** Visualization, Writing – review & editing.

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.

Data availability

The datasets are linked in https://github.com/betterlll/MOSS_data.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grants 42030102, 42192583, and 42001406), and the Major special projects of Guizhou [2022] 001.

References

Aguilera, C.A., Sappa, A.D., Toledo, R., 2015. LGHD: A feature descriptor for matching across non-linear intensity variations. IEEE International Conference on Image Processing (ICIP) 2015, 178–181. https://doi.org/10.1109/ICIP.2015.7350783.

Brown, M., Szeliski, R., Winder, S., 2005. Multi-image matching using multi-scale oriented patches, 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 510-517 (511). http://doi.org/10.1109/CVPR.2005.235.

Chen, M., Li, W., Fang, T., Zhu, Q., Xu, B., Hu, H., Ge, X., 2023. An adaptive feature region-based line segment matching method for viewpoint-changed images with discontinuous parallax and poor textures. Int. J. Appl. Earth Obs. 117, 103209 https://doi.org/10.1016/j.jag.2023.103209.

Dellinger, F., Delon, J., Gousseau, Y., Michel, J., Tupin, F., 2015. SAR-SIFT: A SIFT-Like Algorithm for SAR Images. IEEE Trans. Geosci. Electron. 53, 453–466. https://doi.org/10.1109/TGRS.2014.2323552.

Dusmanu, M., Rocco, I., Pajdla, T., Pollefeys, M., Sivic, J., Torii, A., Sattler, T., 2019. D2-Net: A Trainable CNN for Joint Description and Detection of Local Features. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2019, 8084–8093. https://doi.org/10.1109/CVPR.2019.00828.

Fan, Z., Liu, Y., Liu, Y., Zhang, L., Zhang, J., Sun, Y., Ai, H., 2022b. 3MRS: An Effective Coarse-to-Fine Matching Method for Multimodal Remote Sensing Imagery. Remote Sens. https://doi.org/10.3390/rs14030478.

Fan, J., Wu, Y., Li, M., Liang, W., Cao, Y., 2018. SAR and Optical Image Registration Using Nonlinear Diffusion and Phase Congruency Structural Descriptor. IEEE Trans. Geosci. Electron. 56, 5368–5379. https://doi.org/10.1109/TGRS.2018.2815523.

Fan, J., Ye, Y., Li, J., Liu, G., Li, Y., 2022a. A Novel Multiscale Adaptive Binning Phase Congruency Feature for SAR and Optical Image Registration. IEEE Trans. Geosci. Electron. 60, 1–16. https://doi.org/10.1109/TGRS.2022.3206804.

Feng, R., Shen, H., Bai, J., Li, X., 2021. Advances and Opportunities in Remote Sensing Image Geometric Registration: A systematic review of state-of-the-art approaches and future research directions. IEEE Geosci. Remote Sens. Mag. 9 (4), 120–142. https://doi.org/10.1109/MGRS.2021.3081763.

Foroosh, H., Zerubia, J.B., Berthod, M., 2002. Extension of phase correlation to subpixel registration. IEEE Trans. Image Process. 11, 188–200. https://doi.org/10.1109/83.988953.

Gao, C., Li, W., Tao, R., Du, Q., 2022. MS-HLMO: Multiscale Histogram of Local Main Orientation for Remote Sensing Image Registration. IEEE Trans. Geosci. Remote Sens. 60, 1–14. https://doi.org/10.1109/TGRS.2022.3193109.

Gong, M., Zhao, S., Jiao, L., Tian, D., Wang, S., 2014. A Novel Coarse-to-Fine Scheme for Automatic Image Registration Based on SIFT and Mutual Information. IEEE Trans. Geosci. Electron. 52, 4328–4338. https://doi.org/10.1109/TGRS.2013.2281391.

Huang, J., Yang, F., Chai, L., 2023. Robust Registration of Multimodal Remote Sensing Images With Spectrum Congruency. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 16, 5103–5114. https://doi.org/10.1109/JSTARS.2023.3281029.

Ji, X., Yang, B., Tang, Q., Xu, W., Li, J., 2022. Feature fusion-based registration of satellite images to airborne LiDAR bathymetry in island area. Int. J. Appl. Earth Obs. 109, 102778 https://doi.org/10.1016/j.jag.2022.102778.

Ji, S., Zhang, T., Guan, Q., Li, J., 2013. Nonlinear intensity difference correlation for multi-temporal remote sensing images. Int. J. Appl. Earth Obs. 21, 436–443. https://doi.org/10.1016/j.jag.2012.06.009.

Ji, S., Zeng, C., Zhang, Y., Duan, Y., 2023. An evaluation of conventional and deep learning-based image-matching methods on diverse datasets. Photogram. Rec. 38, 137–159. https://doi.org/10.1111/phor.12445.

Li, L., Han, L., Ding, M., Cao, H., 2023b. Multimodal Image Fusion Framework for End-to-End Remote Sensing Image Registration. IEEE Trans. Geosci. Electron. 61, 1–14. https://doi.org/10.1109/TGRS.2023.3247642.

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. https://doi.org/10.1109/TIP.2019.2959244.

Li, J., Hu, Q., Zhang, Y., 2023a. Multimodal image matching: A scale-invariant algorithm and an open dataset. ISPRS J. Photogramm. Remote Sens. 204, 77–88. https://doi.org/10.1016/j.isprsjprs.2023.08.010.

Li, L., Liu, M., Ma, L., Han, L., 2022. Cross-Modal feature description for remote sensing image matching. Int. J. Appl. Earth Obs. 112, 102964 https://doi.org/10.1016/j.jag.2022.102964.

Liu, M., Zhou, G., Ma, L., Li, L., Mei, Q., 2023. SIFNet: A self-attention interaction fusion network for multisource satellite imagery template matching. Int. J. Appl. Earth Obs. 118, 103247 https://doi.org/10.1016/j.jag.2023.103247.

Lowe, D.G., 1999. Object recognition from local scale-invariant features. In: Proceedings of the Seventh IEEE International Conference on Computer Vision. https://doi.org/10.1109/ICCV.1999.790410.

Ma, W., Wen, Z., Wu, Y., Jiao, L., Gong, M., Zheng, Y., Liu, L., 2017. Remote Sensing Image Registration With Modified SIFT and Enhanced Feature Matching. IEEE Geosci. Remote Sens. Lett. 14, 3–7. https://doi.org/10.1109/LGRS.2016.2600858.

Meng, L., Zhou, J., Liu, S., Ding, L., Zhang, J., Wang, S., Lei, T., 2021. Investigation and evaluation of algorithms for unmanned aerial vehicle multispectral image registration. Int. J. Appl. Earth Obs. 102, 102403 https://doi.org/10.1016/j.jag.2021.102403.

Revaud, J., Weinzaepfel, P., Souza, C.R., Pion, N., Csurka, G., Cabon, Y., Humenberger, M., 2019. R2D2: Repeatable and Reliable Detector and Descriptor. CoRR abs/1906.06195.

Sedaghat, A., Ebadi, H., 2015a. Distinctive Order Based Self-Similarity descriptor for multi-sensor remote sensing image matching. ISPRS J. Photogramm. Remote Sens. 108, 62–71. https://doi.org/10.1016/j.isprsjprs.2015.06.003.

Sedaghat, A., Ebadi, H., 2015b. Remote Sensing Image Matching Based on Adaptive Binning SIFT Descriptor. IEEE Trans. Geosci. Electron. 53, 5283–5293. https://doi.org/10.1109/TGRS.2015.2420659.

Sedaghat, A., Mokhtarzade, M., Ebadi, H., 2011. Uniform Robust Scale-Invariant Feature Matching for Optical Remote Sensing Images. IEEE Trans. Geosci. Electron. 49, 4516–4527. https://doi.org/10.1109/TGRS.2011.2144607.

Suetake, N., Sakano, M., Uchino, E., 2008. Image super-resolution based on local self-similarity. Opt. Rev. 15, 26–30. https://doi.org/10.1007/s10043-008-0005-0.

Sun, J., Shen, Z., Wang, Y., Bao, H., Zhou, X., 2021. LoFTR: Detector-Free Local Feature Matching with Transformers. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2021, 8918–8927. https://doi.org/10.1109/CVPR46437.2021.00881.

Suri, S., Reinartz, P., 2010. Mutual-Information-Based Registration of TerraSAR-X and Ikonos Imagery in Urban Areas. IEEE Trans. Geosci. Electron. 48, 939–949. https://doi.org/10.1109/TGRS.2009.2034842.

Wu, Y., Ma, W., Gong, M., Su, L., Jiao, L., 2015. A Novel Point-Matching Algorithm Based on Fast Sample Consensus for Image Registration. IEEE Geosci. Remote Sens. Lett. 12, 43–47. https://doi.org/10.1109/LGRS.2014.2325970.

Xiang, Y., Wang, F., You, H., 2018. OS-SIFT: A Robust SIFT-Like Algorithm for High-Resolution Optical-to-SAR Image Registration in Suburban Areas. IEEE Trans. Geosci. Electron. 56, 3078–3090. https://doi.org/10.1109/TGRS.2018.2790483.

Xiang, Y., Tao, R., Wang, F., You, H., Han, B., 2020. Automatic Registration of Optical and SAR Images Via Improved Phase Congruency Model. IEEE J. Sel. Top. Appl.

Earth Obs. Remote Sens. 13, 5847–5861. https://doi.org/10.1109/JSTARS.2020.3026162.
Xiong, B., Li, W., Zhao, L., Lu, J., Zhang, X., Kuang, G., 2016. Registration for SAR and optical images based on straight line features and mutual information, 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2582-2585. http://doi.org/10.1109/IGARSS.2016.7729667.
Xiong, X., Jin, G., Xu, Q., Zhang, H., 2021. Self-Similarity Features for Multimodal Remote Sensing Image Matching. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 12440–12454. https://doi.org/10.1109/JSTARS.2021.3131489.
Xu, W., Yuan, X., Hu, Q., Li, J., 2023. SAR-optical feature matching: A large-scale patch dataset and a deep local descriptor. Int. J. Appl. Earth Obs. 122, 103433 https://doi.org/10.1016/j.jag.2023.103433.
Yao, Y., Zhang, B., Wan, Y., Zhang, Y., 2022a. MOTIF: MULTI-ORIENTATION TENSOR INDEX FEATURE DESCRIPTOR FOR SAR-OPTICAL IMAGE REGISTRATION. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLIII-B2-2022, 99-105. http://doi.org/10.5194/isprs-archives-XLIII-B2-2022-99-2022.
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. https://doi.org/10.1109/TIP.2022.3157450.
Ye, Y., Shan, J., Bruzzone, L., Shen, L., 2017a. Robust Registration of Multimodal Remote Sensing Images Based on Structural Similarity. IEEE Trans. Geosci. Electron. 55, 2941–2958. https://doi.org/10.1109/TGRS.2017.2656380.
Ye, Y., Shen, L., Hao, M., Wang, J., Xu, Z., 2017b. Robust Optical-to-SAR Image Matching Based on Shape Properties. IEEE Geosci. Remote Sens. Lett. 14, 564–568. https://doi.org/10.1109/LGRS.2017.2660067.
Ye, Y., Bruzzone, L., Shan, J., Bovolo, F., Zhu, Q., 2019. Fast and Robust Matching for Multimodal Remote Sensing Image Registration. IEEE Trans. Geosci. Electron. 57, 9059–9070. https://doi.org/10.1109/TGRS.2019.2924684.
Yi, Z., Zhiguo, C., Yang, X., 2008. Multi-spectral remote image registration based on SIFT. Electron. Lett 44, 107–108. https://doi.org/10.1049/el:20082477.
Yoo, J.-C., Han, T.H., 2009. Fast Normalized Cross-Correlation. Circuits Systems Signal Process. 28, 819–843. https://doi.org/10.1007/s00034-009-9130-7.
Zhang, X., Liao, P., Chen, G., Zhu, K., Tan, X., Wang, T., Li, X., Wu, H., 2022. Distinguishable keypoint detection and matching for optical satellite images with deep convolutional neural networks. Int. J. Appl. Earth Obs. 109, 102795 https://doi.org/10.1016/j.jag.2022.102795.
Zhang, H., Ni, W., Yan, W., Xiang, D., Wu, J., Yang, X., Bian, H., 2019. Registration of Multimodal Remote Sensing Image Based on Deep Fully Convolutional Neural Network. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12, 3028–3042. https://doi.org/10.1109/JSTARS.2019.2916560.
Zhang, Y., Wan, Y., Shi, W., Zhang, Z., Li, Y., Ji, S., Guo, H., Li, L., 2021a. Technical framework and preliminary practices of photogrammetric remote sensing intelligent processing of multi-source satellite images. Acta Geodaetica Et Cartographica Sinica 50, 1068–1083.
Zhang, M., Wang, Z., Bai, R., Jia, H., 2020. A coarse-to-fine optical and SAR remote sensing image registration algorithm. J. Geo-Inf. Sci 22, 2238–2246.
Zhang, Y., Zhang, Z., Gong, J., 2021b. Generalized photogrammetry of spaceborne, airborne and terrestrial multi-source remote sensing datasets. Acta Geodaetica Et Cartographica Sinica 50, 1–11.
Zhang, Y., Yao, Y., Wan, Y., Liu, W., Yang, W., Zheng, Z., Xiao, R., 2023. Histogram of the orientation of the weighted phase descriptor for multi-modal remote sensing image matching. ISPRS J. Photogramm. Remote Sens. 196, 1–15. https://doi.org/10.1016/j.isprsjprs.2022.12.018.
Zhu, B., Yang, C., Dai, J., Fan, J., Qin, Y., Ye, Y., 2023. R_2FD_2: Fast and Robust Matching of Multimodal Remote Sensing Images via Repeatable Feature Detector and Rotation-Invariant Feature Descriptor. IEEE Trans. Geosci. Electron. 61, 1–15. https://doi.org/10.1109/TGRS.2023.3264610.