**MS-HLMO: Multiscale Histogram of Local Main Orientation for Remote Sensing Image Registration**

Chenzhong Gao®, Wei Li®, Senior Member, IEEE, Ran Tao®, Senior Member, IEEE, and Qian Du®, Fellow, IEEE

**Abstract**—Multisource image registration is challenging due to intensity, rotation, and scale differences among the images. Considering the characteristics and differences in multisource remote sensing images, a feature-based registration algorithm named multiscale histogram of local main orientation (MS-HLMO) is proposed. Harris corner detection is first adopted to generate feature points. The HLMO feature of each Harris feature point is extracted on a partial main orientation map (PMOM) with a generalized gradient location and orientation histogram-like (GGLOH) feature descriptor, which provides high intensity, rotation, and scale invariance. The feature points are matched through a multiscale matching strategy. Comprehensive experiments on 17 multisource remote sensing scenes demonstrate that the proposed MS-HLMO and its simplified version MS-HLMO+ outperform other competitive registration algorithms in terms of effectiveness and generalization.

**Index Terms**—Histogram of local main orientation (HLMO), image registration, multimodal, multiscale, multisource, remote sensing.

**I. INTRODUCTION**

Remote sensing registration is a key preprocessing step in remote sensing applications. With Earth observation systems being developed rapidly in recent years, to achieve in-depth ground analysis, the use of multisource remote sensing images becomes popular in achieving in-depth ground analysis [1]–[5]. The accurate alignment of the multisource images is a prerequisite of subsequent applications such as data fusion, change detection, joint analysis, and other techniques.

The purpose of image registration is to spatially align multiple images for the same scene. In general, remote sensing images are spatially corrected by reference coordinates (such as fictitious graticule) or image control points. However, there are inconsistent or missing spatial references. In particular, in multisource remote sensing, the inconsistency of spatial correction between different sources is more obvious. This phenomenon can be seen from the examples shown in Fig. 1, which is about the correction of the multisensor image solely through the spatial reference information provided by the data source. It is obvious that the ground covers correspond inconsistently. In addition, for some unmanned aerial vehicle (UAV) or ground-based images, there is no spatial reference information. It is very important to correct multisource remote sensing images automatically by registration algorithms to provide accurately registered data for the same scene space.

The existing automatic image registration algorithms are generally divided into two categories, i.e., the traditional methods and the deep-learning-based methods. Among them, the deep-learning-based methods [6] are relatively advanced, but a large amount of annotated samples are needed for model training, and the model is often highly targeted. The traditional ones are systematically classified into area-based and feature-based methods [7]. The area-based, also called intensity-based methods, register images by establishing the similarity measure of the intensity values or in a transform domain [8]–[15], and optimal geometric transformation parameters between the image pairs are found through optimization [16]. Apart from efficiency and other factors, for multisource images, this type algorithm is prone to fail when multimodal differences are present.

Registration based on feature matching is a relatively mature technique, in which the transformation parameters are estimated through the coordinate correspondence of matched features. Scale-invariant feature transform (SIFT) [17]–[19] is one of the most classic, effective, and commonly used feature extraction and matching methods and became the basis of many improved algorithms [20]–[26]. SIFT uses Gaussian scale space, supported by scale-space theory [27], for feature

---

**Fig. 1.** Examples of alignment failure of multisource remote sensing images with only calibration information.
points’ extraction, which also provides scale invariance. The feature points are then described by statistical local gradient information for points’ matching. This also lays the framework for the subsequent algorithms. In addition, many feature extraction algorithms have been proposed and applied to image registration [28]–[37]. However, most algorithms are only applicable to single-modal image registration. The algorithms for multimodal image registration are more critical. Focusing on this, the multimodal image registration algorithms have been proposed. Chen et al. [38] proposed partial intensity invariant feature descriptor (PIIFD) based on the SIFT algorithm for multisource retinal image registration, which overcame the problem of intensity difference problems. PSO-SIFT [39] presented a new gradient definition and an enhanced feature matching method for the registration of remote sensing images with intensity differences. Ye et al. developed histogram of oriented phase congruency (HOPC) [40] and local HOPC (LHOPC) [41] for multimodal remote sensing registration, in which minimum moment of phase congruency (MMPC)-Lap is used for keypoints’ detection, and an extended phase congruency model is used for feature description. Radiation-variation insensitive feature transform (RIFT) [42] used a maximum index map (MIM) as the feature map, which is invariant to intensity difference. The MIM is obtained by assigning the strongest response of the log-Gabor filtering at several predetermined orientations as the maximum index at each pixel. Multiscale PIIFD (MS-PIIFD) [43] improved PIIFD by adopting Gaussian scale space, in which feature extraction and matching are completed in a multiscale strategy. It achieves scale robustness and performs well on multimodal images with scale differences.

The multimodal differences in multisource images result in multiple inconsistency of local features in the same area of image pairs, which makes it difficult to match correctly unless all inconsistencies are conquered. It is needed to handle the differences and enhance the similarity of the features. Therefore, through deep-going analysis of practical work and real data, we refine and model the differences in multisource remote sensing images as

$$I_2 = \mathcal{F}_{\text{Cut}}(\mathcal{F}_{\text{Tran}}(\mathcal{F}_{\text{Sc}}(\mathcal{F}_{\text{Rot}}(\mathcal{F}_{\text{Chg}}(I_1))))))$$  \hspace{1cm} (1)$$

where $I_1$ and $I_2$ are a multisource image pair, and $\mathcal{F}_{\text{Chg}}(\bullet)$ represents the changes in the real ground object content sensed in the images, which is often caused by multimodal differences. $\mathcal{F}_{\text{Int}}(\bullet)$ denotes the intensity transformation. $\mathcal{F}_{\text{Rot}}(\bullet)$, $\mathcal{F}_{\text{Sc}}(\bullet)$, and $\mathcal{F}_{\text{Tran}}(\bullet)$ are the spatial transformation of rotation, scale change, and translation, respectively, which are the basic operation of similarity transformation. $\mathcal{F}_{\text{Cut}}(\bullet)$ represents the spatial cutting operation, which results in images covering different ranges of vision. There may also be large perspective differences or severe local geometric distortions between images, which leads to more complex image distortion and registration modeling. Our goal is to first resolve the most common and fundamental problem of image differences. Among the above multimodal differences, $\mathcal{F}_{\text{Cut}}(\bullet)$, $\mathcal{F}_{\text{Tran}}$, and $\mathcal{F}_{\text{Chg}}(\bullet)$ have been solved by the feature-based process. Thus, the primary goal is to find a feature robust to $\mathcal{F}_{\text{Sc}}(\bullet)$, $\mathcal{F}_{\text{Rot}}(\bullet)$, and $\mathcal{F}_{\text{Int}}(\bullet)$ in multisource images.

Based on the above analysis, a novel feature-based registration framework for multisource remote sensing images is proposed through designing effective strategies for robust feature extraction that overcomes multiple image differences. It is able to automatically register remote sensing images under various conditions without human intervention. The main contributions are summarized as follows:

1) An image registration algorithm named multiscale histogram of local main orientation (MS-HLMO) is designed to cope with various multisource remote sensing images with multimodal differences. Through comprehensive experimental verification, MS-HLMO is able to effectively deal with common multisource remote sensing image registration, including multisensor, multi-temporal, and multiviewpoint image pairs with different resolutions and data size, which reflects strong robustness and generalization.

2) A local feature named histogram of local main orientation (HLMO) is proposed, in which a basic feature map called partial main orientation map (PMOM) is developed for local feature extraction, which handles multimodal differences and provides robust orientation information. In addition, HLMO uses a generalized GLOh-like (GGLOH) feature descriptor to extract local features. The proposed HLMO is invariant to intensity and rotation and has outstanding ability in robust feature extraction from multisource images.

3) To overcome the problem of scale differences in multisource images, a multiscale feature extraction and matching strategy based on Gaussian scale space is applied in MS-HLMO, which has a good effect on solving scale differences and optimizing feature matching. MS-HLMO has a complete processing flow without manual intervention and is able to directly applied to the practical multisource remote sensing joint analysis. To further research and assessment the proposed method, the code will be released on https://github.com/MrPingQi.

The rest of this article is arranged as follows. Section II describes the overall framework of the proposed MS-HLMO and each critical process in detail. Section III presents the experiments and discussion in terms of discrete and overall testing. Finally, this research is summarized in Section IV.

II. PROPOSED REGISTRATION METHOD

The framework of the proposed MS-HLMO registration algorithm is shown in Fig. 2. The input multisource image pair to be registered is preprocessed, which includes data normalization and basic denoising. Then, the preprocessed single-band images are used for feature points’ detection and feature extraction. Harris corner point detection, which contains detail treatments for multisource images, is adopted to generate feature points between the image pair for matching. The key process of the proposed HLMO feature extraction is carried out in a multiscale strategy, in which Gaussian
pyramids are built to create a scale space of the images. The HLMO feature descriptors of each Harris corner point are extracted on the PMOM of the images. The feature points between the image pair are then matched according to the descriptors, and outlier removal is carried out to remove the mismatches. The matching results in the scale space are combined through a multiscale matching strategy. Finally, the spatial transformation between the original image pair is determined by the coordinate relationship between matched feature points according to a selected transformation model.

A. Harris Feature Point Detection

Harris corner \([29]\) is one of the most stable feature points, which is slightly affected by intensity and scale difference and has high computational efficiency \([43], [44]\). It has the advantage in multisource remote sensing images with multimodal properties and large data size. Here, the similar strategy \([43]\) is used for feature points’ detection. The Harris corner response of each pixel is calculated by

\[
\text{cornerness} = \frac{\det(M)}{\text{tr}(M)} \quad (2)
\]

\[
M = \begin{bmatrix}
\sum_{w_x} G_x^2 & \sum_{w_x} G_x G_y \\
\sum_{w_y} G_x G_y & \sum_{w_y} G_y^2
\end{bmatrix} \quad (3)
\]

where \(\det(M)\) and \(\text{tr}(M)\) are the determinant and trace of \(M\), respectively. \(G_x\) and \(G_y\) are the image’s gradient along the \(x\) and \(y\) directions, respectively, and \(W_x\) is a Gaussian window with variance \(\sigma\). Pixels with strong response are considered to be feature points with distinct structure and stability between multisource images.

An important issue in practical multisource remote sensing image registration is that the data size and scale relationships between images to be registered are diverse. An example is shown in Fig. 3, which is the Harris corner detection result of an optical–infrared image pair with only threshold filtering. It is obvious that in the optical image, only a small number of feature points are distributed in the corresponding area of the infrared image, which are also nonuniform with low repeatability. Many of the existing algorithms only deal with the ideal case that the image pair has the same scale and size. This article focuses on solving several key problems at the same time, that is, two uncertain factors of image scale and size should be considered simultaneously. Local nonmaximum suppression (LNMS) is adopted to solve this problem. Since the size and scale difference in the image pairs are uncertain, in Harris corner detection, it is expected that the feature points in the image pair are distributed as uniformly as possible with the use of local non-maximum suppression (LNSM). Then, the ratio of the window size in LNMS is set depending on the ratio of the data size of the image pair

\[
\text{ratio} = \sqrt{\frac{M \times N}{m \times n}} \quad (4)
\]

where \(M, N\) and \(m, n\) are the length and width of the two images, respectively.

Consequently, the distribution of feature points is consistent with the images’ scale proportion when there are scale differences, as shown in Fig. 4(a). When there is a size difference, the feature points are far more uniformly distributed, and the repeatability is higher, as shown in Fig. 4(b).

B. Histogram of Local Main Orientation

1) Partial Main Orientation Map: The feature-point-based registration algorithms often use a local descriptor to extract the neighborhood information of the keypoints and generate their feature vectors for similarity matching. For example, SIFT \([19]\), histogram of oriented gradient (HOG) \([30]\), speeded up robust features (SURF) \([22]\), and PIIFD \([38]\) use gradient
information as the basic feature. However, the performance of these algorithms is greatly degraded when processing multisource, especially multisensor images. So, it is critical to extract invariant feature that is robust to intensity difference. For comparison purposes, the images have been registered manually, basically eliminating the spatial differences in scale, rotation, and size. The magnitude and orientation of images’ gradient are shown in Fig. 6, which observes that due to the multimodal properties of the original data are a visible–infrared image pair, which contains obvious intensity difference. For comparison purposes, the images have been registered manually, basically eliminating the spatial differences in scale, rotation, and size.

The ASG is a gradient weighting method. The elementary gradient of the image along the $x$ and $y$ directions, i.e., $G_x$ and $G_y$, are calculated as

$$
\begin{bmatrix}
G_x(x, y) \\
G_y(x, y)
\end{bmatrix} = \begin{bmatrix}
\frac{\partial}{\partial x}I(x, y) \\
\frac{\partial}{\partial y}I(x, y)
\end{bmatrix}
$$

(5)

where $I(x, y)$ represents the single-layer gray-scale image. The magnitude and orientation of its gradient, i.e., $G_p$ and $G_\phi$, are

$$
\begin{bmatrix}
G_p \\
G_\phi
\end{bmatrix} = \begin{bmatrix}
\sqrt{G_x^2 + G_y^2} \\
\arctan \frac{G_y}{G_x}
\end{bmatrix}
$$

(6)

In the ASG, a locally weighted squared gradient [45] along the $x$ and $y$ directions, $GW_{x,s}$ and $GW_{y,s}$, are obtained as

$$
\begin{bmatrix}
GW_{x,s,x} \\
GW_{y,s,y}
\end{bmatrix} = \begin{bmatrix}
\sum_{s} W_s G_{s,x}^2 - G_{s,y}^2 \\
\sum_{s} 2G_{s,x}G_{s,y}
\end{bmatrix}
$$

(7)

where $W_s$ is a Gaussian window with variance $\sigma$. Accordingly, the orientation of this gradient is

$$
GW_{x,s,\phi} = \angle \left( GW_{x,s,x}, GW_{x,s,y} \right)
$$

(8)

where $\angle(X, Y)$ is defined as

$$
\angle(X, Y) = \begin{cases}
\arctan \left( \frac{X}{Y} \right), & X \geq 0 \\
\arctan \left( \frac{X}{Y} \right) + \pi, & X < 0, Y \geq 0 \\
\arctan \left( \frac{X}{Y} \right) - \pi, & X < 0, Y < 0
\end{cases}
$$

(9)

making $GW_{x,s,\phi}$ within $(-\pi, \pi)$. According to [45], this gradient is obtained by doubling the angle of the original gradient, so the orientation of the ASG is

$$
GW_{x,\phi} = \frac{1}{2} GW_{x,s,\phi}.
$$

(10)

Compared with the classical gradient operator, ASG orientation $GW_{x,\phi} \in (-\pi/2, (\pi/2))$ reflects the weighted gradient orientation of a local region $W_s$, which is more robust and stable. In addition, the $x$ direction gradient is positive, and this characteristic meets the requirement that not affected by the reversal of gradient in intensity difference. Note that when $\sigma$ increases, the scale of ASG increases, which makes the local orientation more invariant to intensity difference and noise, but the uniqueness of local features decreases. From this, the following function is defined:

$$
G_{PMOM} = \frac{1}{2} \left( \sum_{\sigma} \sum_{W_s} G_{x,\phi}^2 - G_{y,\phi}^2, \sum_{\sigma} \sum_{W_s} 2G_{x,\phi}G_{y,\phi} \right)
$$

(11)

where a series of scale $\sigma$ are preset, the weighted responses $GW_{x,s,\phi}$ and $GW_{y,s,\phi}$ at each scale are added, and the ASG orientation is obtained. By filtering the image with (11), the PMOM is obtained, where its value reflects the overall orientation of multiple scales in each partial area of the image. Fig. 5 shows the schematic of PMOM, in which it is seen that different scales provide orientation information of different levels, and PMOM represents the effect of information combination of multiple scales.

A visualized comparison of PMOM with other feature maps of typical multisource data is shown in Fig. 6. The original data are a visible–infrared image pair, which contains obvious intensity difference. For comparison purposes, the images have been registered manually, basically eliminating the spatial differences in scale, rotation, and size. The magnitude and orientation of images’ gradient are shown in Fig. 6, which are obtained using (5) and (6). These are the basic feature information in most algorithms [19], [22], [30], [38], [43]. It is observed that due to the multimodal properties of the original image pair, these two feature maps have large differences and instability, which is the main reason for the failure of most algorithms. The MIM [42] of RIFT shown in Fig. 6 also focuses on the local orientation of the image, where the maximum index is the main orientation among several ones. Compared with the directional gradient, Gabor transformation has a more stable response, which leads to RIFT robust to...
intensity difference. However, its value will also mutate due to local changes in images, and the rotation invariance is slightly poor. The image pair’s PMOMs are shown at the bottom of Fig. 6. Compared with MIM, the proposed PMOM is not only more robust and stable between multimodal images but also continuous in value, which is conducive to achieving effective rotation invariance. In HLMO, PMOM is used as the unique feature information to extract local features of keypoints that are invariant to multimodal properties.

2) Descriptor Extraction: After determining the feature points and the specific feature for discrimination, the next step is to make use of the local feature information around each point and generate descriptors. Gradient location and orientation histogram (GLOH) has shown excellent ability through experiments [21] and has been applied in multisource remote sensing image registration [25], [39]. The original GLOH descriptor is a circular region divided by three circles, similar to that shown in Fig. 8, in which the two outer circular regions are divided into four parts, and the radius of the circular region divided is 5, 9, and 11. The partition size and the number are then improved [21], [25], [39]. However, different parameters have various effects when treating multifarious types of images. In addition, if the number of outer ring regions is too small, the character of feature points is not significant, which makes it difficult to match accurately. If it is too large, the features are unstable, and the dimension of the descriptor is too high, which increases the burden of redundant calculation. To deal with this, a generalized GLOH-like (GGLOH) descriptor is proposed, and its structure is shown in Fig. 7(a).

Let \( A^0 \) denote the central circular region, and \( A^i_j, (i = 1, 2, j = 1, \ldots, N_A) \) represents the sector subregion \( j \) in the outer ring region \( i \). Let \( N_A \) be the number of the subregions in each outer ring region, which is even, \( \theta_0 \) be the main orientation of the feature point, and \( R_0, R_1, \) and \( R_2 \) be the radii of the central and outer regions, respectively. Note that the orientations of pixels’ gradient in each region are counted as feature information, and therefore, fair use of information in each region is expected. The number of pixels in each region should be roughly the same, and the weight of the outer regions should not change due to the change in \( N_A \). So the area of each region should be the same, that is,

\[
N_A \cdot \pi R_0^2 = \pi (R_1^2 - R_0^2) = \pi (R_2^2 - R_1^2)
\]

(12)

which also fixes the relationship between \( R_0, R_1, R_2, \) and \( N_A \). When \( N_A \) is given different values, the stability and importance of each region’s feature remain the same. In HLMO, GGLOH is used to extract local features on PMOM, where the orientation values within \((-\pi/2, \pi/2)\) are uniformly quantified to \( N_O \) values, as shown in Fig. 7(b), where \( \phi_k (k = 1, 2, \ldots, N_O) \) is the angle after quantization. A histogram with \( N_O \) bins is obtained in each region.

It is simple to achieve rotation invariance of HLMO. For each keypoint, the PMOM value at its location is the main orientation, that is, the reference orientation \( \theta_0 \) of GGLOH. Then, all the PMOM values within the local area of GGLOH also take \( \theta_0 \) as the reference (0°), that is, all the angle values minus \( \theta_0 \), and those beyond \((-\pi/2, \pi/2)\) are flipped to their opposite angles.
Another key problem is that the rotation and nonlinear intensity difference may cause the jump of the main orientations of some feature points near \((-\pi/2)\) and \((\pi/2)\). An example is shown in Fig. 8. In PIIFD [38], a similar problem has been discovered and improvement has been made for SIFT. Then, a similar strategy is adopted to process GLOH-like descriptors.

![Image](Image 9-5626714)

Fig. 9. Structure of the Gaussian pyramid used in MS-HLMO.

\[
D_1 = \begin{bmatrix}
H_1^1 & H_2^1 & \cdots & H_{N_A/2}^1 \\
H_1^2 & H_2^2 & \cdots & H_{N_A/2}^2 \\
\vdots & \vdots & \ddots & \vdots \\
H_1^{N_A/2} & H_2^{N_A/2} & \cdots & H_{N_A}^{N_A/2}
\end{bmatrix}
\]

(13)

\[
D_2 = \begin{bmatrix}
H_1^{N_A/2+1} & H_2^{N_A/2+2} & \cdots & H_{N_A}^{N_A/2} \\
H_1^{N_A/2+2} & H_2^{N_A/2+4} & \cdots & H_{N_A}^{N_A/2+2} \\
\vdots & \vdots & \ddots & \vdots \\
H_1^{N_A} & H_2^{N_A} & \cdots & H_{N_A}^{N_A}
\end{bmatrix}
\]

(14)

\[
D = \left[D_1 + D_2 \right] / \left[|D_1| - |D_2| \right]
\]

(15)

where \(H_i\) is the histogram vector of gradient orientation of region \(A_i\). In this way, no matter whether the main orientation of feature points is reversed 180° or not, descriptor \(D\) is composed of the addition and subtraction of the upper and lower parts of GGLOH according to the main orientation axis, without changing the regions’ order. Finally, a descriptor vector \(D_P\) is generated for the feature point \(P\), whose dimension is \((2 \cdot N_A + 1) \cdot N_O\).

### C. Multiscale Registration Strategy

The scale difference \(F_{sc}(\bullet)\) of multisource images has a great influence on local features. Some algorithms have quantitative scale judging methods, such as SIFT [19] and LHOPC [41]. However, it is found that these methods are invalid in images with large modal differences [16], [42], [43]. The reason is that when images do not belong to the same degradation model, it is not credible to judge the scale quantitatively through local image feature information. Multisource images often have scale differences, and sometimes the scale proportion is unknown. To deal with this key problem and realize scale robustness, a multiscale feature extraction and matching strategy is designed in MS-HLMO.

Local information of feature points is extracted in the scale space of the images. Based on the scale-space theory [27], the method of building image’s Gaussian pyramids is adopted. The schematic of establishing Gaussian pyramid of the image is shown in Fig. 9. The original image is first sampled down step by step to obtain a series of images with different resolutions, that is, the first layer in each octave. Then in each octave, a series of Gaussian blurs are performed.

\[
L = G \ast I
\]

(16)

\[
G = \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(x^2+y^2)}{2\sigma^2}}
\]

(17)

where \(I\) is the original image, \(G\) is a Gaussian kernel with a standard deviation of \(\sigma\), and \(L\) is the Gaussian blur image.

#### Algorithm 1 Proposed MS-HLMO Feature Extraction

**Input:** single-band image \(I\), feature point set \(P\), total number of octaves \(N_O\) and layers \(N_OL\) in Gaussian pyramid, subregion and angle partition parameters \(N_A\), \(N_O\) in GGLOH, patch size \(S\) of HLMO.

Through downsampling and (16) and (17), the Gaussian pyramid \(G_1(O, L)\) of image \(I\) is built with \(N_O\) octaves and \(N_OL\) layers in each octave. In each layer of \(G_1(O, L)\):

1. Calculate the PMOM of this layer to get \(F_1(O, L)\) according to (5)\(11\)
2. For each feature point \(P\) in \(P\):
   - Calculate the corresponding position
   - Take the PMOM value at the position as the main direction \(\theta_0\)
   - Taking the main orientation as the reference direction \((0^\circ)\), establish a GGLOH window with size (diameter) of \(S\)
   - Statistics the PMOM value within each region of GGLOH to obtain the basic feature descriptor \(D_1(P, O, L)\) and \(D_2(P, O, L)\)
   - Obtain the descriptor \(D(P, O, L)\) of \(P\) with 15.

**Output:** feature descriptor set \(D_P(O, L)\)

After the scale space of the images is established, for each Harris corner point, the HLMO descriptor is calculated by obtaining the local information at the corresponding location of each feature point in the scale space. The proposed multiscale HLMO feature extraction method is provided in Algorithm 1, where \(O\) is the octave number in the Gaussian pyramid, and \(L\) is the layer number. The algorithm outputs the feature point descriptor set \(D_P(O, L)\), which contains \((2 \cdot N_A + 1) \cdot N_O\)-dimensional vectors for each feature point at each scale.

The next step is to match the feature point sets of the image pair according to the descriptor sets. The process of the multiscale feature matching is provided in Algorithm 2. In the process, each scale is matched in turn. Then the matching results are merged step by step while the outlier removal is carried out to realize the optimization of matching points. The outlier removal method is optional to be any of the existing effective ones, which does not affect the overall registration process. The final matching results are used to determine the spatial transformation between images. The most critical is to combine all the matching results of feature points and remove the outliers, so as to maximize the correct matches of all the scales. Fig. 10 shows this process visually.

The Gaussian scale space is the classic and widely used one, the same as the one in SIFT. The biggest difference is that
Algorithm 2 Proposed MS-HLMO Feature Matching

**Input:** feature point set of the image pair $P_{I1}$, $P_{I2}$, feature descriptor set of the image pair $D_{P_{I1}}(O_1, L_1)$, $D_{P_{I2}}(O_2, L_2)$

Take each layer of $D_{P_{I1}}(O_1, L_1)$:
- Take each layer of $D_{P_{I2}}(O_2, L_2)$:
  - Match $P_{I1}$ and $P_{I2}$ using Euclidean distance of the descriptors
  - Remove outliers, producing the matching result of a single scale $M(O_1, O_2, L_1, L_2)$

The matching results of all layers in each octave of the scale space are union and then optimized with outlier removal, producing the final matching result $M_{OL}(P_{I1}, P_{I2})$

**Output:** feature points’ matching set $M_{OL}(P_{I1}, P_{I2})$

---

**III. EXPERIMENTAL RESULTS AND DISCUSSION**

To validate the registration performance of the proposed framework, a set of typical multisource remote sensing images are carefully selected, and six effective registration algorithms are used for comparison, including SIFT [19], SAR-SIFT [39], PSO-SIFT [39], PIIFD [38], MS-PIIFD [43], and RIFT+ [42], all of which contain processes dealing with multimodal properties. The experiments are implemented using MATLAB2021a on a platform with Advanced Micro Devices, Inc. (AMD)-Ryzen-7-5800X 3.80-GHz CPU, and 32-GB RAM.

**A. Experimental Data and Preprocessing**

The data used in experiments are multisource remote sensing images of 17 scenes, labeled as $\sim q$. These images include hyperspectral image (HSI), multispectral (MSI), synthetic aperture radar (SAR), visible images, infrared images, depth maps such as light detection and ranging (LiDAR)-derived digital surface model (DSM), and even artificial maps. They contain day–night image pair and images from different years. The scenes cover spaceborne, airborne, and ground-based remote sensing data. There are various intensity and geometric distortions with the images. Table I shows the types and sizes of each set of data, as well as whether they have significant intensity difference, rotation, or scale difference. The specific form of the scenes is shown in Figs. 16 and 17. Except for artificial maps, these are all real data without manual manipulation.

| Scene | Type   | Size       | Intensity Difference | Rotation | Scale Difference |
|-------|--------|------------|----------------------|----------|------------------|
| a     | MSI    | 3555 x 4026| ✓                    | ✓        | ✓                |
| a     | HSI    | 1185 x 1342| ✓                    | ✓        | ✓                |
| a     | SAR    | 3555 x 4026| ✓                    | ✓        | ✓                |
| b     | MSI    | 3529 x 1756| ✓                    | ✓        | ✓                |
| b     | HSI    | 1175 x 585 | ✓                    | ✓        | ✓                |
| c     | MSI    | 2461 x 4139| ✓                    | ✓        | ✓                |
| c     | HSI    | 2271 x 3152| ✓                    | ✓        | ✓                |
| d     | PAN    | 1000 x 1000| ✓                    | ✓        | ✓                |
| d     | HSI    | 550 x 450  | ✓                    | ✓        | ✓                |
| e     | visible| 614 x 614  | ✓                    | ✓        | ✓                |
| e     | infrared| 611 x 611 | ✓                    | ✓        | ✓                |
| f     | visible| 617 x 593  | ✓                    | ✓        | ✓                |
| g     | visible| 1920 x 1080| ✓                    | ✓        | ✓                |
| g     | infrared| 667 x 504 | ✓                    | ✓        | ✓                |
| h     | visible| 4056 x 3040| ✓                    | ✓        | ✓                |
| h     | infrared| 640 x 512 | ✓                    | ✓        | ✓                |
| i     | visible| 633 x 460  | ✓                    | ✓        | ✓                |
| i     | infrared| 670 x 508 | ✓                    | ✓        | ✓                |
| j     | HSI    | 349 x 1905 | ✓                    | ✓        | ✓                |
| j     | LiDAR  | 349 x 1905 | ✓                    | ✓        | ✓                |
| k     | HSI    | 166 x 600  | ✓                    | ✓        | ✓                |
| k     | LiDAR  | 166 x 600  | ✓                    | ✓        | ✓                |
| l     | HSI    | 325 x 220  | ✓                    | ✓        | ✓                |
| l     | LiDAR  | 325 x 220  | ✓                    | ✓        | ✓                |
| m     | visible| 500 x 500  | ✓                    | ✓        | ✓                |
| m     | depth  | 500 x 500  | ✓                    | ✓        | ✓                |
| n     | visible| 500 x 500  | ✓                    | ✓        | ✓                |
| n     | map    | 500 x 500  | ✓                    | ✓        | ✓                |
| o     | visible| 500 x 500  | ✓                    | ✓        | ✓                |
| o     | visible| 500 x 500  | ✓                    | ✓        | ✓                |
| p     | SAR    | 600 x 500  | ✓                    | ✓        | ✓                |
| p     | SAR    | 600 x 500  | ✓                    | ✓        | ✓                |
| q     | visible| 300 x 300  | ✓                    | ✓        | ✓                |
| q     | visible| 300 x 300  | ✓                    | ✓        | ✓                |
registration, the images may have multimodal properties. Due to its enhancing pixel-level features, data dimension reduction could break the stable and unified spatial structures of each type of ground covers in the image and increase the structural differences between images. Then, registration would be more difficult as \( F_{CG}(\bullet) \) is exacerbated, which is undesired. Note that HLMO is not subject to the changes in image’s grayscale, so it is optional to select any one of the bands as the input, or just simply add up all the bands, which has little effect on the registration result.

### B. Experiment Settings

The proposed MS-HLMO contains seven main parameters that greatly affect the registration, which are the number of feature points, the scale of PMOM, \( R_2 \), \( N_A \), and \( N_0 \) of GGLOH, and \( N_{GO} \) and \( N_{GL} \) of the Gaussian pyramid. In general, the lower the number of subregions \( N_A \) and angles \( N_0 \), the higher the robustness of features, but the lower the separability of the feature points. Conversely, the more distinct the features, the lower the stability, which also increases the computational burden. In the experiment, both \( N_A \) and \( N_0 \) are set to 12, and \( R_2 \) is set to 48, which has the best performance and stability in the majority of multisource data. Then, \( R_1 \) and \( R_0 \) in GGLOH are determined by 12, and the patchsize of HLMO is \( 2 \times R_2 = 96 \). Through a simple set of tests, the threshold of the Harris corner points’ number is set at 2000, and \( N_{GO} \) and \( N_{GL} \) are set to 3 and 4, respectively, which handles the registration effectively enough with high efficiency.

In the calculation of PMOM, ten scales are set for fusion, of which the radii are evenly spaced between \( R_0 \) and \( R_2 \), and \( \sigma \) is \( 1/3 \) of each radius. In this way, when GGLOH is applied to a feature point, the main orientation of the patch, that is, the PMOM value of the center point, is only determined by the pixels’ values within the patch, excluding information beyond the GGLOH region. Ten scales are sufficient to fully obtain the local multiscale weighted main orientation information in various multisource images, which balances the robustness and uniqueness with a low computational burden.

The number of correct matches (NCM) is used as the main evaluation metric of feature matching. The correct matches refer to the matched points that are basically at the same position in the real space

\[
NCM = \left| \left\{ \| p^1_i - Tp^2_i \| < 5 \right\}_{i=1}^{N_m} \right|
\]

where \( p^1_i \) and \( p^2_i \) are the coordinates of a pair of matched points, \( T \) is the ground-truth spatial transformation between the image pair, and \( N_m \) is the total number of all the matches. NCM is positively correlated with the accuracy of the transformation model. Note that if the NCM of an image pair is less than 3, the parameters of the spatial transformation model cannot be solved, then the registration process is considered to fail.

To ensure the fairness of the comparison, except for feature point detection and descriptor extraction methods, all the processes are the same, including preprocessing, feature point matching, outlier removal, transformation, etc. The feature points used in PIIFD, MS-PIIFD, and RIFT\(^+\) are the same as the proposed algorithm with the strategy in [43]. Different algorithms use various feature matching and outlier removal methods in their original process, which interferes with the comparison of registration effects. In this experiment, Euclidean distance measurement is used for all feature matching, and fast sample consensus (FSC) is used for outlier removal for a fair comparison. As FSC has random property, the result fluctuates in the experiment, which is not very large but does have an impact. So each test is repeated 50 times and the result with the most NCM is taken.

### C. Invariance Tests

The crux of multisource image registration based on feature matching lies in various multimodal differences, and the robustness of local features to these differences is crucial. According to the analysis, intensity difference \( F_{Int}(\bullet) \), rotation \( F_{Rot}(\bullet) \), and scale difference \( F_{Sc}(\bullet) \) are the most common problems. So we design experiments to singly test the invariance of MS-HLMO from these three aspects, so as to test whether it deals with each multimodal problem between multisource images.

1) **Intensity Invariance**: To independently test the intensity invariance of MS-HLMO, the influence of other factors is excluded. Several image pairs are selected and manually registered in advance, providing image pairs of the same size that basically eliminated the differences in scale and rotation as the input image of registration. The NCMs of the feature points matching results are listed in Table II, and the corresponding visualization of each group is shown in Fig. 11. MS-HLMO\(^+\) represents the proposed algorithm without the operation of rotation invariance, that is, the main orientation is not assigned to each feature point, and the reference orientation \( \theta_0 \) in feature description is set to 0. In this way, the robustness to intensity difference is evaluated more accurately, and the influence of descriptor rotation and PMOM value changes is excluded. As RIFT is an effective feature focusing on intensity difference, it is chosen for comparison.

The eight scenes basically cover common remote sensing image types, including various optical images, SAR, depth map, and artificial map. From the numerical results, each group has obtained sufficient correct matches, especially in scenes (a1) and (d), about 1400 of 2000 detected Harris corner points have been successfully matched. In most scenes, the NCMs of the MS-HLMO are far more than RIFT\(^+\), reflecting stronger robustness to intensity difference. In scene (j), the MS-HLMO with rotation invariance obtained fewer matches than RIFT\(^+\). Through analysis, this is because scene (j) is an

| Method       | Scene | \( a_1 \) | \( a_2 \) | d | j | g | m | \( a_4 \) | p |
|--------------|-------|--------|--------|---|---|---|---|--------|---|
| RIFT\(^+\)   |       | 1032   | 385    | 393| 176| 104| 189   | 33    | 92 |
| MS-HLMO      |       | 1417   | 620    | 1321| 149| 280| 374   | 220   | 575 |
| MS-HLMO\(^+\) |       | 1434   | 1017   | 1454| 771| 351| 846   | 386   | 868 |
Fig. 11. Feature matching results with only intensity difference by MS-HLMO and RIFT. (a) RIFT⁺. (b) MS-HLMO. (c) MS-HLMO⁺.

Fig. 12. NCMs of MS-HLMO on different types of multisource remote sensing scenes as the rotation angles from 0° to 359°. (a) MSI–HSI. (b) MSI–SAR. (c) MSI–map. (d) panchromatic film (PAN)–HSI. (e) Visible–infrared. (f) HSI–LiDAR. (g) Visible–depth. (h) SAR–SAR.

HSI–LiDAR image pair with large spatial size and complex content, where the modal difference between them is large. Thus, the stability of the feature points’ main orientation is not high. The inconsistent reference direction leads to large differences in some feature points’ descriptors, while RIFT⁺ does not have influence caused by the feature point’s orientation. However, without rotation interference, MS-HLMO⁺ obtains a significant NCM, more than four times to RIFT⁺. The experiment shows that MS-HLMO with PMOM as its basic feature has strong intensity invariance and is very effective for multimodal image registration.

2) Rotation Invariance: To test rotation invariance, for each one of the above image pairs without scale and rotation differences, one of them is fixed and the other is rotated. The rotation angles are from 0° to 359° with an interval of 1°, from which a total of 360 image pairs of each scene are obtained. Then, the registration algorithms are performed on each image pair, and the corresponding NCMs are plotted in Fig. 12. The samples of feature points’ matching results under different rotation angles are shown in Fig. 13.

It is observed that the NCMs of the eight scenes vary as the rotation angles change, but are all greater than 100. None of the image pairs fails to be registered at any angles, indicating that the proposed MS-HLMO has rotation invariance over the entire angle interval of 0°–360°. In most scenes, as the images rotate, the NCMs fluctuate with a cycle of 90°, and the reason of this phenomenon is that the image rotation involves data interpolation as a digital image is stored a matrix. In other words, the change in the local gradient orientation is not linear, which is inevitable in digital image processing. As a result, the main orientation of each feature point is not completely stable with a certain offset from the image rotation angle. Relatively, the rotations of 0°, 90°, 180°, and 270° do not change the data value, and the main orientation of each feature point will not shift, resulting in the four peaks in the curve. Theoretically, the four intervals evenly divided by 90° from 0° to 360° are completely consistent, that is, the descriptors of a feature point are completely the same when rotated by 90°, 180°, and 270° which is confirmed by the experimental results. MS-HLMO has a strong rotation invariance and fully copes with image rotation from various angles. In addition, rotation invariance is not restricted by the types of imaging sensors.

3) Scale Invariance: To evaluate the scale invariance of MS-HLMO, the same strategy is used by fixing one image
TABLE III
NCMs OF THE 17 REMOTE SENSING SCENES’ COMPARISON BY EIGHT REGISTRATION METHODS

| Method      | a1 | a2 | a3 | a4 | b  | c  | d  | e  | f  | g  | h  | i  | j  | k  | l  | m  | n  | o  | p  | q  |
|-------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| SIFT        | 794| 30 | 25 |    | 145| 86 |    |    |    |    |    |    |    |    |    |    |    |    | 15 | 14 | 256|
| SAR-SIFT    | 16 | 6  |    |    | 66 | 4  |    |    |    |    |    |    |    |    |    |    |    |    | 9  | 36 | 262|
| PSO-SIFT    | 3  |    |    |    |    |    |    | 171| 13 |    |    |    | 9  | 6  | 10 |    |    |    | 19 | 24 | 259|
| PIIFD       |    |    |    | 53 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    | 493|
| MS-PIIFD    | 1105| 481| 444| 145| 999| 460| 78 |    |    |    |    |    | 25 | 5  | 339| 30 | 74 | 165| 322| 860|
| RIFT+       | 21 | 34 |    |    |    |    |    |    |    | 39 | 176| 109| 188| 189| 96 | 179|    |    |    |    |
| MS-HLMO     | 1126| 112| 865| 119| 603| 394| 1030| 689| 161| 130| 231| 79 | 149| 511| 420| 374| 161| 97 | 432| 883|
| MS-HLMO+    | 1194| 321| 996| 149| 984| 564| 1163| 693| 162| 337| 145| 761| 789| 861| 846| 273| 451|    |    |    |

Fig. 14. NCMs of MS-HLMO and MS-HLMO+ in scene (m) with the scaling ratios from 1 to 2.

and scaling the other. The scaling ratios are from 1 to 2 with an interval of 0.1, providing a total of 11 image pairs in each scene. Examples of the feature points’ matching results of MS-HLMO and MS-HLMO+ are shown in Fig. 14-15, respectively.

Even if there are scale differences between images with integer or noninteger proportions, the algorithm remains effective and obtains accurate matching. By observing the NCM curve, with the gradual change in the scale ratio from 1 to 2, the number of matching points gradually decreases and reaches the lowest when the scale ratio was 1.5, and then gradually increases. This is because, in the experiment, the downsampling ratio in building Gaussian scale space is set as 2. The interference caused by scale difference is well-resisted when the sampling scale is an integer multiple of 2. In other scale proportions, a significant difference in the receptive field in ground covers of the feature descriptor is caused between the image pair. So the local windows used for feature extraction between the two images are different in the receptive field. In practice, the sampling step in Gaussian pyramid is further refined to achieve better scale invariance.

This experiment demonstrates that MS-HLMO features are also robust to a certain extent when images are at scale proportion between two octaves in scale space. The proposed MS-HLMO has good robustness to scale difference in multisource images, which plays an extremely positive role in the actual multisource remote sensing image registration.

D. Registration Performance

In this section, registration is performed directly on each original data of the 17 scenes, to evaluate the overall registration performance. The NCMs of the eight registration methods with all the 17 scenes are listed in Table III. Fig. 16 shows the corresponding visualization of feature point matching results of each scene. Due to space limitation, the one with the better result between MS-HLMO and MS-HLMO+ is selected to display in each group.

In most scenes, MS-HLMO obtains the most NCM, and basically far more than other algorithms. In scenes (j) and (o), RIFT+ has more correct matches. The cases in the two scenes have already been discussed in the former section of intensity invariance test. MS-PIIFD obtains almost the same NCM as MS-HLMO does in scene (n). In this optical–map image pair, due to the simple content and stable obvious contour in the image, the advantage of MS-HLMO is not well-demonstrated. However, the proposed MS-HLMO still obtains a considerable NCM that are sufficient for accurate registration, and without the interference of orientation, MS-HLMO+ obtains far more matches.

SIFT, SAR-SIFT, PSO-SIFT, and PIIFD fail in registration in most multisource scenes, which cannot well cope with multisource remote sensing data images in various situations. Among the compared methods, MS-PIIFD is available for most scenes. Note that some of the test data selected in this article are also used in [43] for MS-PIIFD, and the registration is successful when the number of Harris corners is set to 1000. However, when the number is increased to 2000 in this experiment, the effect of MS-PIIFD is greatly reduced and even fails. This is because the proportion of correct matching of feature points obtained by MS-PIIFD before going through FSC is low, and more outliers may be introduced.
after increasing the number of feature points, which makes it difficult to calculate a consistent transformation model and leads to performance degradation. It indicates that although MS-PIIFD is able to handle multiscale and multimodal image registration, its feature robustness and stability are limited. Under the same conditions, MS-HLMO obtains a large NCM, which indicates that its feature robustness and stability are relatively high. The RIFT+ works well in scenes with only intensity difference, but the NCMs are relatively low, and it cannot cope with the images containing rotation and scale differences.

The last scene is a single-source image with only rotation. This scene proves that MS-HLMO can not only deal with multisource images effectively but also has a good effect on single-source images, which still performs better than other algorithms. Among all the methods, MS-HLMO is the only algorithm that does not fail in any scenes and obtains sufficient feature matches. In the scenes without obvious image rotation difference, MS-HLMO+ achieves the most feature point matching, far outperforming other algorithms. Through the above experiments, the ability of MS-HLMO is verified from a detailed and comprehensive perspective, and it can effectively handle the registration task of multisource remote sensing images. Compared with the results of the existing effective algorithms, the proposed MS-HLMO has much more significant effect than other algorithms in most cases. MS-HLMO+ better deals with the situation without obvious rotation difference, which is also the majority of practical multisource image registration tasks. MS-HLMO and its simplified version MS-HLMO+ have obvious advantages in robust feature extraction and matching.

Fig. 17 illustrates the registration results of the 17 scenes, in which the transformation parameters are obtained based on the feature matching shown in Fig. 16. For each image pair, the one with the higher resolution is taken as the reference, and the other is aligned with it through spatial affine transformation and data interpolation. Image pairs with large spatial offset are shown in a fusion form, and the areas with prominent
structures in the overlapping parts are selected and shown in an alternation form. All the operations are performed autonomously without human intervention.

It is observed that most of the image pairs are well-registered, and the outlines and textures of the ground covers stay continuous and consistent. The deviations are mostly within one pixel. The only obvious deviations are found in some areas of scenes (g) and (i) shown in Fig. 17. This phenomenon does not result from inaccurate feature extraction or matching but is caused by the characteristics of the data. Scenes (g) and (i) are ground-based images with relatively close imaging distance, in which there are obvious viewpoint differences. In addition, nonrigid geometric distortions exist in the infrared images, which may be caused by the imaging capability of the sensors. Due to the above reasons, more complex spatial differences exist between the image pairs, which is difficult to be solved with only an affine model globally transforming the images. In practice, this problem can be solved using more sophisticated nonlinear transformation models, or by dividing the image into blocks for adaptive local registration. Apart from that, most areas are well-registered, where the objects are accurately aligned.

E. Ablation Experiment

In the former experiment, the proposed MS-HLMO outperforms other algorithms in multisource image registration. However, the feature-based methods are composed of multistep. For example, the proposed algorithm is regarded as Harris–PMOM–GGLOH–MsS (multiscale strategy), and the final registration result is obtained by the interaction and complementarity of all the steps. Therefore, a set of ablation experiments are performed to more clearly show the validity of the proposed feature extraction methods of PMOM and GGLOH. The experiments are subdivided into two groups,
namely, registration tests with rotation invariance and without rotation invariance, for a fairer comparison. NCMs of all the matching results are listed in Table IV and Table V, where SIFT, GLOH, and LogPolar are the nonrotation invariance SIFT [19], GLOH [21], and LogPolar [39] descriptors. The same Harris corner detection and multiscale strategy (MsS) used in MS-HLMO are fixed for the algorithms including SIFT in Table III in the last section, it is found that the proposed method. Meanwhile, compared with the results of SIFT [19], GLOH [21], and LogPolar [39] descriptors. The same Harris corner detection and multiscale strategy (MsS) used in MS-HLMO are fixed for the algorithms including SIFT to independently test the effects of features.

In the case of rotation invariance, PMOM–GLOH (MS-HLMO) is the only valid one across all the scenes, whose NCMs are the highest in most scenes. In the absence of rotation invariance, PMOM–GLOH+ (MS-HLMO+) also has the highest NCMs in most scenes. Compared with the original gradient, the use of PMOM eliminates the failure cases and improves NCMs in most scenes. It is proven that PMOM has better similarity and stability between multisource images. Compared with GLOH and LogPolar, GLOH significantly improves NCMs in each group, proving the effectiveness of this feature descriptor. In addition, the comparison with Harris–SIFT–MsS shows that the NCMs are high for MS-HLMO which is not solely because more feature points are detected by Harris corner detection, which further highlights the importance of robust feature extraction in the proposed method. Meanwhile, compared with the results of SIFT in Table III in the last section, it is found that the proposed MsS greatly improves SIFT’s ability to deal with scale differences and NCM.

Through ablation experiments, although PMOM–GLOH is not the most effective in every scene, its robustness and generalization are fully proved. All the proposed processes are independently effective.

### IV. Conclusion

In this article, an image registration algorithm has been proposed for multisource remote sensing image registration. The proposed MS-HLMO uses PMOM based on ASG as the basic feature map, a generalized GLOH-like (GGLOH) descriptor structure, and multiscale feature extraction and matching strategy, which has the benefit of high robustness to multimodal properties and effectively deals with the key problems of intensity distortion, rotation, and scale differences in registration of multisource images. Comprehensive experiments on 17 datasets of multisource remote sensing scenes reveal the effectiveness and generalization of the proposed MS-HLMO and its simplified version MS-HLMO+ compared with other state-of-the-art registration methods.

### Table IV

| Harris-Method-MsS | a1 | a2 | a3 | a4 | b | c | d | f | g | h | i | j | k | l | m | n | o | p | q |
|-------------------|----|----|----|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SIFT              | 757| 53 | 187| 453| 413| 715|    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Gradient- GLOH    | 712| 24 | 270| 7 | 289 | 53 | 642| 150| 177 |    |    |    |    |    |    |    |    |    |    |    |    |
| PMOM - GLOH       | 592|    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| PMOM - LogPolar   | 735| 33 | 441| 7 | 402 | 333| 826| 71 |    |    |    |    |    |    |    |    |    |    |    |    |    |
| PMOM - GGLOH      | 1126| 112| 865| 119| 605 | 394| 1030| 689| 161 |    |    |    |    |    |    |    |    |    |    |    |    |

### Table V

| Harris-Method-MsS | a1 | a2 | a3 | a4 | b | c | d | f | g | h | i | j | k | l | m | n | o | p | q |
|-------------------|----|----|----|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SIFT+             | 785| 86 | 260|    |    |    |    | 595| 506| 828|    |    |    |    |    |    |    |    |    |    |    |
| Gradient+ GLOH    | 757| 108| 394| 41 | 603 | 203| 915 | 333| 237| 229| 66 | 464| 821| 775| 550| 60 | 25 |    |    |
| PMOM + GLOH       | 801| 14 | 512|    |    |    |    | 218| 108| 830|    |    |    |    |    |    |    |    |    |    |    |
| PMOM + LogPolar   | 752| 131| 439| 91 | 533| 510 | 935 | 139| 224| 77 | 449 | 471| 625| 517| 113| 239|    |    |    |    |
| PMOM + GGLOH      | 1194| 321| 996| 149| 984| 564| 1163| 693| 162| 337| 145| 761| 789| 861| 846| 273| 451|    |    |

### References

[1] M. Zhang, W. Li, Q. Du, L. Gao, and B. Zhang, “Feature extraction for classification of hyperspectral and LiDAR data using patch-to-patch CNN,” IEEE Trans. Cybern., vol. 50, no. 1, pp. 100–111, Jan. 2020.

[2] X. Zhao et al., “Joint classification of hyperspectral and LiDAR data using hierarchical random walk and deep CNN architecture,” IEEE Trans. Geosci. Remote Sens., vol. 58, no. 10, pp. 7355–7370, Oct. 2020.

[3] C. Liu, R. Tao, W. Li, M. Zhang, W. Sun, and Q. Du, “Joint classification of hyperspectral and multispectral images for mapping coastal wetlands,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 14, pp. 982–996, 2020.

[4] M. Zhang, W. Li, R. Tao, and S. Wang, “Transfer learning for optical and SAR data correspondence identification with limited training labels,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 14, pp. 1545–1557, 2020.

[5] Y. Gao et al., “Hyperspectral and multispectral classification for coastal wetland using depthwise feature interaction network,” IEEE Trans. Geosci. Remote Sens., vol. 60, no. 1–15, 2021.

[6] G. Litjens et al., “A survey on deep learning in medical image analysis,” Med. Image Anal., vol. 42, pp. 60–88, Dec. 2017.

[7] B. Zitová and J. Flusser, “Image registration methods: A survey,” Image Vis. Comput., vol. 21, no. 11, pp. 977–1000, Oct. 2003.

[8] W. K. Pratt, “Correlation techniques of image registration,” IEEE Trans. Aerosp. Electron. Syst., vol. AES-10, no. 3, pp. 353–358, May 1974.

[9] E. De Castro and C. Morandi, “Registration of translated and rotated images using finite Fourier transforms,” IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-9, no. 5, pp. 700–703, Sep. 1987.

[10] P. Viola and W. M. Wells III, “Alignment by maximization of mutual information,” Int. J. Comput. Vis., vol. 24, no. 2, pp. 137–154, Sep. 1997.

[11] J. Le Moigne, W. J. Campbell, and R. P. Cromp, “An automated parallel image registration technique based on the correlation of wavelet features,” IEEE Trans. Geosci. Remote Sens., vol. 40, no. 8, pp. 1849–1864, Aug. 2002.

[12] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, “Mutual-information-based registration of medical images: A survey,” IEEE Trans. Med. Imag., vol. 22, no. 8, pp. 986–1004, Aug. 2003.

[13] A. Mahmoud and S. Khan, “Correlation-coefficient-based fast template matching through partial elimination,” IEEE Trans. Image Process., vol. 21, no. 4, pp. 2099–2108, Apr. 2011.
