

Article

Not peer-reviewed version

Multi-Modal Remote Sensing Image Registration Using Curvature Scale Space Contour Point Features

Jianhua Zhu , [Changjiang Liu](#) ^{*} , Danling Liang

Posted Date: 2 April 2026

doi: 10.20944/preprints202604.0149.v1

Keywords: contour point features; remote sensing; image registration; gradient definition



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Multi-Modal Remote Sensing Image Registration Using Curvature Scale Space Contour Point Features

Jianhua Zhu, Changjiang Liu * and Danling Liang

School of Mathematics and Statistics, Sichuan University of Science and Engineering, Zigong, 643000, China

* Correspondence: liuchangjiang@189.cn

Abstract

Multi-modal remote sensing image registration is a challenging task due to differences in resolution, viewpoint, and intensity, which often leads to inaccurate and time-consuming results with existing algorithms. To address these issues, we propose an algorithm based on Curvature Scale Space Contour Point Features (CSSCPF). Our approach combines multi-scale Sobel edge detection, dominant direction determination, an improved curvature scale space corner detector, a new gradient definition, and enhanced SIFT descriptors. Test results on publicly available datasets show that our algorithm outperforms existing methods in overall performance. Our code will be released at <https://github.com/JianhuaZhu-IR>.

Keywords: contour point features; remote sensing; image registration; gradient definition

1. Introduction

Multi-modal remote sensing image registration has become a key area in computer vision and image processing, as single-modal images are insufficient for comprehensive applications [1]. Captured by various sensors with distinct imaging principles, multi-modal images offer unique information that enhances earth observation. The goal is to accurately align images from different times, cameras, or perspectives [2].

Multi-modal remote sensing image registration algorithms are mainly classified into region-based and feature-based methods [3]. Region-based algorithms use image intensity information and optimization to align regions by minimizing a cost function. Feature-based algorithms, which better handle geometric distortions, extract and match features, then select a transformation model based on geometric relationships [4]. Techniques like Scale-Invariant Feature Transform (SIFT) [5] address key multi-modal registration challenges.

Gao et al. [6] proposed the Multi-Scale Partial Intensity Invariant Feature Descriptor (MS-PIIFD) algorithm to address multi-source remote sensing image differences, showing good registration accuracy but being time-consuming and producing fewer matching point pairs, sometimes failing to complete registration. Gao et al. [7] developed the Multi-Scale Histogram of Local Main Orientations (MS-HLMO) algorithm, which handles intensity, scale, and rotation differences, but is also very slow. Li et al. [8] introduced the Radiation-Invariant Feature Transform (RIFT) algorithm to address Non-linear Radiometric Distortion (NRD), offering reliable feature matching but with lower accuracy and longer processing time than MS-PIIFD. Zhu et al. [9] discovered that existing image registering algorithms for infrared and visible images of power equipment suffer from low accuracy and long processing times. To address these issues, they proposed an image registering algorithm based on Large-Gap Fracture Contours (LGFC). This algorithm has shown success in the registering of power equipment images. However, its applicability to multi-modal remote sensing images is limited because remote sensing images generally lack rich LGFC information, which leads to poor registering performance in such images.

Öfverstedt et al. [10] found that intensity-based image registration relies on similarity metrics, which are crucial for robustness and accuracy. They proposed an affine registration framework combining intensity and spatial information using symmetric, non-intensity interpolation. Jiang et al. [11] addressed challenges in multi-modal power equipment images with their Contour Angle Orientation (CAO) method and Coarse-to-Fine (C2F) algorithm (CAO-C2F). Although effective, it still faces issues with accuracy and time consumption.

In response to the above analysis, this paper develops a multi-modal remote sensing image registering algorithm based on Curvature Scale Space (CSS) point features. The main contributions of this paper are as follows.

- (1) The number of feature points is crucial for image registration quality. To ensure an adequate number of feature points, we have enhanced the CSS corner detection algorithm. Since CSS extracts feature points based on contours, and the number of edges correlates with contour quantity, maintaining a sufficient number of edges is vital. To address this, we propose a multi-scale Sobel edge detection algorithm.
- (2) Given the significant differences in intensity, resolution, and viewpoint between multi-modal remote sensing image pairs, we propose a new gradient definition and a method to determine the dominant direction of feature points for rotation invariance. This gradient definition is applied to SIFT descriptors, with segmented normalization to enhance the similarity between feature point descriptors.

The following sections are structured as follows: Section 2 details the proposed registering algorithm. Section 3 conducts an extensive experimental analysis of the algorithm. Lastly, Section 4 provides the conclusion, outlining the key findings and contributions.

2. Proposed Image Registering Algorithm

The study begins by applying the proposed multi-scale Sobel edge detection algorithm to extract image edges, followed by contour tracking to outline these edges. The improved CSS algorithm detects and extracts feature points, with their dominant direction determined for rotation invariance. Enhanced SIFT descriptors are then used to characterize the feature points for image registration, ensuring alignment between the two images.

2.1. Edge Detection

The gradient components in the x and y directions can be calculated using the convolution operation as follows:

$$G_x = S * I(x, y), \quad G_y = S^T * I(x, y). \quad (1)$$

$$S = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix},$$

where $*$ denotes the convolution operation, S represents the convolution kernel and $I(x, y)$ denotes the multi-modal remote sensing image.

The gradient magnitude image G can then be obtained using the formula:

$$G = \sqrt{G_x^2 + G_y^2}. \quad (2)$$

We introduce a multi-scale Sobel edge detection approach. Initially, the input color image undergoes a transformation into a grayscale image. Subsequently, we apply a Gaussian filter with a scale parameter of σ_0 to the grayscale image, yielding the first layer of the scale space, denoted as $L_1(x, y, \sigma_0)$. Following this, $L_2(x, y, \sigma_1)$ is derived by convolving a Gaussian filter with a scale of σ_1 over $L_1(x, y, \sigma_0)$, and $L_3(x, y, \sigma_2)$ is generated by convolving a Gaussian filter with a scale of σ_2

over $L_2(x, y, \sigma_1)$. This iterative process persists until a set of N layers, each of identical dimensions, is established, constituting the scale space. The Gaussian filter function and the relationship between consecutive layers within the scale space are depicted as follows:

$$\begin{cases} g_k(x, y, \sigma_{k-1}) = \frac{1}{2\pi\sigma_{k-1}^2} e^{-\frac{x^2+y^2}{2\sigma_{k-1}^2}}, \\ L_k(x, y, \sigma_{k-1}) = g_k(x, y, \sigma_{k-1}) * I(x, y), k = 1, 2, \dots, N, \end{cases} \quad (3)$$

where $L_k(x, y, \sigma_{k-1})$ signifies the image obtained through Gaussian filtering.

The selection of an appropriate value for σ_{k-1} is crucial in this process. Increasing σ_{k-1} leads to more blurring, resulting in the loss of fine details and nuances. On the other hand, using a smaller σ_{k-1} value produces a clearer image with more details.

To ensure the required relationship $\sigma_{k-1} = f(k)$ with $f'(k) < 0$, we can assume a function $f(k) = \sigma_0 k^u$. Calculating the first derivative of $f(k)$ yields $f'(k) = u\sigma_0 k^{u-1} < 0$, indicating that $u < 0$. Consequently, the following relationship between σ_{k-1} and σ_0 is established:

$$\sigma_{k-1} = \sigma_0 k^u, u < 0, k = 1, 2, 3, \dots, N. \quad (4)$$

After constructing the scale space $L_k(x, y, \sigma_{k-1})$, $k = 1, 2, \dots, N$. We apply the single-scale Sobel edge detection method to each scale space image $L_k(x, y, \sigma_{k-1})$, resulting in N edge images $E_k(x, y, \sigma_{k-1})$. Then, we perform an overlay operation on these edge images $E_k(x, y, \sigma_{k-1})$ to obtain an overlaid edge image $E(x, y)$. Last but not least, the overlaid edge image $E(x, y)$ is normalized using Equation (5):

$$\tilde{E}(x, y) = \frac{E(x, y) - a}{b - a}, \quad (5)$$

where $\tilde{E}(x, y)$ represents the normalized image. The minimum value in matrix $E(x, y)$ is denoted as a , and the maximum value is denoted as b . $E(x, y) - a$ refers to subtracting a from each element in $E(x, y)$, and then dividing each element in the resulting matrix $E(x, y) - a$ by $b - a$.

Considering the binary nature of the edge image, the following technique can be employed:

$$\hat{E}(x, y) = \begin{cases} 1, & \text{if } \tilde{E}(x, y) > T_1, \\ 0, & \text{otherwise,} \end{cases} \quad (6)$$

where $\hat{E}(x, y)$ represents the pixel value at coordinate (x, y) , and T_1 is a threshold value. By performing the aforementioned operations, we obtain the resulting edge image $\hat{E}(x, y)$.

2.2. Contour Extraction

Contours offer more continuity than edges, providing geometric details like area and perimeter. Before corner detection, using a contour tracking method is recommended to extract contour sets from multi-modal remote sensing images.

$$L_c = \left\{ l^i \mid l^i = \left\{ (x_1, y_1)^i, (x_2, y_2)^i, \dots, (x_n, y_n)^i \right\} \right\}, \quad (7)$$

where $i = 1, 2, \dots, M$, L_c denotes the set of extracted contours, l represents the elements within each contour, i signifies the i -th contour, and M is the total number of contours in L_c .

2.3. Feature Point Detection

The classical corner detection algorithm, CSS [12], computes the absolute curvature of identified corners along the contour at a low scale. It uses the local maximum of the absolute curvature as initial candidate corner points and employs an adaptive algorithm to eliminate rounded corners, thus reducing false corner detections.

Each contour extracted from a multi-modal remote sensing image can be treated as a curve, represented by an arc length parameter v :

$$\beta(v) = [x(v), y(v)], \quad (8)$$

where $x(v)$ and $y(v)$ denote the sequence of horizontal and vertical coordinates constituting each curve.

The curve $\beta(v)$ is smoothed using a Gaussian function with a scale parameter α to derive the smooth filter curve $\beta_\alpha(v)$.

$$\begin{cases} \beta_\alpha(v) = [x(v) * g(v, \alpha), y(v) * g(v, \alpha)] = [X(v, \alpha), Y(v, \alpha)], \\ g(v, \alpha) = \frac{1}{\sqrt{2\pi}\alpha} e^{-\frac{v^2}{2\alpha^2}}, \end{cases} \quad (9)$$

where the convolution operator symbol $*$ is employed in the context, and $g(v, \alpha)$ represents a one-dimensional Gaussian function.

The absolute curvature of each point on the curve can be determined by:

$$K(u, \alpha) = \frac{|X_v(v, \alpha)Y_{vv}(v, \alpha) - X_{vv}(v, \alpha)Y_v(v, \alpha)|}{[X_v(v, \alpha)^2 + Y_v(v, \alpha)^2]^{\frac{3}{2}}}. \quad (10)$$

where $X_v(v, \alpha) = x(v) * g_v(v, \alpha)$, $Y_v(v, \alpha) = y(v) * g_v(v, \alpha)$, $X_{vv}(v, \alpha) = x(v) * g_{vv}(v, \alpha)$, and $Y_{vv}(v, \alpha) = y(v) * g_{vv}(v, \alpha)$.

By utilizing Equations (8)-(10), we can calculate the absolute curvature for each point along the contour. Subsequently, the CSS [12] selects initial corner candidates at a lower scale based on the maximum value of the absolute curvature $K(v, \alpha)$.

Since the CSS is based on contours, it cannot detect a sufficient number of feature points if the image lacks rich edge detail information. Therefore, we have improved CSS to increase the detection of corner information. The possible relationships between the line segments connecting adjacent feature points on each contour $l^i (i = 1, 2, \dots, M)$ are illustrated in Figure 1).

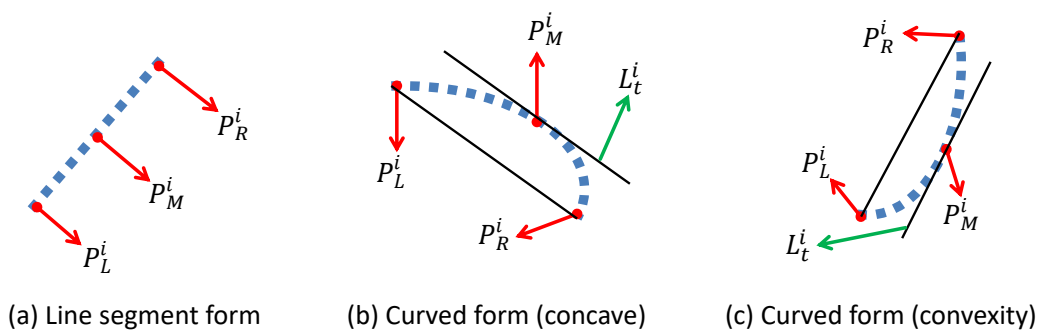


Figure 1. Schematic diagram of possible connecting lines between adjacent feature points on the each contour.

As shown in Figure 1(a), for the case where the line segment connecting feature points $P_L^i(x_L^i, y_L^i)$ and $P_R^i(x_R^i, y_R^i)$ is a straight line, we use the midpoint $P_M^i(x_M^i, y_M^i)$ of the line segment $P_L^i P_R^i$ as a feature point. The midpoint P_M^i can be calculated using the following formula:

$$P_M^i = (x_M^i, y_M^i) = \frac{1}{2}(P_L^i + P_R^i) = \frac{1}{2}(x_L^i + x_R^i, y_L^i + y_R^i). \quad (11)$$

As shown in Figures 1(b) and 1(c), the connecting curve between feature points P_L^i and P_R^i is a curved segment, exhibiting concave and convex shapes respectively. This connecting line segment is continuous between feature points P_L^i and P_R^i . According to Lagrange's theorem, as long as the

function is continuous on the closed interval and differentiable on the open interval, there must be at least one point P_M^i where Equation (12) holds.

$$F(x_M^i) = \frac{y_L^i - y_R^i}{x_L^i - x_R^i}, \quad (12)$$

where $F(\cdot)$ is the derivative of the function $f(\cdot)$ that represents the curved segment between feature points P_L^i and P_R^i .

From Figures 1(b) and 1(c), it can be observed that the straight line segment $P_L^i P_R^i$ connecting P_L^i and P_R^i is parallel to the tangent L_t^i at point P_M^i , and there is only one such tangent. Therefore, the tangent point P_M^i is used as a feature point, and P_M^i is calculated as shown in Equation (13).

$$P_M^i = (x_M^i, y_M^i) = (F^{-1}(x_m^i), f(F^{-1}(x_m^i))), \quad (13)$$

where $F^{-1}(\cdot)$ is the inverse function of $F(\cdot)$.

After the above processing, the number of feature points extracted by the refined CSS corner detection algorithm is $\sum_{i=1}^M (2N^i - 1)$. Here, N^i represents the number of feature points detected by the original CSS algorithm in the i -th contour, and M denotes the total number of contours extracted from the image.

2.4. Dominant Direction

Assigning a dominant direction linked to the local gradient for each feature point is essential for achieving rotation, translation, and scaling invariance, which helps in generating effective descriptors. We propose a novel method to determine the dominant direction, specifically optimized to handle significant intensity variations in multi-modal remote sensing image pairs. For the image $I(x, y)$, the gradient magnitude image G is computed using Equation (2). Subsequently, Equation (1) is utilized to calculate the first horizontal and vertical steps, G_{1x} and G_{1y} , of the gradient magnitude image G , and the gradient magnitude G_1 is determined as:

$$G_1 = \sqrt{G_{1x}^2 + G_{1y}^2}, \quad (14)$$

where G_1 denotes the gradient magnitude image of G .

Next, proceed to normalize the gradient magnitude of the gradient magnitude image G_1 , as defined by the operation:

$$\bar{G} = U(G_1), \quad (15)$$

where the symbol U denotes the normalization operation, which involves dividing all values in G_1 by its maximum value. The resulting \bar{G} represents the normalized gradient magnitude image.

Subsequently, calculate the horizontal and vertical partial derivatives based on the normalized gradient image using Equation (1). This can be expressed as:

$$\begin{pmatrix} \bar{G}_x \\ \bar{G}_y \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial x}(\bar{G}) \\ \frac{\partial}{\partial y}(\bar{G}) \end{pmatrix}, \quad (16)$$

where ∂ denotes the operation of partial derivative.

Equation (16) represents the new gradient definition proposed in this paper. The weighted squared gradient in the average squared gradient method is defined as follows:

$$\begin{bmatrix} G_{w_r, s, x} \\ G_{w_r, s, y} \end{bmatrix} = \begin{bmatrix} \sum w_r \bar{G}_x^2 - \bar{G}_y^2 \\ 2 \sum w_r \bar{G}_x \bar{G}_y \end{bmatrix}, \quad (17)$$

where w_r represents a Gaussian window with variance r . Subsequently, the dominant direction of the feature points can be determined by:

$$D = |\angle(G_{w_r,s,x}, G_{w_r,s,y})|. \quad (18)$$

In Equation (18), the absolute value operation $|\cdot|$ is applied to confine the dominant direction range of feature points from $(-\pi, \pi)$ to $(0, \pi)$. The variable D denotes the dominant direction of the specific feature point. The calculation of $\angle(G_{w_r,s,x}, G_{w_r,s,y})$ is determined using:

$$\angle(X, Y) = \begin{cases} \arctan\left(\frac{Y}{X}\right), & X < 0, Y < 0, \\ \arctan\left(\frac{Y}{X}\right) + \pi, & X \geq 0, \\ \arctan\left(\frac{Y}{X}\right) + 2\pi, & X < 0, Y \geq 0. \end{cases} \quad (19)$$

2.4.1. Feature Descriptor Construction

Due to variations in resolution, intensity, and other attributes, the SIFT descriptor struggles with multi-modal images. The traditional SIFT, based on the first-order gradient, fails to effectively address these disparities. To improve this, we propose a novel gradient definition within the SIFT framework, using Equation (16) to calculate a more robust gradient magnitude G_2 , enhancing the similarity of feature descriptors across multi-modal remote sensing images.

$$G_2 = \sqrt{\bar{G}_x^2 + \bar{G}_y^2}. \quad (20)$$

Utilizing a methodology akin to that outlined in Equation (15), the process of normalization is executed as represented by:

$$\bar{G}_2 = U(G_2). \quad (21)$$

In addition, we applied piecewise normalization to the gradient magnitude of the normalized gradient image \bar{G}_2 . Specifically, the gradient magnitudes were sorted in descending order and then normalized accordingly.

2.4.2. Coarse-to-Fine Feature Matching

We use the Best Bin First (BBF) [13] matching algorithm with generated feature descriptors, performing two-sided matching to establish initial correspondences between feature points.

In the context of bilateral matching of two multi-modal remote sensing images, denoted as $R(x, y)$ (reference image) and $F(x, y)$ (floating image), we identify the nearest p and second-nearest q neighbors related to a point o in $R(x, y)$, as well as the nearest f and second-nearest v neighbors corresponding to a point e in $F(x, y)$. The distance between two points is computed:

$$d_{ij} = \sqrt{\sum_{k=1}^{128} [R_i^k(x, y) - F_j^k(x, y)]^2}. \quad (22)$$

We enforce a distance ratio threshold T_2 , as defined in Equation (23), where satisfying the conditions ensures successful matching between points from both images.

$$\begin{cases} \frac{d_{op}}{d_{oq}} \leq T_2, \\ \frac{d_{ef}}{d_{ev}} \leq T_2. \end{cases} \quad (23)$$

We use the line consistency theory [11] to filter out most of the incorrect matches. Finally, to achieve more robust feature point pairs, we apply the Random Sample Consensus (RANSAC) algorithm [14] for fine-tuning the matches.

3. Experimental Results and Analysis

The experimental setup includes an 11th Gen Intel(R) Core(TM) i3-1115G4 processor running at 3.00GHz, accompanied by 8GB of memory. The system operates on a 64-bit Windows 11 operating system and an x64 based processor architecture. The software utilized is developed on the MATLAB R2019b platform.

3.1. Data Set and Parameter Setting

The test dataset for our method has been sourced from the literature [15], encompassing seven distinct types of multi-modal remote sensing image pairs, such as Cross-Season, Day-Night, Depth-Optical, Map-Optical, and Infrared-Optical. Within this study, each parameter is assigned the following values: $u = \frac{1}{2}$, $\sigma_0 = 1.6$, $N = 5$, $T_1 = 0.5$, $r = 5$, $T_2 = 0.95$.

3.2. Experimental Results of Multi-scale Sobel Edge Detection

Figure (2) shows the original multi-modal remote sensing images, single-scale Sobel edge detection results, and multi-scale Sobel edge detection results. Our multi-scale Sobel approach offers superior edge detection with enhanced continuity, supporting tasks like image registration, contour extraction, and feature point identification. The algorithm's effectiveness is further demonstrated in Figures 2(b), 2(c), 2(h), 2(i), 2(e), 2(f), 2(n), and 2(o), where more edges are extracted compared to the single-scale method, proving its value for advanced image analysis in multi-modal remote sensing.

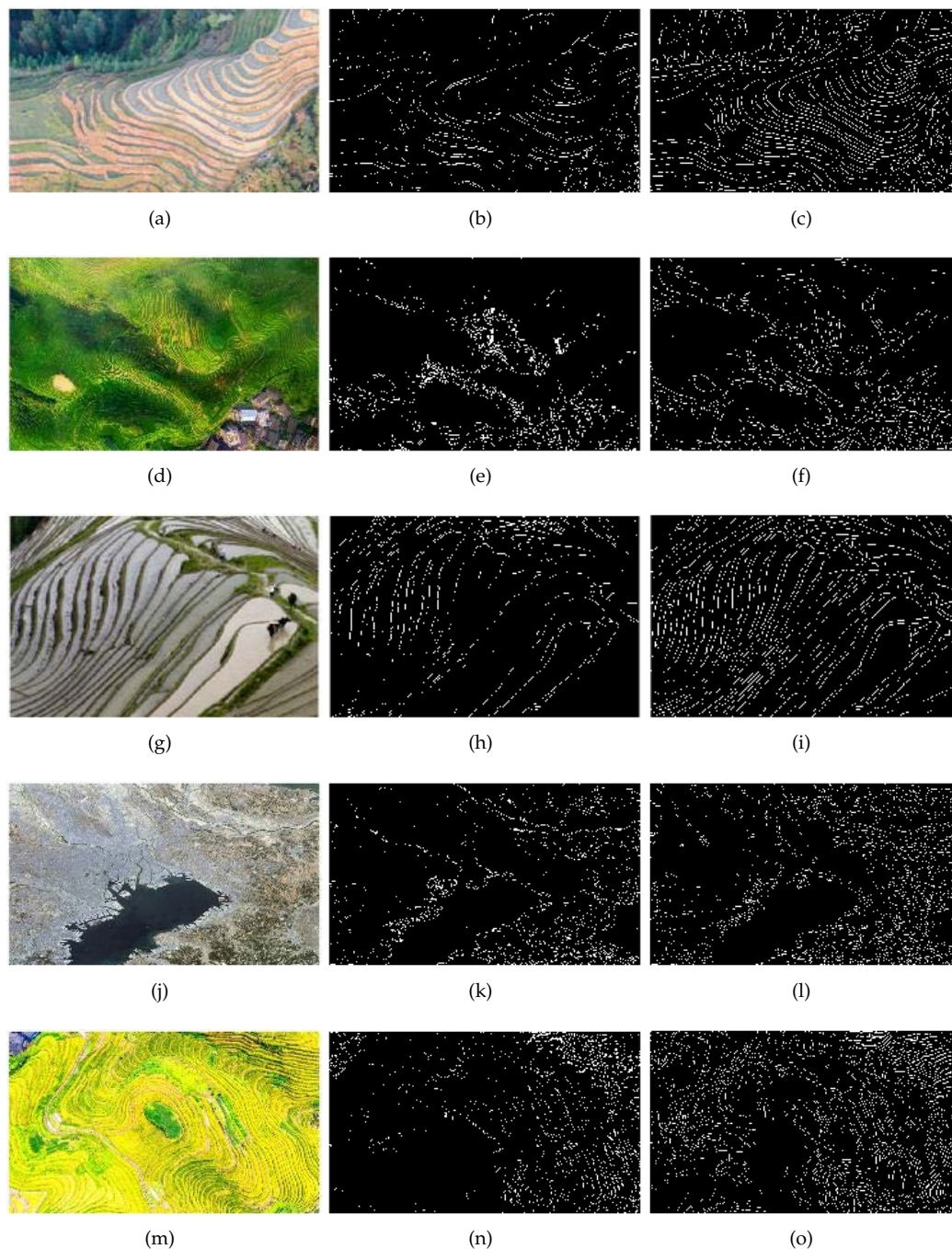


Figure 2. Comparison of images resulting from single-scale and multi-scale Sobel edge detection.

3.3. Subjective Evaluation of the Registration Results

We conducted tests and evaluated our algorithm through subjective assessment against several established classical multi-modal image registering techniques, namely LGFC [9], MS-PIIFD [6], CAO-C2F [11], RIFT [8], and MS-HLMO [7].

Figure (3) presents partial test results of our algorithm compared to five advanced multi-modal image registration techniques across twelve image pairs. Each row shows the results for one algorithm. The LGFC algorithm, while correctly matching some feature points, fails when fewer than three pairs are matched, as at least three valid pairs are required to solve the transformation matrix. This failure occurs due to LGFC's reliance on large gap fracture contours. The MS-PIIFD algorithm struggles with intensity differences, as seen in Figure 3(f), but succeeds in some cases. In contrast, MS-HLMO, CAO-C2F, and RIFT perform better, yielding more visible matching pairs due to their design for

handling intensity and resolution differences. While our algorithm doesn't visually match as many pairs as these methods, it still effectively identifies correct pairs, demonstrating its adaptability to image pairs with intensity differences. Since some subtle differences may be undetected by the human eye, objective evaluations were conducted in the following section.

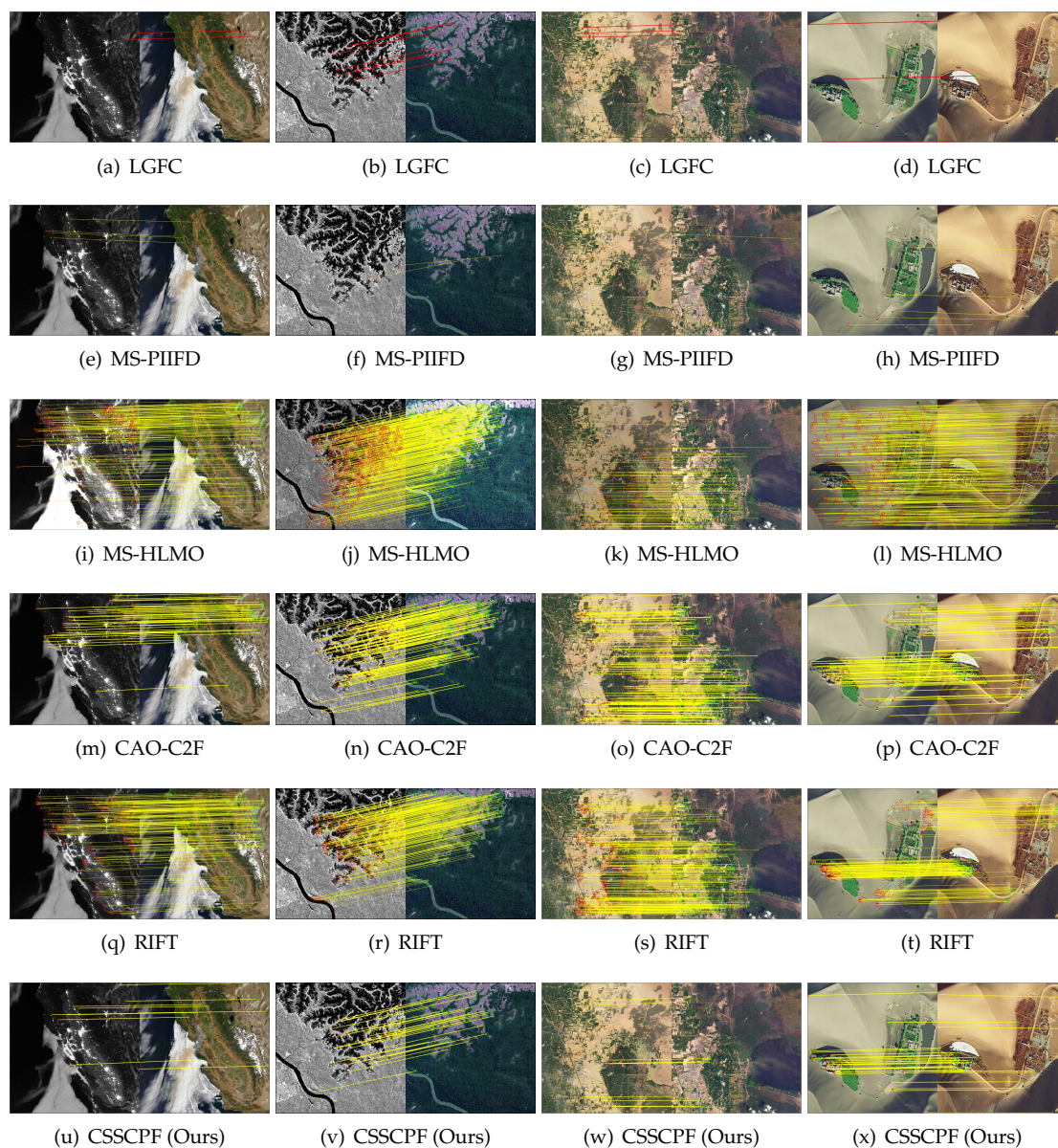


Figure 3. The proposed algorithm and the intuitive test results of different multi-modal image registering algorithms, where each row represents the test results of one multi-modal image registering algorithm.

3.4. Objective Registering Results and Analysis

Table (1) compares various algorithms with the proposed registering algorithm.

Table 1. Average running time, average Number of Correct Matching (NCM) points, average Root Mean Square Error (RMSE), and average registering accuracy of different image registering algorithms.

Algorithms	Average running time (s)	Average NCM	Average RMSE	Average registering accuracy (%)
LGFC	4.1215	3	8.4699	25.7
MS-PIIFD	18.3924	4	1.8334	29.08
CAO-C2F	25.2271	43	1.8796	20.84
RIFT	10.2207	147	1.9241	10.09
MS-HLMO	214.514	267	2.2151	18.75
CSSCPF (Ours)	5.5795	14	1.8542	46.74

3.4.1. Running Time of Different Algorithms

Table (1) shows the average running times of various algorithms. The MS-HLMO algorithm has the longest running time, nearly 4 minutes, followed by CAO-C2F. MS-PIIFD and RIFT perform slightly better than CAO-C2F, yielding relatively favorable results. The LGFC algorithm has the shortest running time, while our algorithm ranks second with an average time of 5.5795 seconds. This is acceptable, though further optimization is possible for improvement.

3.4.2. NCM Point Pairs for Different Algorithms

Table (1) presents the average NCM point pairs for various algorithms. Due to matching issues in the LGFC and MS-PIIFD algorithms, only successfully matched pairs were used for calculating the average NCM. Typically, at least three point pairs are needed to compute a transformation matrix. While our algorithms final average NCM of 14 isn't as high as RIFT and MS-HLMO, it is sufficient for computing the transformation matrix. The CAO-C2F algorithm, designed for multi-modal power equipment images, also shows a notable average NCM. In contrast, MS-PIIFD and LGFC perform relatively poorly.

3.4.3. RMSE of Different Algorithms

We calculated the average RMSE and compared registration times, as shown in Table (1). A smaller RMSE indicates higher registration accuracy. Our algorithm achieved an average RMSE of 1.8542, ranking second among all methods. While MS-PIIFD slightly outperforms our method in RMSE, it underperforms in several other metrics. LGFC performed the worst, and CAO-C2F and RIFT had RMSE values similar to ours. MS-HLMO ranked third with an average RMSE of 2.2151. Overall, our algorithm performs excellently in terms of average RMSE.

$$\text{RMSE} = \sqrt{\frac{1}{n} \left[\sum_{j=1}^n (x_{1j} - x_{2j})^2 + (y_{1j} - y_{2j})^2 \right]}, \quad (24)$$

where (x_{1j}, y_{1j}) represent the coordinates of the reference image matching point, and (x_{2j}, y_{2j}) represent the coordinates of the corresponding matching point of the floating image after affine transformation to the coordinate system of the reference image. The parameter n denotes the number of matched point pairs.

3.4.4. Registration Accuracy of Different Algorithms

We assess registration accuracy by calculating the ratio of NCM to the initial number of matching point pairs. As shown in Table (1), RIFT and MS-HLMO exhibit relatively low registering accuracy due to a high number of initial matching point pairs, leading to more erroneous matches. MS-PIIFD and CAO-C2F show better accuracy, with LGFC in second place. Our algorithm achieves the best performance. An ablation experiment comparing single-scale and multi-scale Sobel edge detection showed that the single-scale version had an average NCM 23 lower than the multi-scale

version, validating the effectiveness of our approach. Overall, our algorithm demonstrates excellent performance.

4. Conclusion

Multi-modal remote sensing image registration remains a complex challenge, with existing algorithms still in early stages of development. Due to differences in intensity, spectral characteristics, and viewing angles, current methods face limitations in both registration time and accuracy. To address these issues, we propose a novel approach that introduces a new gradient definition and technique for determining the dominant direction of feature points, along with improvements to the CSS corner detection algorithm and SIFT descriptor. Experimental results on public datasets demonstrate superior performance.

Author Contributions: J.Z. wrote the main manuscript and conducted the related experiments, C.L. revised and proofread the paper and D.L. provided the software.

Data Availability Statement: Code and data will be made available on request.

Acknowledgments: This work was supported in part by the Open Project of the Key Lab of Enterprise Informationization and Internet of Things of Sichuan Province grant number 2022WZJ01, Postgraduate course construction project of Sichuan University of Science and Engineering grant number YZ202103, Research Project on Teaching Reform at Sichuan University of Science and Technology grant number JG-2302, Talent Introduction Program at Sichuan University of Science and Engineering grant number 2023RC22 and Graduate Innovation Fund of Sichuan University of Science and Engineering grant number Y2023338.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhang, X.; Leng, C.; Hong, Y.; Pei, Z.; Cheng, I.; Basu, A. Multimodal remote sensing image registration methods and advancements: A survey. *Remote Sensing* **2021**, *13*, 1–31.
2. Ye, Y.; Shan, J.; Bruzzone, L.; Shen, L. Robust registration of multimodal remote sensing images based on structural similarity. *IEEE Transactions on Geoscience and Remote Sensing* **2017**, *55*, 2941–2958.
3. Huang, Q.; Guo, X.; Wang, Y.; Sun, H.; Yang, L. A survey of feature matching methods. *IET Image Processing* **2024**, *18*, 1385–141.
4. Ma, J.; Jiang, X.; Fan, A.; Jiang, J.; Yan, J. Image matching from handcrafted to deep features: A survey. *International Journal of Computer Vision* **2021**, *129*, 23–79.
5. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* **2004**, *60*, 91–110.
6. Gao, C.; Li, W. Multi-scale PIIFD for registration of multi-source remote sensing images. *Journal of Beijing Institute of Technology* **2021**, *30*, 113–124.
7. Gao, C.; Li, W.; Tao, R.; Du, Q. MS-HLMO: Multiscale histogram of local main orientation for remote sensing image registration. *IEEE Transactions on Geoscience and Remote Sensing* **2022**, *60*, 1–14.
8. Li, J.; Hu, Q.; Ai, M. RIFT: Multi-modal image matching based on radiation-variation insensitive feature transform. *IEEE Transactions on Image Processing* **2020**, *29*, 3296–3310.
9. Zhu, J.; Liu, C.; Yang, Y. Robust image registration for power equipment using large gap fracture contours. *IEEE MultiMedia* **2024**, *31*, 53–64.
10. Öfverstedt, J.; Lindblad, J.; Sladoje, N. Fast and robust symmetric image registration based on distances combining intensity and spatial information. *IEEE Transactions on Image Processing* **2019**, *28*, 3584–3597.
11. Jiang, Q.; Liu, Y.; Yan, Y.; Deng, J.; Jiang, X. A contour angle orientation for power equipment infrared and visible image registration. *IEEE Transactions on Power Delivery* **2021**, *36*, 2559–2569.
12. Yung.; Nelson, H.C. Corner detector based on global and local curvature properties. *Optical Engineering* **2008**, *47*, 1–13.
13. Beis, J.S.; Lowe, D.G. Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In Proceedings of the Conference on Computer Vision & Pattern Recognition, 1997.
14. Fischler, M.A.; Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Readings in Computer Vision* **1981**, *24*, 381–395.

15. Jiang, X.; Ma, J.; Xiao, G.; Shao, Z.; Guo, X. A review of multimodal image matching: Methods and applications. *Information Fusion* **2021**, *73*, 22–71.

Short Biography of Authors



Jianhua Zhu received his B.S. degree in mathematics and applied mathematics from Xichang University, and his M.A.Sc. degree in mathematics from Sichuan University of Science and Engineering, Zigong, 643000, China. His research interests include image processing and computer vision. This work was completed during his master's studies. He is the first author of this article. Contact him at tostuhua@qq.com.



Changjiang Liu is an associate professor with the Key Laboratory of Higher Education of Sichuan Province for Enterprise Informationalization and Internet of Things, Sichuan University of Science and Engineering, Zigong, 643000, China. His research interests include image processing and computer vision. Liu received his Ph.D. degree in image segmentation and image registration from Sichuan University. He is the corresponding author of this article. Contact him at liuchangjiang@189.cn.



Danling Liang is currently working toward her M.A.Sc. degree, focused on the image segmentation, with the School of Mathematics and Statistics, Sichuan University of Science and Engineering, Zigong, 643000, China. Her research interests include image processing. Liang received her B.S. degree in mathematics and applied mathematics from Sichuan University of Science and Engineering. She is the co-author of this article. Contact her at liangdanling@163.com.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.