A High Precision Feature Matching Method Based on Geometrical Outlier Removal for Remote Sensing Image Registration

HAN YANG[1], XIAORUN LI[1], YIJIAN MA[2,3], LIAOYING ZHAO[4], AND SHUHAN CHEN[1]

[1]Faculty of Electrical Engineering, Zhejiang University, Hangzhou 310000, China
[2]Zhejiang Academy of Special Equipment Science, Hangzhou 310000, China
[3]Key Laboratory of Special Equipment Safety Testing Technology of Zhejiang Province, Hangzhou 310000, China
[4]School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China

Corresponding author: Xiaorun Li (lxr@zju.edu.cn)

ABSTRACT The reliability of feature matching is required for accurate remote sensing image registration. Outliers reduce the accuracy and effectiveness of feature matching. This paper proposes a novel feature matching method that utilizes the global geometric relationship of feature points in two images to eliminate outliers and preserve inliers. A mathematical model is formulated based on the similarity of the geometric relationship of feature points in the reference image and the sensed image. We also find the optimization solution through analysis and simplification of the mathematical model. The corresponding feature matching algorithm based on outlier removal is proposed according to the optimization solution. The experimental results of several remote sensing images demonstrate that our method can preserve more inliers, remove more outliers and obtain a better registration performance with higher accuracy and robustness than the state-of-the-art methods, such as SIFT, SIFT-RANSAC, SIFT-GTM, SIFT-LPM.

INDEX TERMS Remote sensing, image registration, feature matching, outlier removal, mathematical model.

I. INTRODUCTION

Image registration is the process of aligning two or more overlapping images of the same scene taken from different viewpoints, at different times, or by different image sensors [1]–[4]. Image registration is a critical preprocessing step for super resolution imaging, image mosaicking, remote sensing, medical imaging, and computer vision [5]–[7]. Remote sensing image is captured by the imageing spectrometer with several contiguous and spectral bands, which is different from traditional image [8], [9]. Therefore, image registration for remote sensing image is significant.

Feature-based registration is a robust and effective image registration method [10]. Feature-based registration algorithms include three main steps: (a) extraction of features, (b) matching of the features by searching for a transformation that best aligns them, and (c) resampling the sensed image to construct a new image in the coordinate system of the reference image [11]. Feature point matching is a challenging step in feature-based registration techniques, required to establish reliable correspondences between feature points extracted from the reference image and the sensed image.

Because the number of feature points extracted by feature-based registration techniques is very large, a huge matching space is required, increasing the probability of false matches [12]. Another limitation is that feature point matching based on local descriptor information can produce outliers due to images with duplicate or similar patterns. Finally, the intensity and texture information can be very different in the same region due to slightly different perspectives or lighting conditions. Thus, outliers caused by the above factors are necessary to be eliminated before feature points matching.

To solve the outlier elimination challenge, several different methods have been proposed, which will be introduced in section 2. Above all, local geometrical structure of feature points has been fully utilized to eliminate outliers, which is

The associate editor coordinating the review of this manuscript and approving it for publication was Lefei Zhang[ID].

a smart and efficient method to remove outliers. However, the local information of feature points can change significantly with the change of grayscale or texture, the global geometric relationships of feature points are generally well preserved. Scale relationship of feature points is a significant global geometric relationship, which will not be affected by grayscale or texture. Obviously, inliers are bound to follow the similar scale relationship. In contrast, the scale relationship of outliers is inconsistent. According to this principle, we propose an outlier elimination method based on the global geometric relationship of feature points, which can accurately and efficiently remove outliers and preserve inliers. Additionally, the similarity of the global geometric relationships between inliers are nearly unaffected by the number of outliers, different perspectives or lighting conditions. Based on this fact, we construct a mathematical model to measure the difference of geometric relationships in the feature points of two images. Our method combines the global geometric relationship and the function-based fitting, precisely represents the similarity of the global geometric relationship between feature points, which can help easily eliminate outliers. Experiments using various remote sensing image data demonstrate that the proposed method can eliminate most outliers and preserve most inliers, resulting in improved registration accuracy and effectiveness. The main contributions of this manuscript are given as follows:

1. We propose a high precision feature matching method based on outlier removal, which makes full use of the global geometrical similarity, improving the accuracy and robustness of image registration.

2. An effective mathematical model is formulated based on the similarity of the geometric relationship of feature points in the reference image and the sensed image to remove outliers extracted by feature based method.

3. The proposed clever solution of the mathematical model reduces the computational complexity.

The remainder of this paper is organized as follows. In section 2, the background material and related work is introduced. Section 3 describes the outlier elimination method based on global structure. Section 4 illustrates the experimental results and analysis of various remote sensing images using this method and other popular methods. Finally, conclusions are presented in Section 5.

II. RELATED WORK

In this section, we describe and interpret the existing knowledge of feature-based registration and outlier removal methods. First, the scale-invariant feature transform (SIFT) [13] algorithm will be introduced because this classical algorithm can be used to generate the initial point set. Then we will introduce related studies of outlier elimination in registration.

A. THE SCALE-INVARIANT FEATURE TRANSFORM (SIFT) ALGORITHM

The SIFT algorithm [13] is a powerful feature point matching approach that contains two main parts, feature detection and descriptor extraction. The task of feature detection is to identify some points whose certain characteristics do not change with affine transformation of images. Feature description is required to describe the unchanging characteristics for matching.

Feature detection includes three steps: extracting scale space extremum, improving the accuracy of localization, and eliminating unstable extrema [14]. First, scale space is built by the original image using different Gaussian kernels, and then local extremum points are identified in the scale space. Next, the location of the extremum points is calculated using the difference Gaussian scale space function. Finally, unstable extremum points are eliminated through contrast and curvature.

To make each descriptor have rotational invariant property, it is necessary to determine the main direction of feature points. The main direction is the direction of the largest gradient of the field around the feature points. The gradient distribution of the field around the feature points is recorded as the descriptor.

B. RELATED RESEARCH OF OUTLIER ELIMINATION IN REGISTRATION

Many methods have been proposed in recent years to remove outliers from the initial point set. Different methods rely on different principles: function-based fitting, statistical model-based, and graph-based methods.

According to the size of the image and the geometric deformation, the function-based fitting method selects the most appropriate function model, and determines the optimal solution by mathematical methods. Robust point matching via vector field consensus [15] interpolates a vector field and formulates a maximum a posteriori estimation of a Bayesian model to remove outliers. LLT proposed in 2015 develops a local geometrical constrain and formulates a maximum-likelihood estimation to preserve local structures of feature points and inliers [16]. Locality preserving matching and guided locality preserving feature matching (GLPM) are also two good function-based fitting methods proposed by Jiayi Ma [17], [18], which preserve the neighborhood structures of potential true matches between two images. Methods based on function fitting has advantages of simple calculation and fast speed, but the accuracy of outlier detection completely dependent on the selected function model. Thus, an unsuitable function model will lead to a poor selection result. Additionally, this method is highly dependent on the number of feature points. There are typically insufficient feature matching pairs for real remote sensing image registration, and few feature points belong to correct feature points, so this method will not work well.

Statistical model-based methods count the matching points that satisfy a given condition. Random sample consensus (RANSAC) [19] is the most popular statistical model-based method. RANSAC iterates and estimates the optimal parameters of a given model from data sets containing outliers. Maximum likelihood estimation sample consensus [20] and

progressive sample consensus [21] are improved methods based on RANSAC. RANSAC and related methods utilize simple calculations and have good robustness, making these models widely applied for the detection of outliers in remote sensing image [22]–[24]. However, for actual processing, the number of iterations is difficult to calculate accurately. The use of an insufficient number of iterations may result in bad model parameters, and too many iterations will make the calculations too complicated.

Graph-based methods utilize the geometric relationships between key points and the relationships between key points and their fields. Graph transformation matching (GTM) identifies a consensus nearest-neighbor graph from a candidate point set [25]. Recovery and filtering vertex trichotomy matching (RFVTM) [12] exclude outliers based on the idea that geometrical relations between any vertices and lines are preserved after affine transformations. However, these methods not work well when outliers have similar local structures as those of inliers.

Besides these three kinds of methods, a novel mismatch removal method based on learning is proposed in 2019, which trains a classifier based on a general match representation to distinguish outliers and inliers [26]

Overall, the above-mentioned outlier removal methods have limited accuracy that is affected by the number of feature points, the fraction of outliers, and the local structures of feature points. In this paper, we propose an effective outlier elimination method based on the geometric relationships of feature points, that is not sensitive to the number of the initial feature points, the initial proportion of correct matching pairs or the local information of feature points. This new method is effective and robust which can remove outliers from different condition of feature points.

C. TRANSFORMATION MODELS

Typical transformation models such as rigid, affine and similarity model are wildly used in remote sensing image registration task [16]. Different views of satellites are easy to take remote sensing images with scale, rotation, translation and other transformations. Therefore, we assume a similarity transformation model which is wildly used in the registration of remote sensing image registration [27]. Similarity transformation model is constituted by for parameters: scale, rotation, vertical shifts and horizontal shifts. Under the similarity transformation model, inliers have similar scale information, while outliers have different scale information, which is used to remove outliers in this manuscript.

III. PROPOSED METHOD

In this section, we will introduce an outlier elimination method to register remote sensing images with overlapping areas. As generally accepted, a good matching result relies on the number and proportion of correct matching pairs, therefore, it is important to select as many correct points as possible.

Due to the formulation and solution of the proposed method refers to the LPM method [18], we plan to give a brief introduction of the LPM method.

A. BRIEF INTRODUCTION OF LPM

The main idea of the LPM method is to preserve the local neighborhood structures of feature points. Then a mathematical model is formulated for the local neighborhood structures.

The mathematical model is as follows:

$$I^* = \arg\min_I C(I; S, \lambda) \tag{1}$$

With the cost function C defined as:

$$C(I; S, \lambda) = \sum_{i \in I} \sum_{j \in I} (d(x_i, x_j) - d(y_i, y_j))^2 + \lambda(N - |I|) \tag{2}$$

where I represents the unknown inlier set, S is the initial feature point set λ is a parameter. The operation of d is to decide whether x_j belongs to the neighborhood of x_i. To optimize the objective function, the Eq.(2) can be transferred to:

$$C(p; S, \lambda) = \sum_{i=1}^{N} p_i(c_i - \lambda) + \lambda N \tag{3}$$

where

$$c_i = \sum_{j|x_j \in N_{x_i}} d(y_i, y_j) + \sum_{j|y_j \in N_{y_i}} d(x_i, x_j) \tag{4}$$

Then calculate the c_i of all the feature points, and if $c_i <= \lambda$, the feature point is inlier, otherwise the feature point is outlier.

B. PROBLEM FORMULATION

Here, an similarity deformation model is used to describe the relationship of feature points in the reference image and the sensed image.

$$\begin{bmatrix} x_{ref} \\ y_{ref} \end{bmatrix} = a * \begin{bmatrix} \cos\theta, -\sin\theta \\ \sin\theta, \cos\theta \end{bmatrix} * \begin{bmatrix} x_{sen} \\ y_{sen} \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \tag{5}$$

where (x_{ref}, y_{ref}) represent the coordinates in the reference image, (x_{sen}, y_{sen}) represent the coordinates in the sensed image, a represents the scale factor, *theta* represent the rotation angle, and t_x and t_y are the translations between the reference image and the sensed image, respectively.

The initial matching feature point set I is generated by applying the SIFT algorithm and Euclidean distance-based matching. We arbitrarily select a pair of points (s_i^{ref}, s_j^{ref}) from the initial matching feature point set $S_{ref} = \{s_i^{ref}\}_{k=1}^{N} = \{(x_i^{ref}, y_j^{ref})\}_{k=1}^{N}$ in the reference image, where $(i, j) \in \{k, k = 1, 2, \ldots, N\}$. Then, the corresponding point pair (s_i^{sen}, s_j^{sen}) is selected from the point set $S_{sen} = \{s_i^{sen}\}_{k=1}^{N} = \{(x_i^{sen}, y_i^{sen})\}_{k=1}^{N}$ in the sensed image, where $(i, j) \in \{k, k = 1, 2, \ldots, N\}$, and N is the number of the initial matching feature point set. We define D as the distance of two points.

$$D(s_i^{ref}, s_j^{ref}) = \left\| s_i^{ref}, s_j^{ref} \right\|$$

$$= \sqrt{(x_i^{ref} - x_j^{ref})^2 + (y_i^{ref} - y_j^{ref})^2} \quad (6)$$

Combining Eq.(5) and Eq.(6), it is easy to obtain Eq.(7):

$$D(s_i^{ref}, s_j^{ref}) = a * \left\| s_i^{sen}, s_j^{sen} \right\| = a * D(s_i^{sen}, s_j^{sen}) \quad (7)$$

Therefore, the ratio of $D(s_i^{ref}, s_j^{ref})$ and $D(s_i^{sen}, s_j^{sen})$ is the scale factor. Based on this, we construct an objective function to select inlier from the initial points.

$$S_{in}^* = \arg \min_{S_{in}} F(S_{in}; I, \lambda) \quad (8)$$

The objective function F is defined as

$$F(S_{in}; I, \lambda) = \sum_{i \in S_{in}} \sum_{j \in S_{in}} \sum_{k \in S_{in}} \left| \frac{D(s_i^{ref}, s_j^{ref})}{D(s_i^{sen}, s_j^{sen})} - \frac{D(s_i^{ref}, s_k^{ref})}{D(s_i^{sen}, s_k^{sen})} \right| + \lambda(N - |S_{in}|) \quad (9)$$

In the objective function, S_{in} represents the unknown inlier set, and I represents the initial matching feature point set. λ is the parameter that controls the tradeoff between the inlier set and the outlier set.

C. SOLUTION

In order to find the optimal solution of the objective function, we have to modify Eq. (5) because we do not know the inlier set S_{in}. So, we suppose an $N * 1$ binary vector p, If the point belongs to the inlier set, the value of p_i is 1, else the value of p_i is 0. Then, the number of inliers can be expressed as $\sum_{i=1}^{N} p_i$. In this way, Eq.(9) is converted to Eq.(10) and Eq.(11).

$$F(S_{in}; I, \lambda) = \sum_{i=1}^{N} p_i \left(\sum_{j=1}^{N} \sum_{k=1}^{N} \left| \frac{D(s_i^{ref}, s_j^{ref})}{D(s_i^{sen}, s_j^{sen})} - \frac{D(s_i^{ref}, s_k^{ref})}{D(s_i^{sen}, s_k^{sen})} \right| \right)$$
$$+ \lambda(N - \sum_{i=1}^{N} p_i) \quad (10)$$

$$F(S_{in}; I, \lambda) = \sum_{i=1}^{N} p_i(f_i - \lambda) + \lambda N \quad (11)$$

$$f_i = \sum_{j=1}^{N} \sum_{k=1}^{N} \left| \frac{D(s_i^{ref}, s_j^{ref})}{D(s_i^{sen}, s_j^{sen})} - \frac{D(s_i^{ref}, s_k^{ref})}{D(s_i^{sen}, s_k^{sen})} \right| \quad (12)$$

For the initial set, the coordinates of each point are known, all the cost value $\{f_i\}_{i=1}^{N}$ can be calculated in advance, and the only unknown variable is p_i. Obviously, the solution is that any cost value f_i that is smaller than λ is put into the inlier set, which decreases the objective function. Any cost value f_i that is larger than λ is put into the outlier set, which increases the objective function. Then, the optimal solution of the objective function Eq.(9) is determined by the criteria as follows.

$$p_i = \begin{cases} 1, f_i \leq \lambda \\ 0, f_i > \lambda \end{cases} \quad i = 1, 2, \ldots, N \quad (13)$$

The optimal inlier set S_{in} is:

$$S_{in} = \{i \,|\, p_i = 1, i = 1, 2, \ldots, N\} \quad (14)$$

D. ALGORITHM DESCRIPTION

The outlier elimination algorithm improves registration accuracy by removing outliers based on the local geometric relationships. There are three main steps of the algorithm: extraction the initial matching feature point set, calculation of the optimal solution of the objective function, and calculate the registration result. The procedures of the algorithm are summarized in **Algorithm 1**.

Algorithm 1

Input: An image pair, Initial parameter λ
Output: Correct match set S_{in}^* and registration result

1: Establish the feature point set S_{ref} and S_{sen} using the SIFT algorithm and Euclidean distance-based matching with the distance ratio threshold $\iota = 0.8$;
2: Calculate the cost value $\{f_i\}_{i=1}^{N}$ using Eq.(12) based on S_{ref} and S_{sen};
3: Determine S_{in}^* according to Eq.(13) and Eq.(14);
4: Calculate the root mean square error ($rmse_n$) of the image pair based on feature points set S_{in}^*;
5: Update S_{ref} and S_{sen} based on the feature point set S_{in}^*, repeat step 2,3, and 4
6: Compare the two calculation results ($rmse_n$ and $rmse_{n+1}$);
7: **if** $|rmse_{n+1} - rmse_n| < 0.1$ **then**
8:     output the correct match set S_{in}^* and registration result
9: **else**   Update the parameter λ and **Jump** to step 5;
10: **End**

To selecting many correct points, the initial point set is request to contain enough correct points even if there are also many outliers. Therefore, we select the feature point pairs produced by SIFT algorithm and Euclidean distance-based matching as the initial feature point set.

With a large number of outliers in the initial point set, outliers will affect the cost value of inliers, resulting in a large cost value of inliers. Removing the outliers once may cause some outliers not to be removed cleanly. Thus, it is necessary to implement the optimization process several times until the difference of the two previous registration results remains almost unchanged.

From Eq.(12), it is obvious that the cost value can also be influenced by the number of the point set, and a small value of λ can increase the precision and decrease the recall. If we set λ as a constant regardless of the number of feature points, it is difficult to get a satisfactory result for different images and different numbers of feature points. Therefore, we set a dynamic value according to the number of feature points. Through several experiments, when we set $\lambda = N/100$, the result and efficiency of the algorithm is best, and N is the number of feature points. Due to multiple loops of eliminating outliers, the value of λ can change immediately through different number of feature points in loops, and remove outliers effectively.
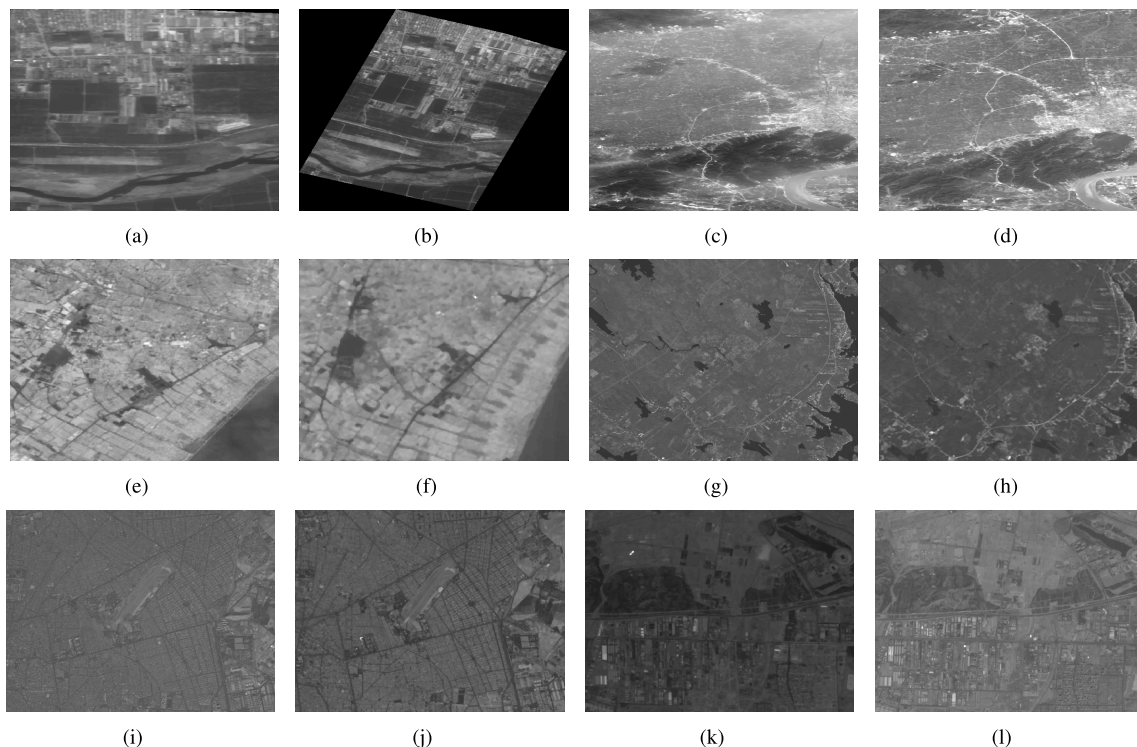
FIGURE 1. Four pairs of original images. image(a) and (b) correspondences to Image pair 1, image(c) and (d) correspondences to Image pair 2, image(e) and (f) correspondences to Image pair 3, image(g) and (h) correspondences to Image pair 4, image(i) and (j) correspondences to Image pair 5, image(k) and (l) correspondences to Image pair 6.

E. COMPUTATIONAL COMPLEXITY

According to Eq. (12), calculating the cost $\{f_i\}_{i=1}^N$ in Lines 2 costs the main operation of the proposed algorithm, and its time complexity is $O(N^3)$. The time complexity of Line 4 and Line 7 involves some addition operation, and its time complexity is $O(N)$. Besides, the determination of p and S_{in} through Eq. (13) and Eq. (14) cost $O(N)$. Therefore, the total time complexity of the proposed method is about $O(N^3)$. The space complexity of the proposed method is $O(N)$ due to the memory requirements for storing the cost value $\{f_i\}_{i=1}^N$. In order to make full use of the global geometric relationship of feature points, the proposed method spend some operation and has a nonlinear time complexity. But our algorithm has a linear space complexity.

IV. DATASET AND EVALUATION CRITERION

A. DATASET

As is known, the difference of scenes, spatial resolution, grayscale and sensors will all affect the result of image registration. To test and verify the high accuracy and robustness of the proposed method, we conduct six experiments. In experiment 1, we adopt two simulated images to verify the feasibility of the proposed method. For the other 5 experiments, the experimental images are all real image. From experiment 2-6, we gradually increase the scale difference between images. Besides, the scenes and gray-scale are all different to verify the high accuracy and robustness of the proposed method.

The details of six pairs of remote sensing images are shown in the next paragraph.

We selected six remote sensing image pairs considering differences of grayscale, resolution, and scene, with a pair of simulated remote sensing images and three pairs of real remote sensing images, which is displayed in Fig 1. Image pair1 includes two simulated hyperspectral remote sensing images with 20° rotation and 4 times scaling, which covers a city area located in Colombia-Cali and correspondences to Fig 1(a) and (b). Image pair 2 has a significant grayscale difference between the sensed image (band 3from Landsat TM) and the reference image (band 1 from SPOT 4), mainly covering mountain area at Hangzhou, which correspondences to Fig 1(c) and (d). Image pair 3 has a significant spatial resolution difference (about 1.875 times) between the sensed image (band 55 from EO-1) and the reference image (from google earth) covering a plain area at Suzhou, which correspondences to Figure 1(e) and (f). Image pair 4 has a significant spatial resolution difference (about 3 times) and texture difference between the sensed image (band 1 from SPOT 5) and the reference image (band 3 from Landsat ETM+), covering a plain area at Halifax, which correspondences to Fig 1(g) and (h). Image Pair 5 has a significant noise and scale difference between the sensed image (SPOT 4) and the reference image (IRS-1C), covering a farmland area in Iran-Tehran, which corresponds to Figure 1(i) and (j). Image Pair 6 has a significant spatial resolution difference (about 2 times) and grayscale difference between the sensed image (IRS-1C)
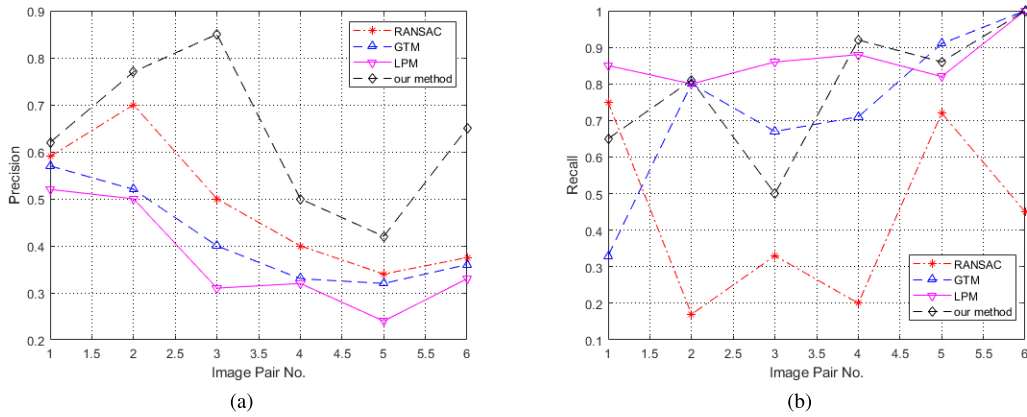
FIGURE 2. Comparison of feature point matching with outlier removal.

TABLE 1. Specifications of image data for experiments.

data type	No.	Satellite	Spectral Mode	Image Size	Pix Size(m/pixel)	Bits per pixel	Acquisition Data
Simulated image	1		Hyperspectral-Band:45	500*325	5	32	2005
			Hyperspectral-Band:45	2324*1906	20	32	2005
Real image	2	SPOT4	Multispectral-Band:1	611*1235	30	8	2001
		Landsat TM	Multispectral-Band:3	648*1230	30	8	2004
	3	EO-1	Hyperspectral-Band:55	239*256	7.5	16	2014
		Google		543*508	4	16	2017
	4	SPOT-5	Panchromatic	1311*1215	10	8	2006
		Landsat ETM+	Multispectral-Band:3	440*410	30	8	1999
	5	IRS-1C	Panchromatic	1346*1135	5	8	1998
		SPOT 4	Panchromatic	700*590	10	8	1996
	6	IRS-1C	Panchromatic	1122*1032	5	6	1998
		SPOT 4	Panchromatic	568*522	10	8	1999

and the reference image (SPOT 4), covering a field and town area at Iran-Tehran, which corresponds to Figure 1(k) and (l). Table 1 lists the details of the four typical image pairs.

B. EVALUATION CRITERION

To evaluate the accuracy and efficiency of the proposed method, we compared the feature point matching of our method with three other matching methods RANSAC, GTM and LPM in terms of RMSE and Precision and Recall [28]. The initial feature point set are extracted using SIFT algorithm with $d_{radio} = 0.8$ for the remote sensing images.

The reliable reference geometric transformation parameters are calculated by manually selecting tie-points via ENVI and manual registration. Correct matches are feature points with a spatial distance less than 1 pixel between the reference image and the sensed image after correction with the manual registration results. The initial value of parameter (λ) is set as the square of the initial number of feature points. Here, RMSE is used as the measurement of the position error of image registration. Precision is the proportion of the correct matches of the residual matches. The greater the precision and the proportion of correct matches, the better the accuracy of the image registration will be. Recall represents the residual correct matches with respect to the initial correct matches, and

indicates the effectiveness and reliability of outlier removal. The calculation formulas are as follow:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}((x_i^1 - x_i^2)^2 + (y_i^1 - y_i^2)^2)} \quad (15)$$

$$Precision = \frac{residual\ correct\ matches}{residual\ matches} \quad (16)$$

$$Recall = \frac{residual\ correct\ matches}{initial\ correct\ matches} \quad (17)$$

In Eq.(11), N points (x,y) are taken in the reference image, then points (x^1, y^1) are obtain through the affine transformation model using the image registration parameter calculated by the algorithm. (x^2, y^2) are obtained through the real affine transformation model as calculated by manual selection.

V. EXPERIMENTAL RESULTS

In this section, experiments are described that serve to evaluate the precision and effectiveness of the proposed method using a laptop with 2-GHz CPU and 8-GB RAM (Intel Core 5). We also compare our method with two other powerful matching algorithms, RANSAC, GTM and LPM. The image data set and evaluation criterion are introduced, and then the experimental results of our method, RANSAC, GTM

TABLE 2. Comparision of Deformation parameter evaluated by the proposed method, three other different methods and ground truth.

Image pair No.	initial matches	initial correct matches	Method	a_{11}	a_{12}	a_{21}	a_{22}	t_x	t_y	RMSE
1	576	291	Ground truth	0.235	0.086	-0.086	0.235	-160.53	59.12	——
			RANSAC	0.235	0.086	-0.086	0.235	-160.79	58.35	0.59
			GTM	0.235	0.086	-0.086	0.235	-160.81	58.37	0.61
			LPM	0.235	0.085	-0.086	0.235	-160.83	58.31	0.62
			the proposed method	0.235	0.086	-0.086	0.235	-160.65	58.24	**0.47**
2	259	102	Ground truth	0.996	-0.074	0.076	0.998	52.27	-42.88	——
			RANSAC	0.995	-0.073	0.076	0.998	51.93	-43.04	0.39
			GTM	0.993	-0.073	0.075	0.997	52.64	-42.76	0.45
			LPM	0.992	-0.073	0.078	0.997	52.6	-42.87	0.58
			the proposed method	0.995	-0.073	0.078	0.997	52.25	-42.98	**0.30**
3	70	12	Ground truth	1.819	-0.438	0.442	1.772	46.664	56.888	——
			RANSAC	1.817	-0.448	0.443	1.772	48.071	56.621	1.72
			GTM	1.846	-0.434	0.437	1.771	44.581	57.207	4.9
			LPM	1.772	-0.34	0.39	1.825	40.647	56.836	13.82
			the proposed method	1.812	-0.444	0.442	1.774	47.864	56.455	**1.16**
4	177	24	Ground truth	2.999	-0.001	0.008	3.014	-9.559	-6.778	——
			RANSAC	2.998	-0.002	-0.006	2.995	-2.548	3.082	3.59
			GTM	3.05	0.011	0.016	2.978	-24.148	1.048	6.74
			LPM	2.726	-0.154	-0.375	2.679	114.082	203.581	65.792
			the proposed method	2.999	-0.001	0.001	2.997	-3.848	0.218	**2.94**
5	59	15	Ground truth	1.998	-0.104	-0.015	2.004	18.428	-13.749	——
			RANSAC	1.977	-0.085	-0.02	2.015	17.608	-13.732	2.816
			GTM	1.991	-0.106	-0.011	2.001	18.76	-12.55	1.778
			LPM	2.022	-0.155	0.136	1.913	14.892	-71.704	36.497
			the proposed method	1.993	-0.109	-0.012	2.004	18.735	-13.184	**1.222**
6	22	8	Ground truth	0.5015	0.027	0.0015	0.5011	-10.543	1.5112	——
			RANSAC	0.505	0.027	0.0025	0.499	-12.242	1.691	1.546
			GTM	0.5	0.0256	0.0031	0.497	-10.0948	2.894	1.8
			LPM	0.494	-0.012	-0.005	0.487	-15.235	14.274	9.386
			the proposed method	0.502	0.0262	0.0023	0.5	-11.174	0.892	**0.569**

and LPM are presented. Finally, we discuss the advantages and disadvantages of our method based on the experimental results.

Table 2 summarizes the initial matches extracted by SIFT, initial correct matches, ground truth and the deformation parameters calculated by image registration algorithm with RANSAC, GTM, LPM and our method respectively for image pair 1-6. Image pair 1 is a pair of simulated remote sensing images to verify the feasibility of the proposed method. It can be found in Table 2 that all these four methods perform well in image pair 1. Besides, the root mean square error(RMSE) of the proposed method is smaller than the other methods. The other five pairs of images are real remote sensing images. It can be seen in Table 2 that there are large number of the initial matches and the initial correct matches in image pair 2, which makes outlier removal easily. Therefore, the root mean square error(RMSE) of these four

methods are small. Also, the root mean square error(RMSE) of the proposed method is smaller than the other methods.

For image pair 3-6, the scale difference of the reference image and the sensed image becomes larger and larger, the number of the initial matcher and initial correct matches decrease rapidly in Table 2, which causes outlier removal difficult. In this case, there are no feature points in the neighborhood of most feature points. Then LPM may perform unsatisfactory. But the proposed method can get better registration results in Table 2, which demonstrates the accuracy and robustness of the proposed method.

The precision and recall of four methods for six pairs of remote sensing images are shown in Figure 3 (a)and (b), respectively. For the simulated experiment, we can see from the data presented in Tables 2 and Figure 2 that the precision of our method is higher than that of other three methods, leading to a smaller RMSE value. This indicates that our
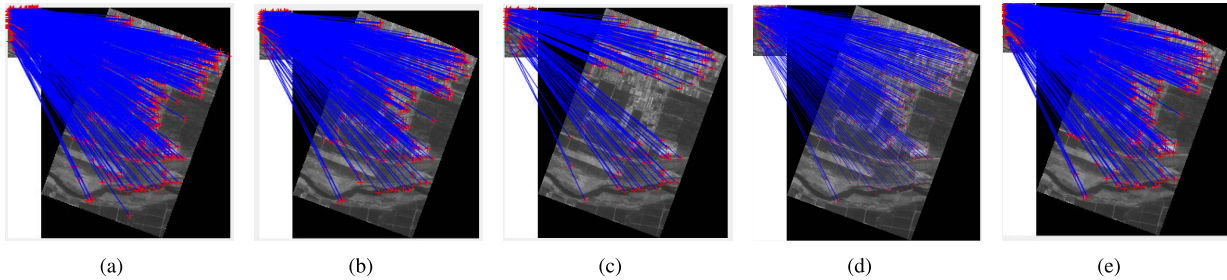
FIGURE 3. Feature point matching result of image pair 1: (a) SIFT. (b) SIFT+RANSAC. (c) SIFT+GTM. (d) SIFT+LPM. (e) SIFT+the proposed method.
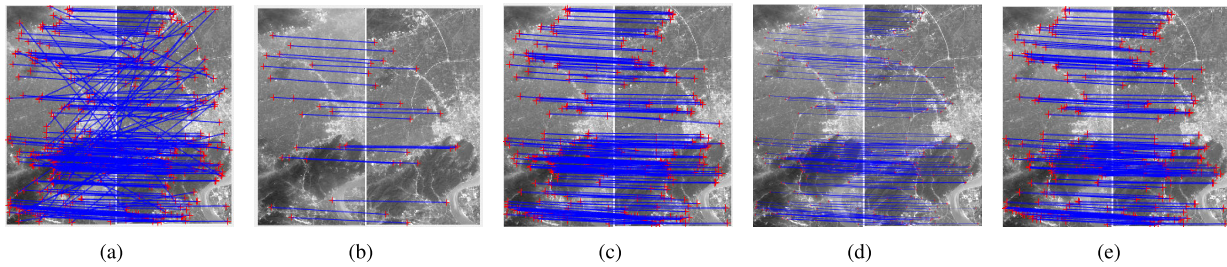

FIGURE 4. Feature point matching result of image pair 2: (a) SIFT. (b) SIFT+RANSAC. (c) SIFT+GTM. (d) SIFT+LPM. (e) SIFT+the proposed method.

method is able to eliminate more outliers from the initial feature points and obtain more accurate registration results. The Recall of our method in image pair-1 is higher than GTM and a little smaller than that calculated by RANSAC and LPM. This illustrates that our method also removes some inliers when there are many initial feature points, but RANSAC and LPM preserve more inliers and outliers. Large recall and small precision of the LPM method can also illustrate that the the root mean square error(RMSE) of is very large in image pair 3-6, because more outliers are preserved and reduce the registration accuracy.

For analysis of the real experiment of image pair-2, it can be seen that the precision of our method is higher than that other three methods, illustrating that the feature points preserved by our method includes fewer outliers, consistent with the low RMSE of our method. The Recall of our method in image pair-2 is also higher than the value for other three methods, illustrating the fact that our method can preserve many inliers. From the real experiments of image pair 3-6, the precision of our method is significantly higher than other three methods, the recall is much higher than RANSAC and GTM and a little smaller than LPM, even for a small number of initial feature points and a low the proportion of inliers in initial feature points, corresponding to the low RMSE of our method. This indicates that our method is robust to the number of initial feature points and the proportion of inliers in the set of initial feature points.

The overall trend of the results, indicates that the RMSE of the three algorithms becomes larger as the geometric differences of real image increases. For the results using our algorithm, Data set 4 has the largest RMSE, the corresponding precision is the smallest, and the recall is the largest.

In combination with the data presented in Tables 2, the initial matching points have the lowest precision, indicating that the original matching feature points contains the most outlier points, corresponding to the result of Figure 2.

The feature point matching results of the six pairs of remote sensing images preserved by different feature point matching algorithms are presented in Fig 3-8. Fig 3-8 are feature point matching results of image pair 1-6. Each Fig includes five images (a-e), where (a) is point correspondences of the initial feature point set by SIFT, (b) is point correspondences preserved by RANSAC, (c) is point correspondences preserved by GTM, (d) is point correspondences preserved by LPM, (e) is point correspondences preserved by the proposed method. Comparing the feature points of our method and the initial feature point set, we can find that our method removes most of the outliers and preserve most of the inliers from the initial feature point set. In general, it is apparent to find that the number of matching point pairs using the proposed method is always larger than the number of matching pairs preserved by other methods, regardless of the differences in the image pairs. Only one pair of simulated images are compared, considering only the influence of geometrical differences on the registration result. Owing to the simulated data set 1 are acquired from the same band, the number of matching feature points is large, as can be seen from Figure 3. For real images, as the image scale difference increases, the number of final matching feature point pairs is significantly reduced. However, our method can still preserve many feature point pairs.

The results intuitively illustrate that this method can identify preserve more feature point pairs, even for images with large scale difference and a small number of initial matching
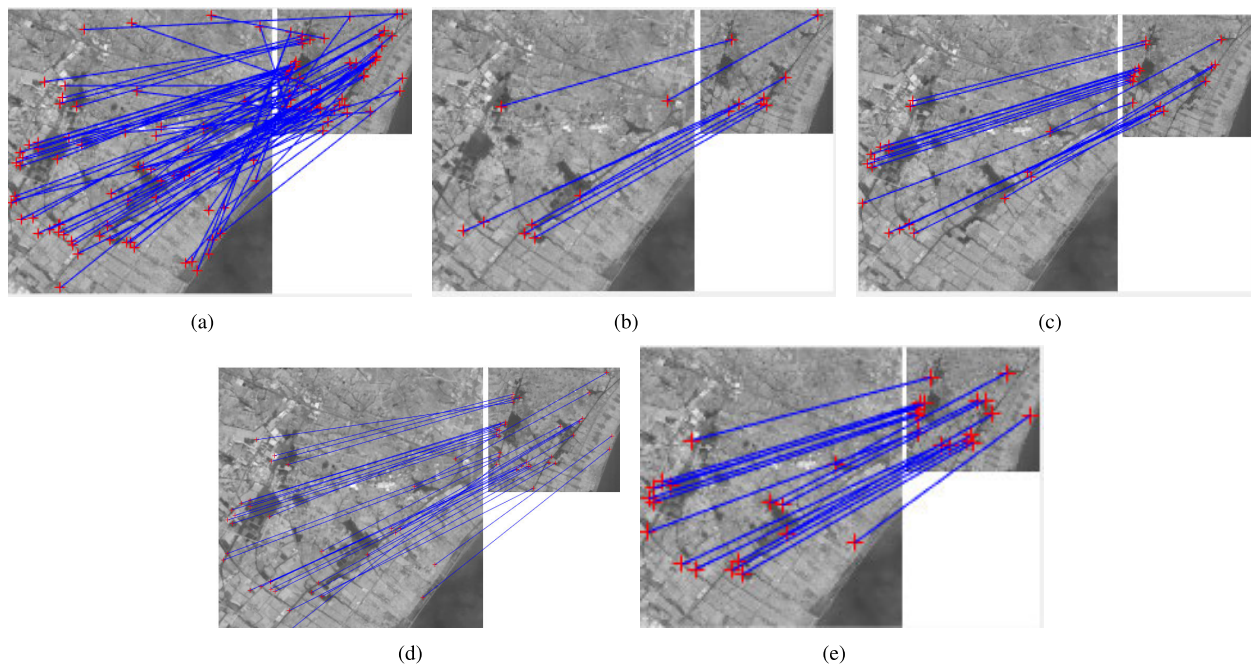
FIGURE 5. Feature point matching result of image pair 3: (a) SIFT. (b) SIFT+RANSAC. (c) SIFT+GTM. (d) SIFT+LPM. (e) SIFT+the proposed method.
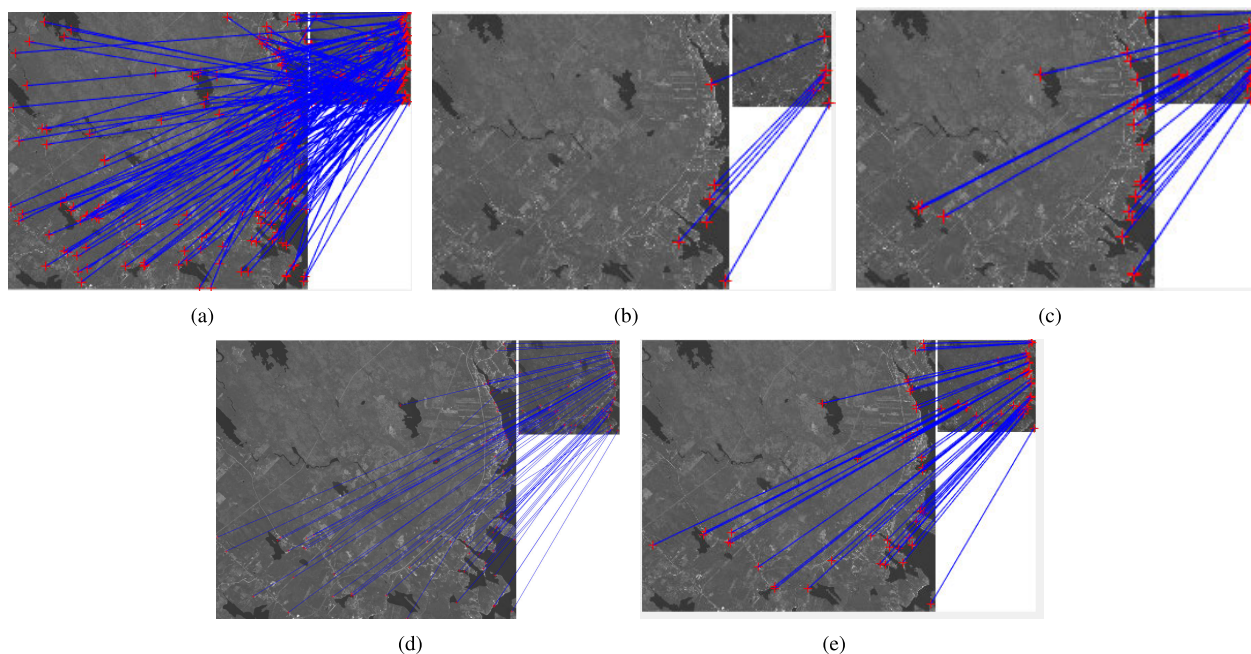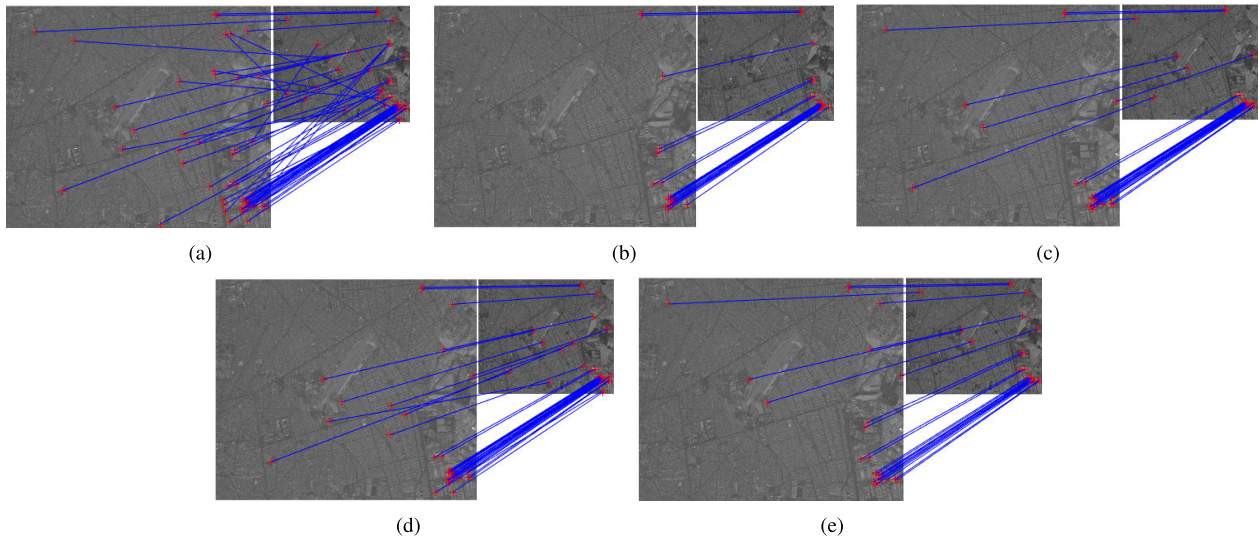

FIGURE 6. Feature point matching result of image pair 4: (a) SIFT. (b) SIFT+RANSAC. (c) SIFT+GTM. (d) SIFT+LPM. (e) SIFT+the proposed method.

feature points. Observing the feature point matching result, it can be found that the proposed method can preserve more correct matching point pairs, and the distribution of the feature matching points is more spatially consistent, especially for the real data-3.

Image registration results of Image pair 1-6 registered by the proposed method are shown in Figure9. Observing the

roads, rivers,etc. in image registration results, it can be found visually that the registration error of the proposed method is small in these five pairs of remote sensing images.

VI. DISCUSSION
The distance ration threshold d_{radio} and λ are main parameters in the proposed model. The distance ration threshold is the

FIGURE 7. Feature point matching result of image pair 5: (a) SIFT. (b) SIFT+RANSAC. (c) SIFT+GTM. (d) SIFT+LPM. (e) SIFT+the proposed method.
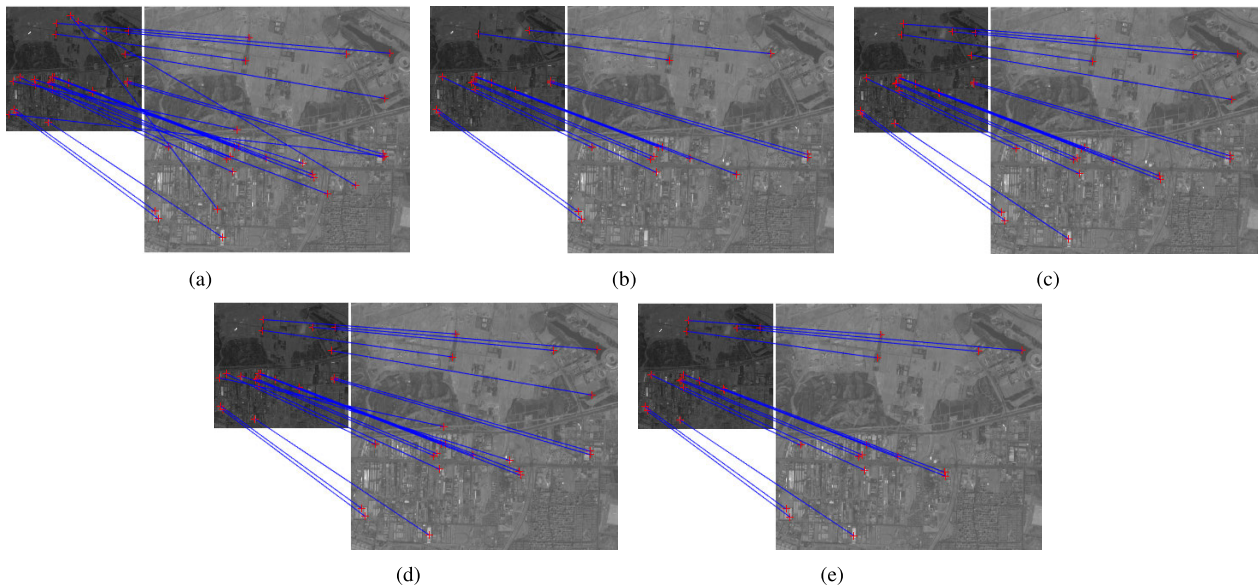

FIGURE 8. Feature point matching result of image pair 6: (a) SIFT. (b) SIFT+RANSAC. (c) SIFT+GTM. (d) SIFT+LPM. (e) SIFT+the proposed method.

ratio of closest to second-closest neighbors of each feature point. As the distance ration threshold increases, the number of matching feature points will increase. Meanwhile, the number of outlier will also increase because larger ration threshold allows more outliers matching. Through the paper (Distinctive image feature from scale-invariant keypoints [13]) and experiments, we found when $d_{radio} = 0.8$, the number of matching feature points is large, and the proportion of outliers is small. Therefore the $d_{radio} = 0.8$ is suitable for experiments.

λ is the parameter to controls the tradeoff between the inlier set and the outlier set. This parameter is the threshold to decide whether the feature point is outlier. While the objective function is calculated through all feature points, so the

threshold should be affected by the number of feature points. Because of the loop in the proposed method, the threshold λ will affect the number of cycles a lot, and affect the precision a little. Through several experiments, we found when $\lambda = N/100$, the precision and the speed of the proposed are best.

Feature point matching through SIFT mainly relies on local similarity of feature points, ignoring the global geometric similarity. The proposed method makes full use of the global geometric similarity to remove outliers, which is not affected by the number of feature points. Compared to RANSAC and LPM, the proposed method can perform well regardless of the small number of matches and uneven distribution, such as experiment 4-6. Compared to GTM, the proposed method considers the geometric relationship of all the feature points,
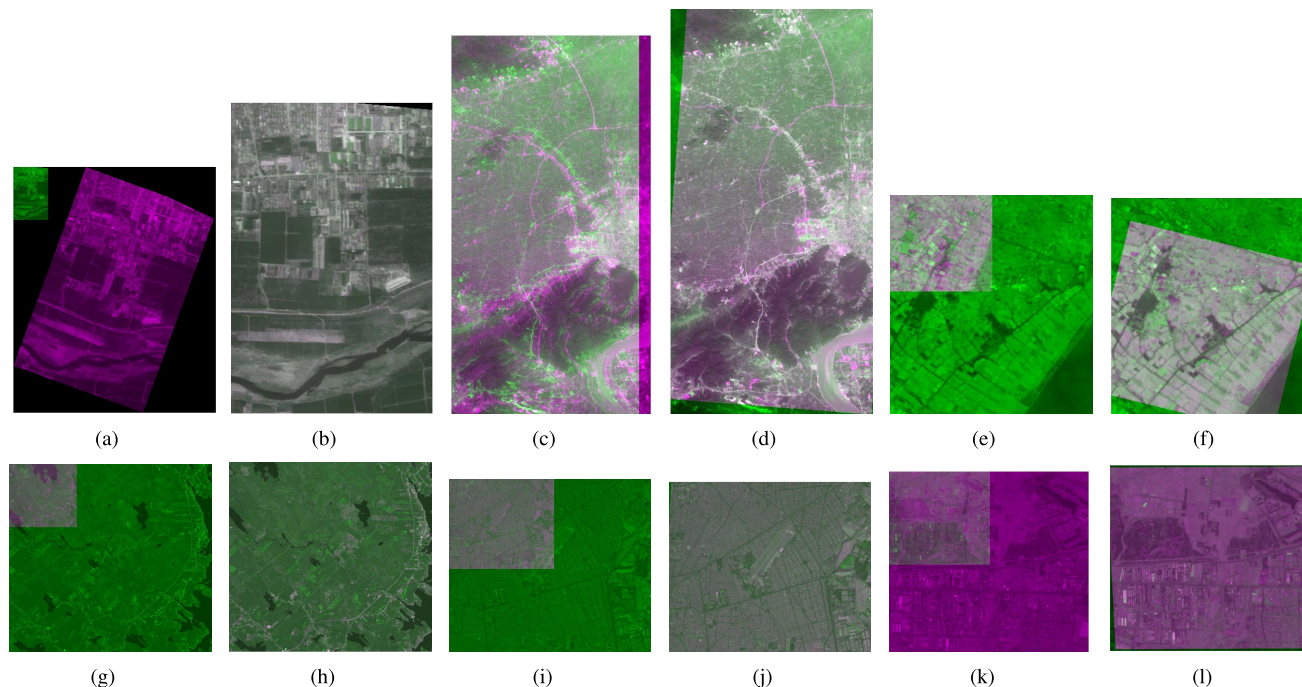
FIGURE 9. Image registration results, image (a-b) correspondence to the registration result of Image pair 1, (a) overlapped image before the registration, (b) the difference after the registration. Image (c-d) correspondence to the registration result of Image pair 2, (c) overlapped image before the registration, (d) the difference after the registration. Image (e-f) correspondence to the registration result of Image pair 3, (e) overlapped image before the registration, (f) the difference after the registration. Image (g-h) correspondence to the registration result of Image pair 4, (g) overlapped image before the registration, (h) the difference after the registration. Image (i-j) correspondence to the registration result of Image pair 5, (i) overlapped image before the registration, (j) the difference after the registration. Image (k-l) correspondence to the registration result of Image pair 6, (k) overlapped image before the registration, (l) the difference after the registration.

while GTM only considers local geometric relationship of feature points. From experiments we can find the precision and robustness of the proposed method is higher than GTM. Therefore, the proposed method is a high precision and robustness image registration method.

VII. CONCLUSION

In this paper, a high precision outlier elimination method is presented for feature matching of remote sensing images. The proposed method is based on the similar global geometric relationship of feature points between two images containing the same scene. A mathematic model was formulated according to this principle. We found the solution of the model by analyzing the optimizing the formula. The feature-matching algorithm based on outlier removal was described based on the solution of the mathematic model. The experimental results of several remote sensing images demonstrate that our outlier elimination method can remove a large number of outliers and preserve many inliers, leading to better registration results. However, the proposed method also has a weakness, which is high time complexity. We will optimize the calculation of the algorithm in the future.

REFERENCES

[1] Y. Dong, T. Long, W. Jiao, G. He, and Z. Zhang, "A novel image registration method based on phase correlation using low-rank matrix factorization with mixture of Gaussian," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 446–460, Jan. 2018.

[2] S. Chen, X. Li, L. Zhao, and H. Yang, "Medium-low resolution multisource remote sensing image registration based on SIFT and robust regional mutual information," *Int. J. Remote Sens.*, vol. 39, no. 10, pp. 3215–3242, Jan. 2018.

[3] Y. Han, F. Bovolo, and L. Bruzzone, "Segmentation-based fine registration of very high resolution multitemporal images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2884–2897, May 2017.

[4] W. Ma, Y. Wu, S. Liu, Q. Su, and Y. Zhong, "Remote sensing image registration based on phase congruency feature detection and spatial constraint matching," *IEEE Access*, vol. 6, pp. 77554–77567, 2018.

[5] X. Li, "High-accuracy subpixel image registration with large displacements," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6265–6276, Nov. 2017.

[6] Z. Li, D. Mahapatra, J. A. W. Tielbeek, J. Stoker, L. J. van Vliet, and F. M. Vos, "Image registration based on autocorrelation of local structure," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 63–75, Jan. 2016.

[7] J. Fan, Y. Wu, M. Li, W. Liang, and Y. Cao, "SAR and optical image registration using nonlinear diffusion and phase congruency structural descriptor," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5368–5379, Sep. 2018.

[8] F. Luo, B. Du, L. Zhang, L. Zhang, and D. Tao, "Feature learning using spatial-spectral Hypergraph discriminant analysis for hyperspectral image," *IEEE Trans. Cybern.*, vol. 49, no. 7, pp. 2406–2419, Jul. 2019.

[9] L. Zhang, L. Zhang, B. Du, J. You, and D. Tao, "Hyperspectral image unsupervised classification by robust manifold matrix factorization," *Inf. Sci.*, vol. 485, pp. 154–169, Jun. 2019.

[10] D. Yu, F. Yang, C. Yang, C. Leng, J. Cao, Y. Wang, and J. Tian, "Fast rotation-free feature-based image registration using improved N-SIFT and GMM-based parallel optimization," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 8, pp. 1653–1664, Aug. 2016.

[11] G.-L. Wu and H.-H. Chang, "An accurate feature point matching algorithm for automatic remote sensing image registration," in *Proc. IEEE Int. Conf. Digit. Image Comput., Techn. Appl. (DICTA)*, Nov./Dec. 2017, pp. 1–8.

[12] M. Zhao, B. An, Y. Wu, H. Van Luong, and A. Kaup, "RFVTM: A recovery and filtering vertex trichotomy matching for remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 375–391, Jan. 2017.

[13] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[14] Z. Hossein-Nejad and M. Nasri, "RKEM: Redundant keypoint elimination method in image registration," *IET Image Process.*, vol. 11, no. 5, pp. 273–284, 2017.

[15] J. Ma, J. Zhao, J. Tian, A. L. Yuille, and Z. Tu, "Robust point matching via vector field consensus," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1706–1721, Apr. 2014.

[16] J. Ma, H. Zhou, J. Zhao, Y. Gao, J. Jiang, and J. Tian, "Robust feature matching for remote sensing image registration via locally linear transforming," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6469–6481, Dec. 2015.

[17] J. Ma, J. Jiang, H. Zhou, J. Zhao, and X. Guo, "Guided locality preserving feature matching for remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4435–4447, Aug. 2018.

[18] J. Ma, J. Zhao, J. Jiang, H. Zhou, and X. Guo, "Locality preserving matching," *Int. J. Comput. Vis.*, vol. 127, no. 5, pp. 512–531, 2019.

[19] O. Chum, J. Matas, and J. Kittler, "Locally optimized RANSAC," in *Pattern Recognition* (Lecture Notes in Computer Science), vol. 2781. Berlin, Germany: Springer, 2003, pp. 236–243.

[20] P. H. S. Torr and A. Zisserman, "MLESAC: A new robust estimator with application to estimating image geometry," *Comput. Vis. Image Understand.*, vol. 78, no. 1, pp. 138–156, 2000.

[21] O. Chum and J. Matas, "Matching with PROSAC—Progressive sample consensus," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 220–226.

[22] J. Yang, Q. Huang, B. Wu, and J. Chen, "A remote sensing imagery automatic feature registration method based on mean-shift," in *Proc. Geosci. Remote Sens. Symp.*, 2012, pp. 2364–2367.

[23] S. Cao, J. Jiang, G. Zhang, and Y. Yuan, "An edge-based scale- and affine-invariant algorithm for remote sensing image registration," *Int. J. Remote Sens.*, vol. 34, no. 7, pp. 2301–2326, 2013.

[24] Y. Yu, K. Huang, W. Chen, and T. Tan, "A novel algorithm for view and illumination invariant image matching," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 229–240, Jan. 2012.

[25] W. Aguilar, Y. Frauel, F. Escolano, M. E. Martinez-Perez, A. Espinosa-Romero, and M. A. Lozano, "A robust graph transformation matching for non-rigid registration," *Image Vis. Comput.*, vol. 27, no. 7, pp. 897–910, Jun. 2009.

[26] J. Ma, X. Jiang, J. Jiang, J. Zhao, and X. Guo, "LMR: Learning a two-class classifier for mismatch removal," *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 4045–4059, Aug. 2019.

[27] W. Ma, Z. Wen, W. Yue, L. Jiao, and L. Liang, "Remote sensing image registration with modified sift and enhanced feature matching," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 1, pp. 3–7, Jan. 2017.

[28] Z. Liu, J. An, and Y. Jing, "A simple and robust feature point matching algorithm based on restricted spatial order constraints for aerial image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 514–527, Feb. 2012.
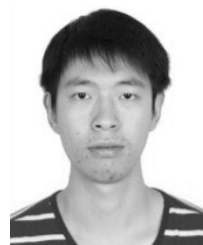
HAN YANG received the B.S degree in automation from Zhejiang University, Hangzhou, China, in 2016, where he is currently pursuing the Ph.D. degree in control theory and control engineering. His research interests include image processing and image registration.

XIAORUN LI received the B.S. degree from the National University of Defense Technology, Changsha, China, in 1992, and the M.S. and Ph.D. degrees from Zhejiang University, Hangzhou, China, in 1995 and 2008, respectively. Since 1995, he has been with Zhejiang University, where he is currently a Professor with the College of Electrical Engineering. His research interests include hyperspectral image processing, signal and image processing, and pattern recognition. His research interests included image processing and image registration.
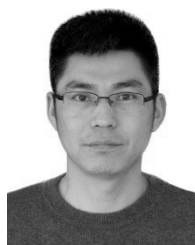
YIJIAN MA received the B.S. degree from the Chongqing University, Chongqing, China, in 2009. Since 2002, he has been a Senior Engineer with the Zhejiang Academy of Special Equipment Science. His main research interest includes lifting appliances safety technology research and inspection.

LIAOYING ZHAO received the B.S. and M.S. degrees from Hangzhou Dianzi University, Hangzhou, China, in 1992 and 1995, respectively, and the Ph.D. degree from Zhejiang University, Hangzhou, in 2004. Since 1995, she has been with Hangzhou Dianzi University, where she is currently a Professor with the College of Computer Science. Her research interests include hyperspectral image processing, signal and image processing, pattern recognition, and machine learning.

SHUHAN CHEN received the B.S. degree from Ludong University, Yantai, China, in 2011, and the M.S. degree from Liaoning Technical University, Huludao, China, in 2014. She is currently pursuing the Ph.D. degree in control theory and control engineering with Zhejiang University, Hangzhou, China, and doing research as a visiting faculty Research Assistant in the Department of Computer Science and Electrical Engineering, University Maryland, Baltimore County, Baltimore, USA. Her research interests include image registration, hyperspectral image processing, and patter recognition.

- - -