Robust Feature Matching for Remote Sensing Image Registration via Guided Hyperplane Fitting

Guobao Xiao, *Member, IEEE*, Huan Luo, Kun Zeng, *Member, IEEE*, Leyi Wei, *Member, IEEE*, and Jiayi Ma, *Member, IEEE*

***Abstract*— Feature matching is a fundamental problem in feature-based remote sensing image registration. Due to the ground relief variations and imaging viewpoint changes, remote sensing images often involve local distortions, leading to difficulties in high-accuracy image registration. To address this issue, in this article, we propose a robust feature matching method called First Neighbor Relation Guided (FNRG) for remote sensing image registration via guided hyperplane fitting. The key idea of FNRG is to exploit the first neighbor relation of feature points between two images for seeking consistent seeds in a parameter-free manner. To boost more consistent matches based on the consistent seeds, we formulate the feature matching problem into an affine hyperplane fitting problem by imposing the motion consistency, and then we design a hyperplane updating strategy to refine the fitting model. We also introduce a locality preserving structure-based cost function to promote the matching performance of the hyperplane updating strategy. Our method can mine consistent matches from thousands of putative ones within only a few milliseconds, and it also can handle the data with a large-scale change, rotation, or severe nonrigid deformation. Extensive experiments on the remote sensing image data sets with different types of image transformations show that the proposed method achieves significant superiority over several state-of-the-art methods.**

***Index Terms*— Feature matching, first neighbor relation, guided feature matching, remote sensing.**

I. Introduction

FEATURE matching, which refers to establishing reliable correspondences between two images of the same regions, is a fundamental and crucial issue in remote sensing and photogrammetry, and it has been widely used in a variety of applications, including image registration and fusion, 3-D reconstruction, Simultaneous Localization And Mapping (SLAM), and image retrieval [1]–[4].

Generally, the feature matching problem includes two steps, i.e., putative match generation and match selection. The first step is usually performed by simply picking out local key-point pairs with similar feature descriptors such as Scale Invariant Feature Transform (SIFT) [5]. However, the generated putative matches often contain a number of false matches (also called mismatches, i.e., outliers) besides the true ones (i.e., inliers) due to various problems, e.g., local key-point localization errors and ambiguities of the local descriptors. Therefore, it is critical to remove mismatches from the generated putative ones in the second step.

A number of mismatch removal methods have been proposed in the past few decades [6]–[15], and most of them assume that the putative matches satisfy an underlying geometrical transformation model (e.g., homography, epipolar geometry, and affine) and then they adopt a geometric constraint to remove mismatches that do not abide by the assumption. However, compared with traditional image registration, remote sensing image registration is more complex and challenging. The reason behind this is that, remote sensing images often suffer from local distortions since these type of images may be captured at low-altitude, which will lead to the ground relief variations and imaging viewpoint changes [16]–[21]. That is, the spatial relationship between image pairs will be more complex and the geometrical transformation model becomes unpredictable, especially for the severe nonrigid deformation. Remote sensing images also often use only local descriptor information to generate putative matches, which inevitably contain a large number of false matches. In addition, the high computational complexity is also a challenging issue due to complex nonrigid transformation models.

To address these issues, we propose a First Neighbor Relation Guided feature matching method (called FNRG) for remote sensing image registration. The proposed method is based on the assumption that a low-dimensional hyperplane can be adopted to estimate many types of deformations for outlier removal, which has been used in [22] and [23]. However, a hyperplane may not be sufficient to remove outliers for remote sensing images due to the ground relief variations and imaging viewpoint changes. As shown in Fig. 1(b), we project the data onto the first-3 principal components from the original

Manuscript received August 17, 2020; revised November 10, 2020; accepted November 25, 2020. Date of publication December 9, 2020; date of current version December 2, 2021. This work was supported by the National Natural Science Foundation of China under Grant 62072223, Grant 61702431, and Grant 61773295. *(Corresponding author: Jiayi Ma.)*

Guobao Xiao is with the Fujian Provincial Key Laboratory of Information Processing and Intelligent Control, College of Computer and Control Engineering, Minjiang University, Fuzhou 350108, China and also with The State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710126, China (e-mail: x-gb@163.com).

Huan Luo is with the College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350108, China (e-mail: hluo@fzu.edu.cn).

Kun Zeng is with the College of Computer and Control Engineering, Minjiang University, Fuzhou 350108, China (e-mail: zengkun301@aliyun.com).

Leyi Wei is with the School of Software, Shandong University, Jinan 250100, China (e-mail: weileyi@tju.edu.cn).

Jiayi Ma is with Electronic Information School, Wuhan University, Wuhan 430072, China (e-mail: jyma2010@gmail.com).

Digital Object Identifier 10.1109/TGRS.2020.3041270

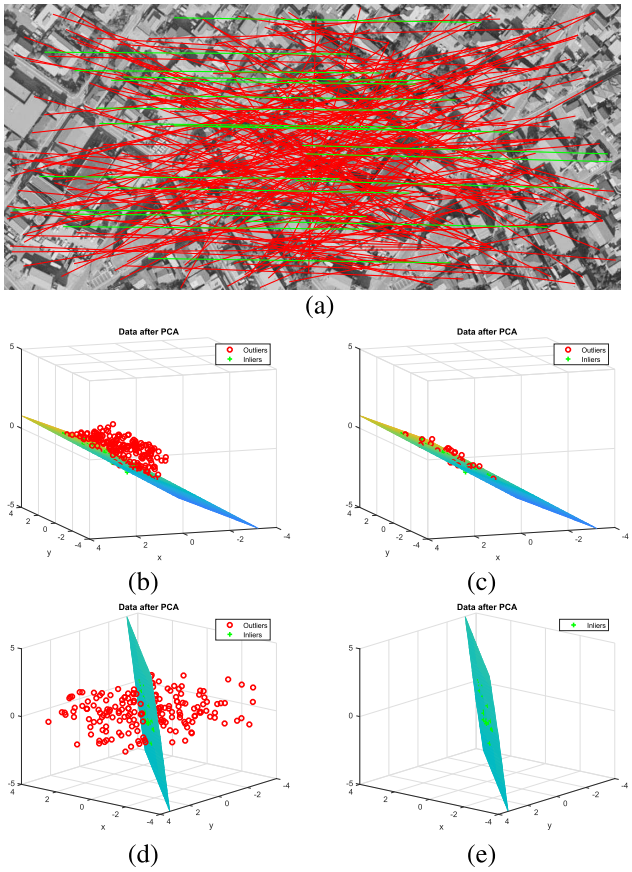

Fig. 1. Illustration of hyperplane with the motion consistency in our method. (a) We show the input matches, which include 200 matches with 20 inliers (green = inliers, red = outliers). (b) We project the data onto the first-3 principal components without the motion consistency and show the estimated inliers in (c). (d) We project the data onto the first-3 principal components with the motion consistency and show the estimated inliers in (e).

data space [see Fig. 1(a)], which is generated by the image coordinates of each two feature points of correspondences, and both inliers and many outliers are distributed on an affine hyperplane [see Fig. 1(c)], making no sense for outlier removal. Thus, we introduce the motion consistency, which is captured by the motion vector of each correspondence, to the projection. The motion consistency is based on the observation that inliers tend to have similar motion behavior, while outliers are randomly distributed across the image domain [24]. To illustrate this observation, we show the motion field of correspondences from Fig. 1(a) in Fig. 2 (we show more examples in Fig. 10). We can see that inliers always tend to have coherent motions, and they share a similar motion with their neighbors; while the outliers tend to be randomly distributed across the motion field. This is because, for two inliers $s_1 = (x_1, y_1)$ and $s_2 = (x_2, y_2)$ from a scene, they will not only have similar lengths, i.e., $|y_1 - x_1|$ and $|y_2 - x_2|$, but also have similar directions, i.e., $y_1 - x_1$ and $y_2 - x_2$; in contrast, for two outliers, they do not have that property.

Then, for a correspondence $s_i = (x_i, y_i) \in \mathbb{R}^4$, we add the motion vector $m_i = y_i - x_i$ into s_i, i.e., $s_i = (x_i, y_i, m_i) \in \mathbb{R}^6$. From Fig. 1(d) and (e), we can see that the inliers will be actually distributed compactly on an affine hyperplane while outliers are distributed randomly. The reason behind this is that

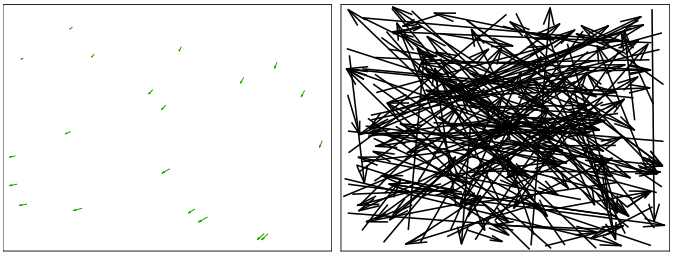

Fig. 2. Illustration of the motion consistency in our method. We show arrows in the motion field corresponding to inliers (Left) and outliers (Right), respectively. The head and tail of each arrow correspond to the positions of two corresponding feature points in the image pair.

the spatial relationship between image pairs for remote sensing images is more complex, and the motion vector can reduce the relationship, and meanwhile, we put the correspondences into a higher-dimensional space, which will significantly increase its separability. Thus, the outlier removal problem for remote sensing images can be formulated as an affine hyperplane fitting problem by the projection with the motion consistency.

To fit the affine hyperplane, we can directly use the popular fitting method, i.e., RANdom SAmple Consensus (RANSAC) [6]. However, RANSAC cannot obtain consistent results due to its randomness, and it also cannot deal with the data with a large proportion of outliers. Even worse, currently existing deterministic fitting methods, such as [25]–[27], cannot be directly used to deal with remote sensing images since they often suffer from high computational complexity.

Fortunately, if we obtain some small number of (at least three) inliers in advance, then the affine hyperplane will be directly estimated. This is because we can use a single and global affine model to fully characterize the given correspondences. To this end, we introduce the local neighborhood structure information to seek a small number of consistent seeds. The local neighborhood structure information has been preserved to deal with different problems [1], [2], [11], [28] due to its efficiency and simplicity. As discussed above, remote sensing images often involve the ground relief variations and imaging viewpoint changes, then the absolute distance between two feature points may change while the spatial neighborhood relationship is generally well preserved. Thus, the local neighborhood structure information can be exploited to seek a small number of consistent seeds.

However, how to define the "local" neighborhood often depends on some parameters or thresholds, which will restrict the generalization of the proposed methods. Unlike [1], [2], [11], which use the local neighborhood to seek all inliers, we only require to seek a small number of seeds. Thus, we propose to preserve the local neighborhood structure by using the first neighbor information. That is, we do not require any parameter or threshold in advance except for the "first" one. Specifically, we compute the first (i.e., 1-nearest) neighbor of each feature point in two images under the Euclidean distance. Then, we merge the feature points from the two images into clusters by using the first neighbor information, respectively. After that, we count the number of correspondences that belong to the same cluster in both two views. For example in Fig. 3, Clusters A and B share three correspondences, and Clusters C and D share four correspondences, and

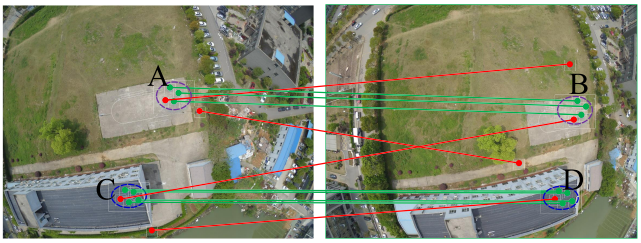

Fig. 3. Illustration of consistent seed seeking strategy based on first neighbor relation for remote sensing images. The point with green and red color represents an inlier and outlier, respectively. A, B, C, and D are four clusters of features points.

Clusters B and C share one correspondence. We can see that, if correspondence is an inlier, it will have more good neighbors whose corresponding feature points belong to the same cluster in both views. Thus, we select the correspondences that have the best neighbors as the consistent seeds.

Then, we estimate an affine hyperplane based on the sought seeds by imposing the motion consistency. To further improve the effectiveness of the estimated affine hyperplane, we introduce a hyperplane updating strategy to refine the hyperplane, where we introduce a locality preserving structure-based cost function to measure the quality of hyperplane models. After that, we remove outliers from all correspondences according to the estimated affine hyperplane.

More concretely, we summarize the contributions of this article as follows:

1) We propose a simple but effective feature matching method to remove mismatches for remote sensing images by fitting an affine hyperplane, where we introduce the motion consistency and a hyperplane updating strategy to promote the matching performance.
2) We propose a novel first neighbor relation-based strategy to seek consistent seeds for remote sensing images. It is worth pointing out that the proposed strategy not only provides a small number of seeds from inliers but it also does not require to provide any parameter or threshold in advance.
3) The proposed feature matching method is able to mine consistent matches from thousands of putative ones within only a few milliseconds. Extensive experiments on the remote sensing image data sets with different types of image transformations also demonstrate that the proposed method can achieve significant superiority over several state-of-the-art methods.

The rest of this article is organized as follows. The proposed method is described in detail in Section III. Experimental results are given in Section IV. Finally, discussion and conclusion are made in Section V.

II. Related Work

In this section, we introduce some feature matching methods that are related to the proposed method. A general feature matching method involves two steps: 1) putative match generation and 2) match selection (also called mismatch or outlier removal).

For the first step, some popular methods, e.g., SIFT, speeded up robust features (SURF) [29] and oriented FAST and rotated BRIEF (ORB) [30], are proposed to detect feature points. SIFT adopts gradient histogram (that is, it compares the distance ratio between the nearest and the second nearest neighborhoods against a threshold) to form descriptors after detecting feature points. SURF introduces the Hessian matrix and an integral image strategy for effectiveness and efficiency, respectively. ORB improves SIFT by using a FAST detector and BRIEF descriptor for a faster speed. There are also some other methods that are proposed to improve the effectiveness or efficiency, but the putative matches are inevitably contaminated by a large population of mismatches due to the use of only local appearance feature.

Therefore, numerous mismatch removal methods have been proposed to address the second step. These methods can be roughly categorized into three groups, i.e., learning-based methods [12], [31], [32], parametric methods [6]–[8], [33] and nonparametric methods [1], [2], [10], [11], [34].

The deep learning techniques [35]–[38] have achieved great success in recent years, thus, some authors introduce it to address the feature matching, e.g., [12], [31], [32]. Learning to Find Good Correspondences (LFGC) [31] is the first one to introduce the PointNet-like architecture to construct a multilayer perceptron-based deep learning framework for outlier removal. Although LFGC is able to achieve appealing solutions, it ignores useful local information. Then, Mining reliable Neighbors Network (MN-Net) [12] improves LFGC by adding the local information derived from a compatibility-specific neighbor mining algorithm to the deep learning framework. Order-Aware Network (OA-Net) [32] also proposes an improved version called OA-Net, by inferring the probabilities of the input matches being inliers. These methods can address the matching problems, however, just as other data-driven methods, they cannot guarantee their performance on the correspondences that are not used for training, which restricts the applications in the real world.

Most of the current feature matching methods, e.g., [6]–[8], [33], use a geometric constraint, such as homography, epipolar geometric, and affine, to remove outliers. They often use a "hypothesis-and-verify" framework (i.e., sample minimal subsets for hypothesis generation and then verify the generated hypotheses). RANSAC [6] is one of the most popular methods due to its simplicity and effectiveness. These methods require to provide the defined model in advance, but remote sensing images often suffer from local distortions, which may result in the ineffectiveness of the methods.

There are some nonparametric methods, e.g., [1], [2], [10], [11], [34], which exploit the relationship between matches to address the general feature matching problem. For example, coherence-based decision boundaries [10] proposes to exploit the coherence-based separability constraints to remove outliers, and grid-based motion statistics (GMS) [34] exploits the motion smoothness constraints to do that. Locality Preserving Matching (LPM) [11] proposes to exploit the local neighborhood information to seek consistent matches. Guided LPM (GLPM) [2] improves LPM by introducing a guided matching strategy for effectiveness. Multi-scale top k rank preservation (mTopKRP) [1] also improves LPM by introducing ranking lists derived from the local neighborhood information for effectiveness.

There are also some transform-based methods, such as, [39]–[41], for image registration. Zavorin and Moigne [41] developed an automatic image registration technique based on wavelets and wavelet-like pyramids. Alam *et al.* [40] developed a conditional entropy-based objective function based on a probabilistic model of the curvelet coefficients of images. Murphy *et al.* [39] developed a shearlet feature algorithm to produce distinct features for automatic image registration.

The proposed method (FNRG) is one of the parametric methods, however, unlike most of them, which involves random nature due to the randomly sampling, FNRG can provide consistent solutions by proposing a novel FNRG strategy. FNRG also proposes to introduce the motion consistency to the affine hyperplane model for addressing the general remote sensing image matching problem. It is worth pointing out that, although FNRG also uses a geometric constraint, it fully exploits the relationship between matches by preserving the first neighbor relations for seeking consisting seeds and preserving local neighborhood structure for the cost function of the proposed hyperplane updating strategy.

III. Methodology

In this section, we describe the details of the proposed feature matching method for remote sensing images.

A. Problem Statement

Given two remote sensing images, we extract a set of n putative feature correspondences $S = \{s_i = (x_i, y_i)\}_{i=1}^{n}$ from the images, where $x_i = (x_i^1, x_i^2)$ and $y_i = (y_i^1, y_i^2)$ are the pixel coordinates of the corresponding feature points in the two images, respectively. Then, we decompose each correspondence s_i on a manifold as follows [23]:

$$s_i = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} x_i^1 \\ x_i^2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ a_{13} \\ a_{23} \end{bmatrix}}_{(a)} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ \mu_1 L\left(x_i^1, x_i^2\right) \\ \mu_2 L\left(x_i^1, x_i^2\right) \end{bmatrix}}_{(b)} \tag{1}$$

where the manifold involves (a) a 2-D affine hyperplane and (b) the nonlinear deviation. a_{pq} denotes the qth component of the pth affine parameter vector, and μ_1 and μ_2 are two coefficients, and $L(x_i^1, x_i^2)$ represents a nonlinear lifting function.

Then, given at least three correspondences $s' = [s_1, s_2, s_3]$, we characterize an affine hyperplane by computing the mean value of the pixel coordinates of the given correspondences, and their first-two left singular vectors, i.e., an affine hyperplane $\theta \in \mathbb{R}^{4\times 3}$ is defined as:

$$\theta = [\text{mean}(s') \ D(s')] \tag{2}$$

where $D(s')$ is the first-two left singular vectors of the given correspondences. Then the residual of a correspondence s_i to a subspace $\theta = [\theta_1, \theta_2, \theta_3]$ can be measured by orthogonal distance $r_i(\theta)$:

$$r_i(\theta) = \sqrt{(s_i - [\theta_2 \ \theta_3][\theta_2 \ \theta_3]^T (s_i - \theta_1) + \theta_1)^2}. \tag{3}$$

To improve the effectiveness of the affine hyperplane for remote sensing images, which often contain the ground relief variations and imaging viewpoint changes, we introduce motion consistency to the given correspondence. The motion consistency is able to further increase the separability of each correspondence and reduce the spatial relationship between remote sensing image pairs. Thus, the affine hyperplane will be more effective for outlier removal since two inliers should have similar motion properties, including rotation and scale changes; In contrast, two outliers do not have similar properties.

Specifically, we add the motion vector $m_i = y_i - x_i$ into each correspondence s_i, then the correspondence becomes $s_i = (x_i, y_i, m_i) \in \mathbb{R}^{6\times 1}$ and the affine hyperplane becomes $\theta \in \mathbb{R}^{6\times 3}$. The example in Fig. 1 also can show the effectiveness of the motion consistency.

Therefore, the outlier removal problem for remote sensing images can be formulated as an affine hyperplane θ fitting problem. Then, in the following, we introduce the details of the affine hyperplane fitting for remote sensing images.

B. Proposed Consistent Seed Seeking Strategy

To effectively estimate the affine hyperplane, we first seek some consistent seeds that have a high probability of being an inlier for remote sensing images. Following the work mentioned in [1], [2], [11], [28], where the spatial neighborhood relationship of feature points is generally well preserved to deal with different problems, we further exploit the local neighborhood relationship to seek consistent seeds.

Unlike most existing methods that require some parameters to define "local" structure, we only use the first (i.e., 1-nearest) neighbor of feature points in two images under the Euclidean distance and then merge feature points into clusters to define "local" structure.

Specifically, given a set of n putative feature correspondences $S = \{s_i = (x_i, y_i)\}_{i=1}^{n}$, we compute the first neighbor $\mathcal{N}_{x_i}^1$ and $\mathcal{N}_{y_i}^1$ of two feature points of each s_i based on the corresponding spatial position, respectively. Then, we define two adjacency link matrixes A_x and A_y based on the first neighbor of two views, respectively. Given two integer indices $\mathcal{N}_{x_i}$ and $\mathcal{N}_{x_j}$ of the first neighbor of feature points x_i and x_j (here, i and j are two indices of feature points), the item $A_x(i, j)$ of the adjacency link matrix A_x is defined as

$$A_x(i, j) = \begin{cases} 1, & \text{if } j = \mathcal{N}_{x_i}^1 \text{ or } \mathcal{N}_{x_j}^1 = i \text{ or } \mathcal{N}_{x_i}^1 = \mathcal{N}_{x_j}^1 \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where $A_y(i, j)$ has the similar definition as $A_x(i, j)$.

The adjacency link matrix connects a feature point x_i to its first neighbor x_j by $j = \mathcal{N}_{x_i}^1$, and the first neighbor x_i to the corresponding feature point x_j by $\mathcal{N}_{x_j}^1 = i$, and two feature points x_i and x_j that share the same first neighbor by $\mathcal{N}_{x_i}^1 = \mathcal{N}_{x_j}^1$. Intuitively, the adjacency link matrix can be directly used to construct some clusters. Two feature points are merged into the same cluster when they are the first neighbor of each other or they share the same first neighbor. For example, in Fig. 4, given eight feature points and their first neighbors, the feature points are merged into three clusters. We can see that, x_1 and x_3 are merged since x_3 is the first neighbor of x_1, and x_2 and x_3 are merged since they share the

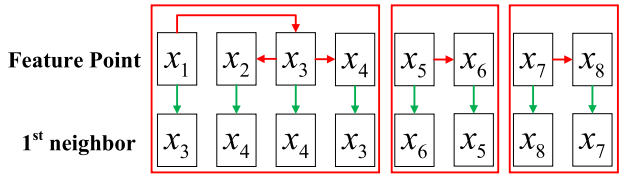

Fig. 4. Illustration of feature point clustering based on first neighbor relation. The green arrow denotes the first neighbor and the red arrow denotes the connected feature point. The red block denotes a cluster.

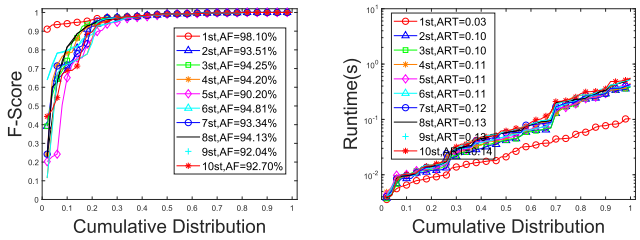

Fig. 5. F-score (Left) and running time (Right) with respect to the cumulative distribution obtained by different settings of the nearest neighbors. A point on the curve with coordinate (x, y) denotes that there are $100 * x$ percents of image pairs which have values no more than y.

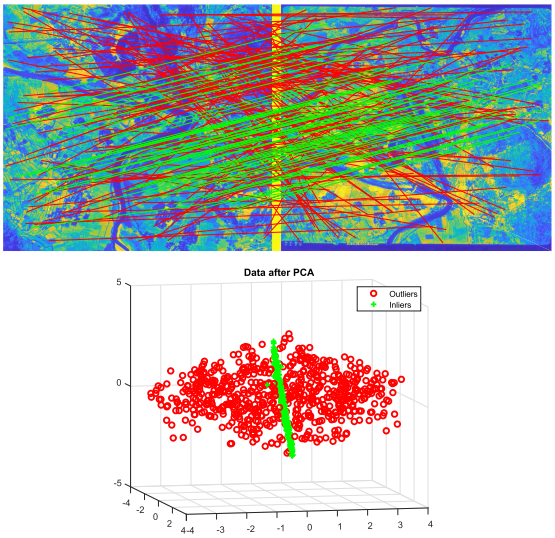

Fig. 6. Example of ground truth for affine hyperplane fitting. We show the input correspondences (Top) and project the data onto the first-3 principal components (Bottom), where green = inliers, red = outliers.

same first neighbor (i.e., x_4), while other feature points are merged due to the similar reasons.

It is worth pointing out that we cannot ignore the negative effect of outliers on the clusters. This is because the cluster is generated based on the first neighbor of each feature point, that is, the elements of a cluster will be ineffective after linking the first neighbor of an outlier. However, although two feature points of a true correspondence may not have many neighbors that belong to the same cluster due to the effects of outliers, the correspondences typically have a high probability of being inliers if their feature points belong to the same cluster in both views. This is because these correspondences preserve local neighborhood structures of feature points.

Thus, our goal is to match clusters that share the most same elements in two views, respectively. Specifically, we count the number of correspondences whose feature points belong to the same clusters in both views; and then select correspondences from the best match that share the most feature points, as the consistent seeds are used to estimate the affine hyperplane for the outlier removal problem. For example, in Fig. 3, clusters A and B share three correspondences while clusters C and D share four correspondences. Thus, the match between C and D is the best one. Then, the four correspondences are considered as consistent seeds.

In this article, we argue that the first neighbor of feature points of each correspondence is a sufficient statistic to discover the consistent seeds for outlier removal. To further illustrate this idea, we test different settings of the nearest neighbors of feature points on 50 remote sensing image pairs and report the f-score and running time in Fig. 5. From the results, we can see that only using the first neighbor of feature points is able to obtain the best performance on both f-score and running time among all cases. The reason behind this is that increasing the nearest neighbors will increase the bad effects of outliers as well. Thus, we only use the first neighbor of feature points in our method.

Note that, the work of Sarfraz [42] also uses the first neighbor relation of data points for clustering as our method. However, they are significantly different: 1) Sarfraz [42] cannot deal with outliers while our method further exploits the consistency of local neighborhood structures to deal with outliers; 2) Sarfraz [42] belongs to the family of hierarchical agglomerative methods and requires to provide the number of clusters; while our method only discovers the chains based on the first neighbor relation of feature points and does not require to provide any parameter; 3) Sarfraz [42] aims to segment all data points, but our method aims to seek some consistent seeds for the following process of outlier removal and we have further verified the effectiveness of these consistent seeds (see Section III-C). Thus, our method is more general and effective for the outlier removal problem than [42].

C. Proposed Hyperplane Updating Strategy

We can obtain stable results of outlier removal for remote sensing images according to the affine hyperplane, which is estimated by the consistent seeds obtained by the strategy in Section III-B. However, the proposed consistent seed-seeking strategy only preserves the consensus of local neighborhood elements while ignores the consensus of global information, which will affect the effectiveness of the estimated affine hyperplane.

For example, in Fig. 6, which includes 1000 matches with 30% inliers, we show the consistent seeds obtained by the proposed consistent seed seeking strategy in Fig. 7(a). We can see that, although the consistent seeds are inliers, the estimated affine hyperplane [shown in Fig. 7(b)] can still be refined to be more close to the ground truth [shown in Fig. 6(b)] and the corresponding matching results [shown in Fig. 7(c) and (d)] are also not desirable.

To address this problem, we introduce a novel hyperplane updating strategy, which exploits the residual information of all given correspondences. The proposed hyperplane updating strategy is based on the assumption that a true hyperplane

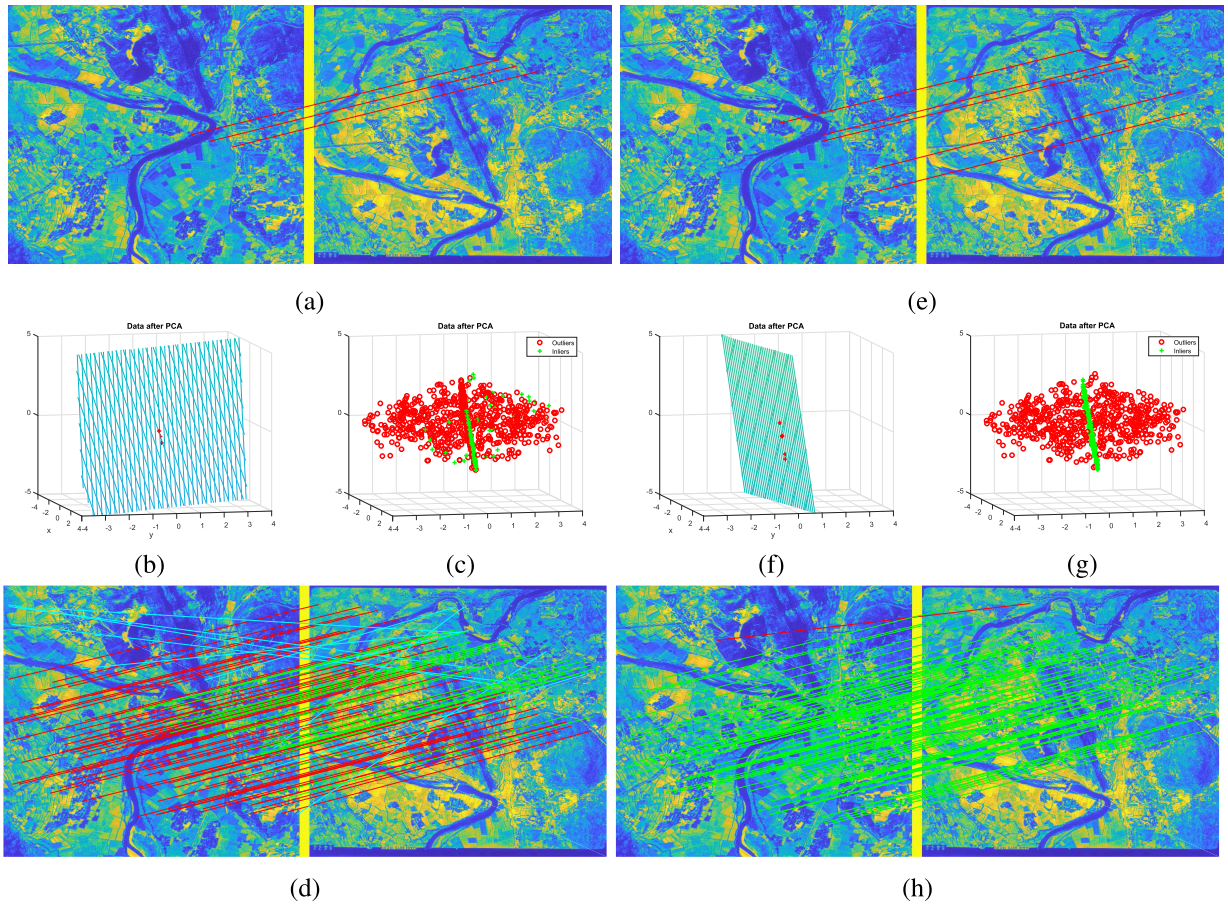

Fig. 7. Example showing the proposed hyperplane updating strategy. From top to bottom: the samples for fitting hyperplane; the fitting hyperplane and the corresponding inliers and outliers; the final feature matching results (green = true positive, red = false negative, cyan = false positive). (a)–(d) are the results obtained by the consistent seeds; (e)–(h) are the results obtained by the samples derived from the proposed hyperplane updating strategy.

has enough inliers that can be used to reestimate a similar hyperplane; otherwise, the hyperplane is estimated by the inliers of a false hyperplane is significantly different from the false hyperplane. This assumption is reasonable for our case in this article since we only use a single and global affine model to fully characterize the given correspondences for remote sense images. That is, if the inliers derived from the estimated by the hyperplane are the truth, then they can be used to estimate a stable hyperplane; otherwise, the estimated hyperplane will vary.

Based on the above assumption, we describe the details of our hyperplane updating strategy as follows: Given the consistent seeds, we first estimate the affine hyperplane; Then, we compute the residual values between the affine hyperplane and the input correspondences; After that, we sort the residual values in ascending order and sample $p+2$ correspondences around the m_kth correspondence in the order of residual values.[1] We repeat these steps until we obtain a converged solution.

Intuitively, if an affine hyperplane θ is the true model, then a new affine hyperplane, which is reestimated by using $p+2$ correspondences around the m_kth correspondence, will be close to θ; Otherwise, the new affine hyperplane that is reestimated by the correspondences of a false hyperplane θ', will be far from θ'. It is worth pointing out that, the reestimated affine hyperplane is not an arbitrary model since it is derived from the correspondences around the m_kth residual order, which means that these correspondences are likely to be around the intersection of the current affine hyperplane with one of those clusters. Therefore, the strategy repeats these steps and as soon as two sequential feature matching results are similar, it is deemed to have converged to a solution.

To handle various types of remote sensing images, we define a locality preserving structure-based cost function to measure the quality of feature matching results. A good feature matching result consists of as more true correspondences as possible, and a true correspondence will well preserve a locality structure. Thus, we search the K nearest neighbors of two feature points of a correspondence $s_i' = (x_i, y_i)$ under the Euclidean distance to measure the locality structure

$$L(s_i') = \frac{1}{2K}\left(\sum_{j|x_j\in\mathcal{N}_{x_i}} \left(d_q(x_i,x_j) - d_q(y_i,y_j)\right)^2 + \sum_{j|y_j\in\mathcal{N}_{y_i}} \left(d_q(x_i,x_j) - d_q(y_i,y_j)\right)^2\right) \quad (5)$$

[1]Here, $p=3$ is the minimal number of correspondences to determine a hyperplane. m_k is the minimum number of inliers, which means that a true affine hyperplane has at least m_k inliers.

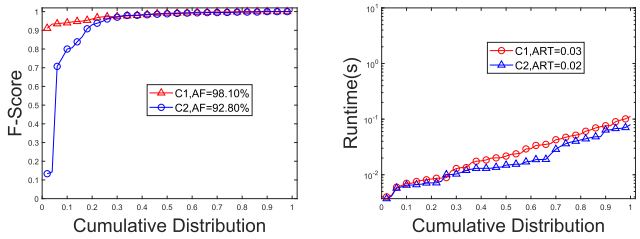

Fig. 8. F-score (Left) and running time (Right) with respect to the cumulative distribution obtained by different settings of the cost function.

where $\mathcal{N}_{x_i}$ and $\mathcal{N}_{y_i}$ are the neighborhoods of the points x_i and y_i, respectively. We normalize the contribution of each element in the neighborhood by $(1/2K)$. To handle scale changes that often happen remote sensing images, we quantize the distance between points into two levels as

$$d_q(x_i, x_j) = \begin{cases} 1, & \text{if } x_j \in \mathcal{N}_{x_i} \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

And $d_q(y_i, y_j)$ has the similar definition. In (5), if s_i' is an inlier, we will obtain a smaller value since s_i' will share more same neighbors in two views, i.e., $d_q(x_i, x_j)$ will be similar to $d_q(y_i, y_j)$; Otherwise, we will obtain a larger value. Thus, we can use (5) to measure the quality of a correspondence.

Then, we also consider the estimated inliers for the cost function, i.e., $n - n'$, where n and n' are the number of input matches and estimated inliers, respectively. This means that the more estimated inliers the feature matching result is better.

Thus, given the feature matches $S' = \{s_i' = (x_i, y_i)\}_{i=1}^{n'}$ derived from the estimated affine hyperplane, we define the cost function $C1(S')$ as follows:

$$C1(S') = \log_{10}\left(\sum_{i=1}^{n'} L\left(s_i'\right)\right) + \log_{10}(n - n') \tag{7}$$

where we normalize the contribution of the quality and the number of the estimated inliers by log function. We can see that, by (7), if the obtained feature matches include more number of true inliers, we will obtain a smaller value of the cost function; Otherwise, we will obtain a larger value.

To illustrate the effectiveness of the locality preserving structure-based cost function, we introduce a density estimate technique-based cost function [43]

$$C2(S') = n \Big/ \sum_{j=1}^{n} \frac{\Psi(r_i(\theta')/b(\theta'))}{\tilde{s}(\theta')b(\theta')} \tag{8}$$

where $\Psi(\cdot)$ is a kernel function (such as the Epanechnikov kernel); θ' is the affine hyperplane generated by S'; $\tilde{s}(\theta')$ and $b(\theta')$ are the inlier scale and bandwidth of θ'; $r_i(\theta')$ is the residual derived from θ' and the ith correspondence; $b(\theta')$ is the bandwidth of the ith model hypothesis.

As discussed in [43], an affine hyperplane with more inliers, and with smaller residuals, is assigned a lower cost function according to (8). Then, we test our method with the two cost functions (i.e., (7) and (8)), on 50 remote sensing image pairs, and report the f-score and running time in Fig. 8. We can see that, although our method with (7) is slightly slower than the version with (8), the former improves the latter about 5.30% of the average f-score. Thus, our proposed cost function is very effective to our method.

Algorithm 1 FNRG Method

Input: putative set $S = \{(x_i, y_i)\}_{i=1}^{n}$, parameter m_k, K, $MaxIter$.

1: Search for the first neighbor of each feature point.
2: Construct two adjacency link matrixes by (4).
3: Generate clusters based on the adjacency link matrixes.
4: Seek consistent seeds by matching the clusters.
5: Estimate an affine hyperplane θ based on the seeds.
6: Initialize $t \leftarrow 1$; $\hat{\theta}^t \leftarrow \theta$;
7: **repeat**
8: Construct the rank list $\{rank_i^t\}_{i=1}^{n} \leftarrow Sort\left(\{r_i(\theta^t)\}_{i=1}^{n}\right)$;
9: Estimate an inlier set S^t based on an inlier scale.
10: Assign a cost $C(S^t)$ to the estimated inlier set by (7);
11: if $C1(S^t) = C1(S^{t-1})$ then
12: break;
13: end if
14: $\hat{\theta}^{t+1} \leftarrow Fit([x_{rank_j}]_{j=m_k-4}^{m_k})$ //Fit a hyperplane.
15: t++;
16: **until** $t \geq MaxIter$
17: $S^* \leftarrow argmin\{C1(S^j)\}_{j=1,2,\cdots}$

Output: The estimated inlier set S^*

To illustrate the process of our proposed hyperplane updating strategy, we show an example in Fig. 7, from which we can see that, after only one iteration, the samples [shown in Fig. 7(e)] are more widely distributed than the seeds [shown in Fig. 7(a)]. Accordingly, the estimated affine hyperplane [shown in Fig. 7(f)] is close to the ground truth and the corresponding matching results [shown in Fig. 7(g) and (h)] are also much better. The precision, recall, and f-score are also improved from 75.00%, 71.55% and 73.24% to 99.80%, 100.00% and 99.90%, respectively, while the running time is only used from 0.0292 to 0.0329 s.

Thus, we can significantly improve the effectiveness of the affine hyperplane by the proposed hyperplane updating strategy with a negligible time cost. Then, we summarize the proposed the FNRG feature matching method (called FNRG) for remote sensing image registration in Algorithm 1.

D. Computational Complexity

Algorithm 1 consists of two parts, i.e., the consistent seed seeking strategy (Lines 1 to 4) and the hyperplane updating strategy (Lines 5 to 17). For the consistent seed seeking Strategy, to search the first nearest neighbor of each feature point, the time complexity is close to $O(n\log(n))$ by using K-D tree [44]. Seeking consistent seeds in Lines 2 to 4 only involves simple operation, and its time complexity is much less than $O(n\log(n))$.

For the hyperplane updating strategy, to compute the cost function (Line 10) by (7), it also requires to search for the K nearest neighbors of each feature point for the estimated inliers, and its time complexity is close to $O((K+n')\log(n'))$ by using K-D tree, where $n' < n$ is the number of the

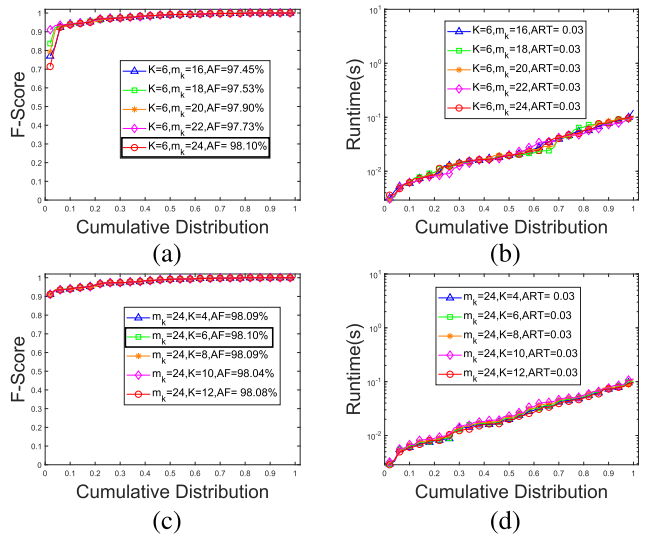

Fig. 9. F-score and running time with respect to the cumulative distribution obtained by different settings of parameters on 50 remote sensing image pairs. (a) and (c) f-score and run time with the same $K = 6$ and different m_k, respectively. (b) and (d) f-score and run time with the same $m_k = 24$ and different K, respectively. The best average f-score is denoted with box in the legend.

estimated inliers. To obtain the rank list $\{\text{rank}_i^t\}_{i=1}^n$ based on the residual values (Line 8), the time complexity is close to $O(n)$. Other steps (Lines 9, 11 to 15) only cost less $O(n)$ complexity. Generally, $(K + n')log(n') < n\log(n)$ and our method can converge in just one or two iterations. Therefore, the total time complexity of our FNRG in one iteration is about $O(n\log(n))$. The space complexity of Algorithm 1 is $O(Kn')$ due to the memory requirement for storing the neighborhoods of the estimated inliers. Thus, Algorithm 1 has linear time complexity and space complexity, which is significant to address large-scale feature matching problems.

E. Implementation Details

The proposed consistent seed seeking strategy is not required to provide any parameter in advance, and the hyperplane updating strategy includes three parameters, i.e., m_k, K, and $MaxIter$. m_k is the minimal inlier to support a hyperplane, and it determines the sample used to reestimate a new hyperplane. K is the number of nearest neighborhoods considered for computing the cost function in (7). $MaxIter$ determines the max iteration of the proposed hyperplane updating strategy. To seek the optimal value of the parameters, i.e., m_k and K (we fix $MaxIter = 10$ since the algorithm often converges to a solution with one or two iterations and it is not necessary to test the value of $MaxIter$), we test different settings on 50 remote sensing image pairs, and report the f-score and running time in Fig. 9.

From the results, we can see that, the f-score does not change significantly as different m_k and K, and the running time is almost constant. Thus, we can set m_k as the value from $\{16, 18, 20, 22, 24\}$, and K as the value from $\{4, 6, 8, 10, 12\}$. In our evaluation, we set the default values of the parameters to $m_k = 24$ and $K = 6$, respectively, due to the slightly better performance.

It is worth pointing out that, the proposed hyperplane updating strategy also includes another parameter, i.e., the inlier scale, to distinguish inliers and outliers. We have not considered it as input parameters since it can be derived from some inlier scale estimators, such as Modified Selective Statistical Estimator (MSSE) [45], which is used to estimate the inlier scale in our evaluation due to its effectiveness.

IV. Experimental Results

In this section, we compare the proposed feature matching method (called FNRG) with six state-of-the-art methods, including, identifying correspondence function (ICF) [46], graph shift (GS) [9], LPM [11], RANSAC++ [23], MTOP [1], and RFMSCAN [24], for the feature matching task on different kinds of remote sensing data sets. We also run RANSAC [6] as a baseline. We fix the parameters of all competing methods according to the original articles. All experiments are run on MS Windows 10 with Intel Core i7-8565 CPU 1.8 GHz and 16 GB RAM.

A. Data Sets and Settings

We perform all competing methods on six remote sensing image data sets as follows:

1) UAV: This data set contains 41 image pairs, which are captured by an unmanned gyroplane, i.e., UAV, over a piece of farmland. These image pairs can be used to handle the automatic crop monitoring problem by feature matching, and they involve projective distortions.

2) SAR: This data set contains 17 image pairs, which are captured by synthetic-aperture radars on a satellite. These image pairs can be used to handle the positioning and navigating problem by feature matching, and they involve affine distortions.

3) CIAP: This data set contains 54 color infrared aerial photographs, which are captured by eastern Illinois, IL, USA. These image pairs can be used to handle the image mosaic problem by feature matching, and they only involve the rigid transformation.

4) PAN: This data set contains 33 image pairs, which are captured by panchromatic aerial from a frame camera at different times. These image pairs can be used to handle the change detection problem, and they involve affine or projective distortions.

5) FE: This data set contains 18 image pairs, which are captured by different scenes from a fisheye camera. These image pairs can be used to handle the nonparametric image matching evaluation, and they involve viewpoint changes and severe nonrigid deformations.

6) MU: This data set contains 10 image pairs [47], which are multiple modalities, including thermal, visible, LiDAR intensity, and depth images from a gaming sensor. These image pairs can be used to handle the multimodal problem, and they involve many same features in the images.

Before the experiments, we have generated the ground truth as a benchmark, and we also have manually check the

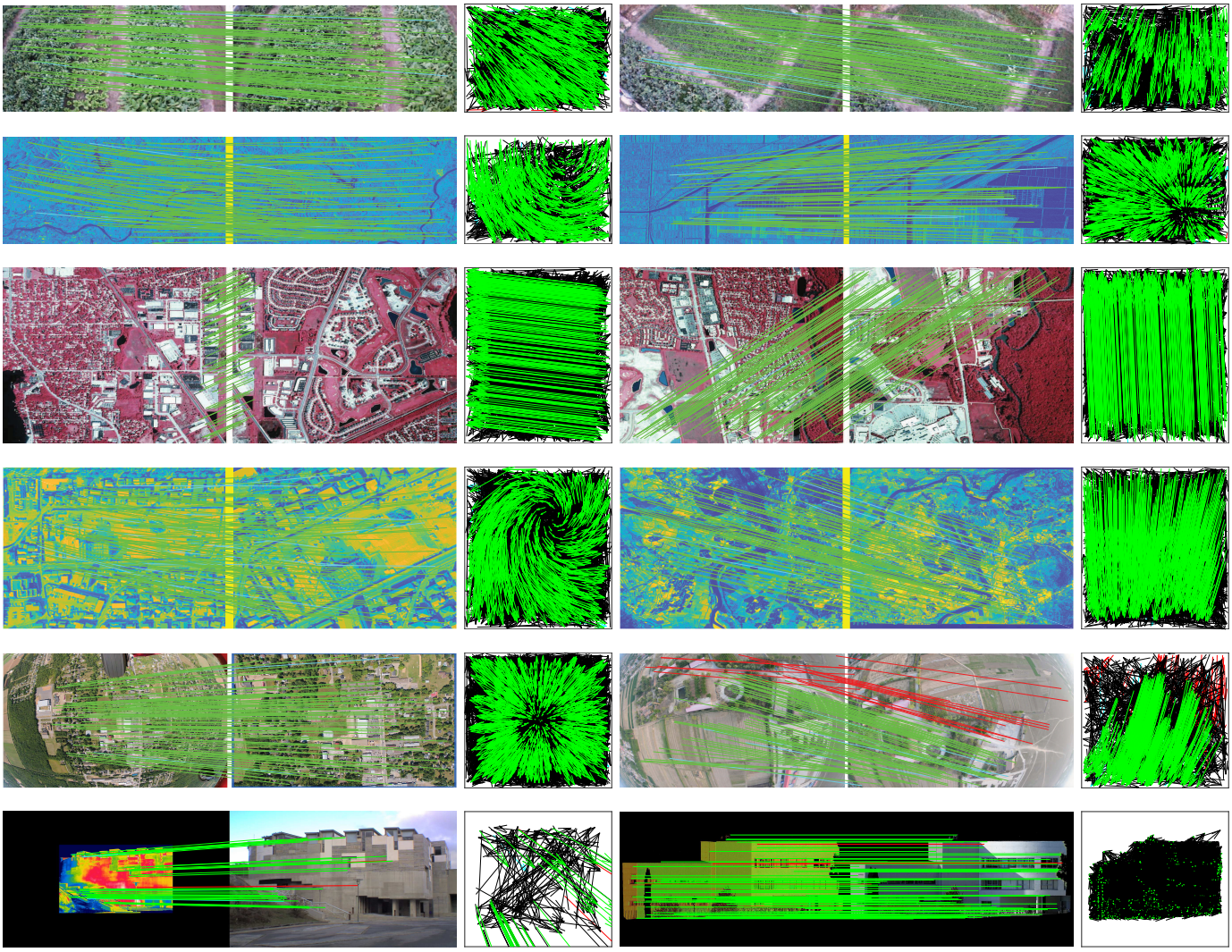

Fig. 10. Qualitative illustration of feature matching performance of our FNRG on 12 representative image pairs. From left to right and top to bottom: UAV1, UAV2, SAR1, SAR2, CIAP1, CIAP2, PAN1, PAN2, FE1, FE2, MU1, and MU2 with 31.97%, 26.56%, 50.97%, 43.44%, 8.78%, 11.82%, 26.56%, 26.99%, 31.97%, 50.30%, 22.35%, and 20.26% inlier ratios, respectively. For visibility, at most 100 randomly selected matches without true negatives are presented. We also show the motion field, where the head and tail of each arrow correspond to the positions of feature points (green = true positive, black = true negative, red = false negative, cyan = false positive).

correspondence set of all image pairs to ensure objectivity. To measure the matching performance, we have computed the value of precision, recall and f-score according the matching results obtained by all competing methods. Here, precision, recall, and f-score are defined as

$$\text{precision} = \frac{\text{tp}}{\text{tp} + \text{fp}} \tag{9}$$

$$\text{recall} = \frac{\text{tp}}{\text{tp} + \text{fn}} \tag{10}$$

$$f - \text{score} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \tag{11}$$

where tp is the number of true positive correspondences; fp is the number of false positive correspondences; fn is the number of false negative correspondences.

B. Qualitative Results

We first provide the matching results obtained by our FNRG on some representative image pairs from the six data sets in Fig. 10. We show two examples for each data set, and for each example, we show the intuitive matching results (we only show 100 randomly selected matches of true positive, false negative and false positive) and the motion field of the matches (the head and tail of each arrow correspond to the positions of feature points).

The representative image pairs involve severe noise, small overlaps, projective distortions, viewpoint changes, nonrigid deformations and multimodal data, thus, it is a very challenging task to mine consistent matches. For the input image pairs, we use SIFT detector to extract feature points and the corresponding descriptor to generate matches. For the twelve image pairs, the number of inliers (and the inlier ratio) are 392(31.97%), 239(26.56%), 421(50.97%), 861(43.44%), 219(8.78%), 313(11.82%), 612(26.56%), 506(26.99%), 985(31.97%), 503(50.30%), 38(22.35%) and 565(20.26%), respectively. We use our FNRG to remove mismatches for the 12 image pairs, and compute the precision, recall, and f-score as follows: (98.04%, 94.66%, 96.32%), (96.56%, 88.52%,

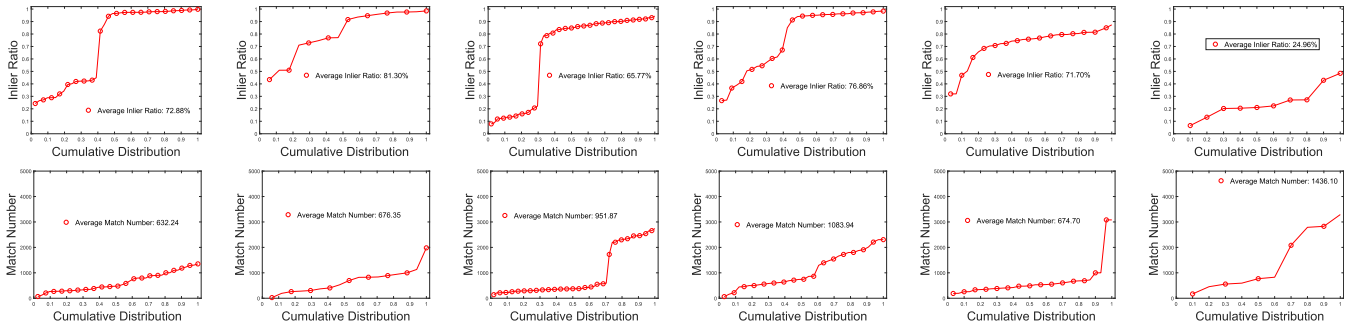

Fig. 11. Initial inlier ratio and match number with respect to the cumulative distribution. From left to right: UAV, SAR, CIAP, PAN, FE, and MU.

92.36%), (98.67%, 97.45%, 98.06%), (96.77%, 93.36%, 95.04%), (99.96%, 99.55%, 99.75%), (99.89%, 99.05%, 99.47%), (97.92%, 92.73%, 95.25%), (96.16%, 87.54%, 91.65%), (99.45%, 98.30%, 98.87%),(86.80%, 91.69%, 89.18%), (96.47%, 92.11%, 94.24%) and (98.28%,100.00%, 99.39%), respectively. We can see that our FNRG is able to achieve good feature matching results for the 12 challenging image pairs. That is, FNRG can successfully mine most of the true matches from the putative correspondences, only fail to wrongly distinguish a few matches. This can show that FNRG is able to handle different types of remote sensing images with different transformations.

C. Quantitative Results

In this section, we implement all competing methods, including RANSAC, ICF, GS, LPM, RANSAC++, MTOP, RFMSCAN, and FNRG, on six data sets. We show the inlier ratio and match number of the six data sets with respect to the cumulative distribution in Fig. 11. Then, we report the precision, recall, f-score, and running time obtained by all competing methods with respect to the cumulative distribution in Fig. 12, and we also summarize the average quantitative comparisons of all competing methods in Table I.

We can see that, for UAV, all competing methods are able to achieve good performance on the three measures. RANSAC, MTOP, and FNRG achieve similar values of f-score, but RANSAC is not very stable due to its random nature (we show the best values by repeating RANSAC 10 times). Although GS and LPM achieve high recall values, their precision values are much lower than other competing methods. ICF, RANSAC++, and RFMSCAN obtain similar values of the three measures. For SAR, FNRG is able to achieve the best values of precision and recall among all seven competing methods, and it also achieves high recall values. For CIAP, ICF, RFMSCAN, and FNRG almost correctly distinguish all true matches from the putative matches while only wrongly distinguishing a few matches, and they achieve the best results on f-score among all competing methods. This data set only involves the rigid transformation, thus, it is easy to handle the feature matching for all methods. For PAN, which involves affine or projective distortions, FNRG is able to good values of precision, recall, and f-score. For FE, which is the most challenging data set, all the seven competing methods cannot perform better results than that for other five data sets.

TABLE I

AVERAGE QUANTITATIVE COMPARISONS OF ALL COMPETING METHODS ON SIX DATA SETS. AP-AVERAGE PRECISION; AR-AVERAGE RECALL; AF = AVERAGE F-SCORE; ART = AVERAGE RUNNING TIME. "–" DENOTES THE METHOD CANNOT WORK FOR THE DATA SET. THE BEST RESULT OF F-SCORE IS BOLDFACED

Method		RANSAC	ICF	GS	LPM	RANSAC++	MTOP	RFMSCAN	FNRG
UAV	AP (%)	97.86	95.86	85.74	91.09	97.79	98.09	95.86	97.03
	AR (%)	96.24	95.61	99.50	99.58	95.00	96.79	95.61	97.16
	AF (%)	97.01	95.69	91.73	94.81	96.33	**97.42**	95.69	97.04
	ART (s)	0.06	0.02	9.19	0.01	0.15	0.07	0.02	0.02
SAR	AP (%)	97.50	94.29	89.64	93.58	96.57	98.09	94.29	98.21
	AR (%)	96.80	95.58	98.95	97.95	95.33	92.92	95.58	97.47
	AF (%)	97.14	94.89	93.88	95.60	95.93	93.47	94.89	**97.83**
	ART (s)	0.03	0.03	26.08	0.01	0.17	0.07	0.03	0.02
CIAP	AP (%)	99.04	99.63	94.93	97.27	97.25	95.12	99.63	99.75
	AR (%)	96.37	99.50	99.95	99.59	90.16	92.53	99.50	99.56
	AF (%)	97.58	99.57	97.32	98.39	93.12	93.76	99.57	**99.66**
	ART (s)	0.18	0.08	5.88	0.01	0.18	0.10	0.08	0.03
PAN	AP (%)	98.44	97.87	88.87	90.68	98.00	97.90	97.87	98.41
	AR (%)	97.36	97.47	99.70	99.69	96.50	97.36	97.47	98.09
	AF (%)	97.89	97.65	93.79	94.59	97.23	97.63	97.65	**98.24**
	ART (s)	0.05	0.07	32.00	0.01	0.20	0.12	0.07	0.04
FE	AP (%)	88.71	82.39	73.91	80.25	88.74	87.62	82.39	88.87
	AR (%)	88.89	90.58	98.45	97.71	86.69	86.13	90.58	90.69
	AF (%)	88.77	86.14	83.92	87.89	87.70	86.87	86.14	**89.72**
	ART (s)	0.05	0.04	6.42	0.01	0.14	0.07	0.04	0.03
MU	AP (%)	90.72	-	71.92	81.14	87.74	91.53	87.86	91.16
	AR (%)	69.81	-	24.96	79.11	63.77	76.80	67.25	96.48
	AF (%)	77.92	-	34.29	74.67	72.54	82.51	75.21	**93.40**
	ART (s)	0.25	-	2.74	0.01	0.22	0.15	0.13	0.05

This is because this data set suffers from viewpoint changes and severe nonrigid deformations. However, our FNRG still obtains the best f-score value among all methods. For MU, which involves multiple modalities, FNRG shows significant superiority over other competing methods, especially for recall and f-score. This is because FNRG achieves the matching results by exploiting the relationship among matches, and it does not depend on the effectiveness of the complex features caused by multimodal data.

Note that, RANSAC can achieve better performance than several compared methods on six data sets, especially for the one with a high inlier ratio (i.e., SAR). However, the results obtained by RANSAC are not very stable due to its randomness. It is also worth pointing out that, given an appropriate inlier scale threshold value, RANSAC can obtain good performance when the data has not a large proportion of outliers; But, for the data with a low inlier ratio (i.e., MU), RANSAC achieves bad performance as several compared methods.

It is worth pointing out that, FNRG has not always shown the best results of precision and recall among all

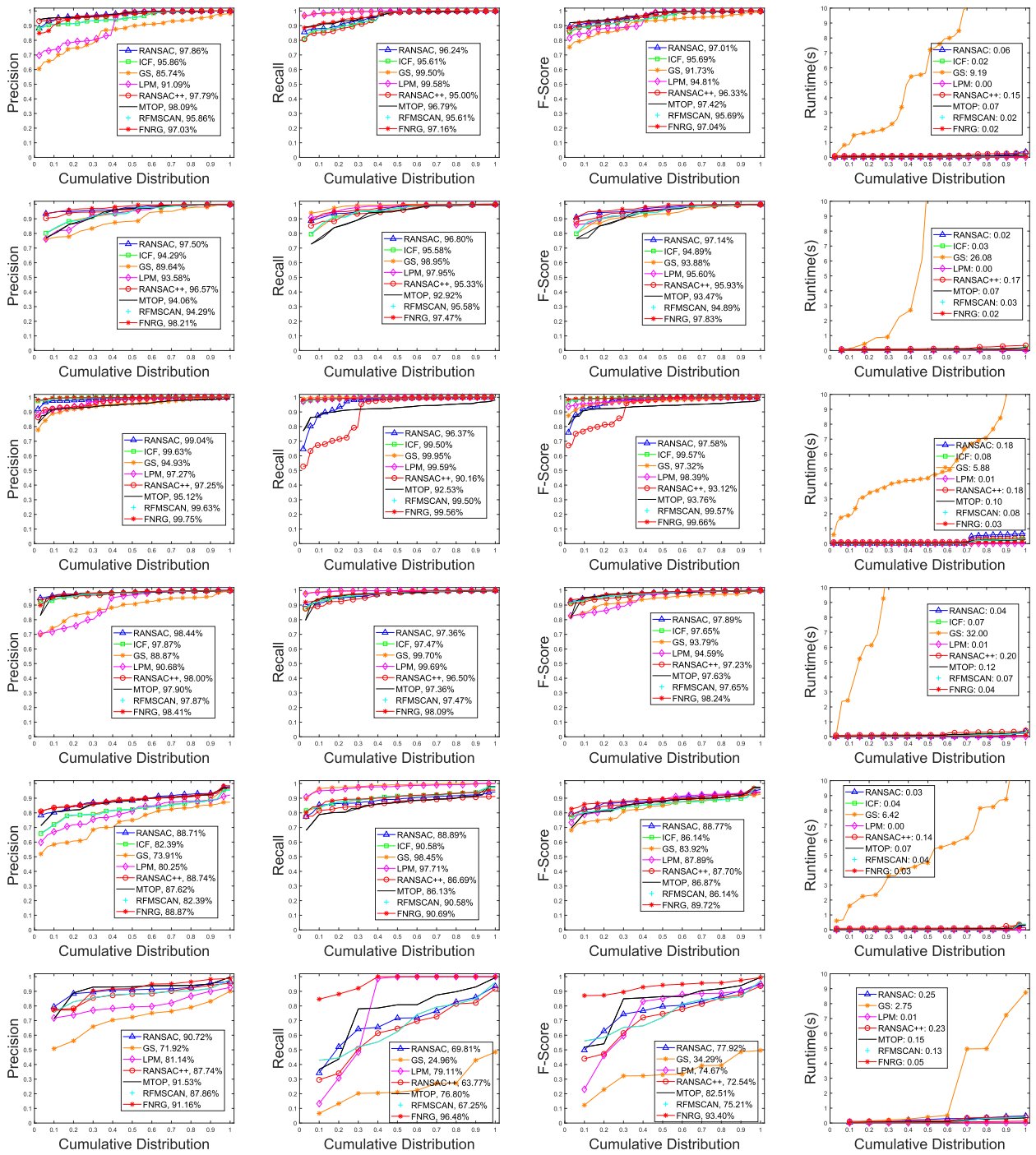

Fig. 12. Quantitative comparisons of RANSAC, ICF, GS, LPM, RANSAC++, MTOP, RFMSCAN, and FNRG on six data sets. From top to bottom: UAV, SAR, CIAP, PAN, FE, and MU. From left to right: Initial inlier ratio, precision, recall, f-score and running time with respect to the cumulative distribution.

competing methods on all data sets. Even so, we cannot make it clear which is more important between precision and recall. Thus, we also use f-score, which is a comprehensive evaluation between precision and recall, to measure the matching performance. From Table I, FNRG is able to achieve the best f-scores in five out of six data sets and the second-best f-score in the remaining data set. For the running time, LPM is fastest than the other six competing methods, while our FNRG also achieves similar speeds as LPM. That is, our FNRG is able to accomplish the feature matching from thousands of matches in only a few milliseconds.

D. Robustness Test

In this section, we first test the robustness of our FNRG in the case of Gaussian noise. To this end, we test FNRG on 50 remote sensing image pairs, and add different variances of Gaussian noise on the image coordinates of correspondences, and report the f-score and running time in Fig. 13.

From the results, we can see that the f-score does not change significantly on the 50 remote sensing image pairs with variances of Gaussian noise from 0.00 to 0.10. The biggest gap in the average f-scores is 1.27%. The running time is very constant. This shows that our FNRG is very robust to Gaussian noise.

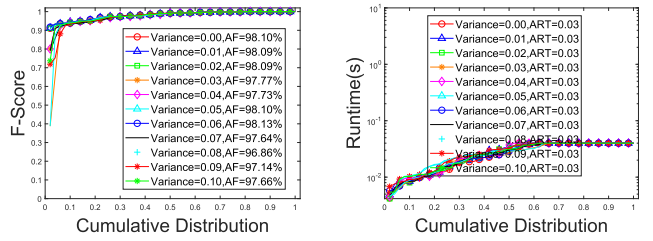

Fig. 13. F-score (Left) and running time (Right) with respect to the cumulative distribution obtained by different variances of Gaussian noise.

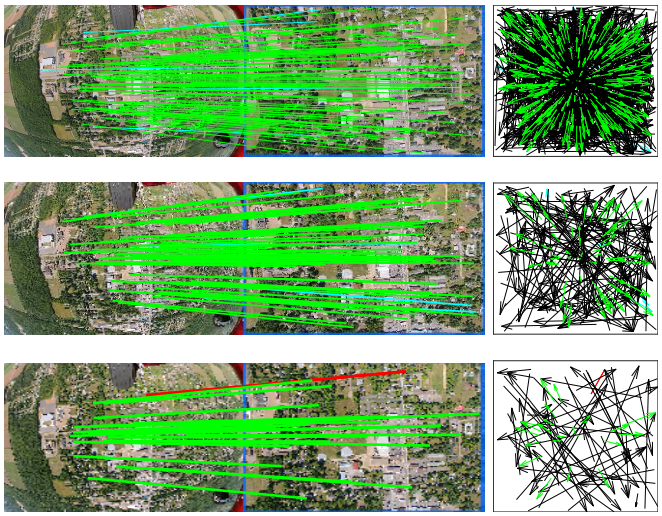

Fig. 14. Qualitative illustration of feature matching performance of our FNRG on an image pair with different GSDs.

We also test our FNRG on an image pair (i.e., FE1) with different ground sampling distances (GSD) since remote sensing images from different imaging devices usually have different GSD. Here, we downsample the image pair with different coefficients to obtain different image resolutions for simulating different GSD. We show the intuitive matching results and the motion field of the matches on Fig. 14, and we also compute the corresponding precision, recall, and f-score as follows: (98.52%, 95.33%, 96.90%), (98.13%, 94.20%, 96.13%), (97.70%, 100.00%, 98.84%). We can see that our method is able to achieve good performance on the image pair with different image resolutions.

V. Discussion and Conclusion

In this article, we propose a feature matching method (called FNRG) for remote sensing image registration. FNRG starts from a novel consistent seed seeking strategy, which exploits the first neighbor relation of feature points between two images to mine consistent matches, without any parameter or threshold. Then, we formulate the feature matching problem into an affine hyperplane fitting problem. That is, based on the seeds, FNRG fits an affine hyperplane for remote sensing image registration after imposing the motion consistency. After that, to further improve the matching performance of FNRG, we propose a novel hyperplane updating strategy to refine the fitting model. For the hyperplane updating strategy, we also introduce a locality preserving structure-based cost function to promote its effectiveness for feature matching. The qualitative and quantitative results of feature matching show that FNRG is able to handle the remote sensing image data sets with a large scale change, rotation, or severe nonrigid deformation, and effectively achieve consistent correspondences.

Compared to several state-of-the-art methods, FNRG also shows significant superiority on the matching performance for remote sensing image data sets with different types of image transformations. More importantly, FNRG is able to accomplish the mine correspondences from thousands of putative matches within a few milliseconds.

It is worth pointing out that, although FNRG is based on the assumption of a global affine model, it can handle different types of deformations (i.e., rigid and no-rigid deformations) and multimodal images due to the effectiveness of the affine hyperplane. However, when the input data includes multiply motion consistency, only a single affine hyperplane is not sufficient to obtain good feature matching performance. Thus, in future work, we shall propose a robust feature matching method to handle this situation.

References

[1] X. Jiang, J. Jiang, A. Fan, Z. Wang, and J. Ma, "Multiscale locality and rank preservation for robust feature matching of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6462–6472, Sep. 2019.

[2] 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.

[3] W. Li, C. Wang, C. Lin, G. Xiao, C. Wen, and J. Li, "Inlier extraction for point cloud registration via supervoxel guidance and game theory optimization," *ISPRS J. Photogramm. Remote Sens.*, vol. 163, pp. 284–299, May 2020.

[4] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, "Image matching from handcrafted to deep features: A survey," *Int. J. Comput. Vis.*, pp. 1–57, Aug. 2020, doi: 10.1007/s11263-020-01359-2.

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

[6] M. A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[7] G. Xiao, H. Wang, Y. Yan, and D. Suter, "Superpixel-guided two-view deterministic geometric model fitting," *Int. J. Comput. Vis.*, vol. 127, no. 4, pp. 323–329, 2019.

[8] H. Wang, G. Xiao, Y. Yan, and D. Suter, "Searching for representative modes on hypergraphs for robust geometric model fitting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 3, pp. 687–711, Feb. 2018.

[9] H. Liu and S. Yan, "Common visual pattern discovery via spatially coherent correspondences," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 1609–1616.

[10] W.-Y. Lin *et al.*, "CODE: Coherence based decision boundaries for feature correspondence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 1, pp. 34–47, Jan. 2018.

[11] 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, May 2019.

[12] C. Zhao, Z. Cao, C. Li, X. Li, and J. Yang, "NM-net: Mining reliable neighbors for robust feature correspondences," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 215–224.

[13] 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.

[14] Z. Shao, M. Chen, and C. Liu, "Feature matching for illumination variation images," *J. Elect. Imag.*, vol. 24, no. 3, pp. 1–11, 2015.

[15] M. Chen, Z. Shao, D. Li, and J. Liu, "Invariant matching method for different viewpoint angle images," *Appl. Opt.*, vol. 52, no. 1, pp. 96–104, 2013.

[16] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016.
[17] X. Zheng, Y. Yuan, and X. Lu, "A deep scene representation for aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4799–4809, Jul. 2019.
[18] H. Sun, S. Li, X. Zheng, and X. Lu, "Remote sensing scene classification by gated bidirectional network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 82–96, Jan. 2020.
[19] F. Luo, L. Zhang, X. Zhou, T. Guo, Y. Cheng, and T. Yin, "Sparse-adaptive hypergraph discriminant analysis for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 6, pp. 1082–1086, Jun. 2020.
[20] F. Luo, L. Zhang, B. Du, and L. Zhang, "Dimensionality reduction with enhanced hybrid-graph discriminant learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5336–5353, Aug. 2020.
[21] 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.
[22] J. Zhu, S. C. H. Hoi, and M. R. Lyu, "Nonrigid shape recovery by Gaussian process regression," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1–9.
[23] Q. Tran, T. Chin, G. Carneiro, M. S. Brown, and D. Suter, "In defence of ransac for outlier rejection in deformable registration," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 274–287.
[24] X. Jiang, J. Ma, J. Jiang, and X. Guo, "Robust feature matching using spatial clustering with heavy outliers," *IEEE Trans. Image Process.*, vol. 29, pp. 736–746, 2020.
[25] D. Liu, A. Parra, and T.-J. Chin, "Globally optimal contrast maximisation for event-based motion estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 1–9.
[26] J. Yang, H. Li, D. Campbell, and Y. Jia, "Go-ICP: A globally optimal solution to 3D ICP point-set registration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 11, pp. 2241–2254, Nov. 2016.
[27] H. Li, J. Zhao, J.-C. Bazin, W. Chen, Z. Liu, and Y. Liu, "Quasi-globally optimal and efficient vanishing point estimation in manhattan world," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2019, pp. 1–9.
[28] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1185–1198, Apr. 2012.
[29] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, Jun. 2008.
[30] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2564–2571.
[31] K. M. Yi, E. Trulls, Y. Ono, V. Lepetit, M. Salzmann, and P. Fua, "Learning to find good correspondences," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2666–2674.
[32] J. Zhang *et al.*, "Learning two-view correspondences and geometry using order-aware network," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2019, pp. 5844–5853.
[33] L. Magri and A. Fusiello, "T-linkage: A continuous relaxation of J-linkage for multi-model fitting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3954–3961.
[34] J. Bian, W.-Y. Lin, Y. Matsushita, S.-K. Yeung, T.-D. Nguyen, and M.-M. Cheng, "GMS: Grid-based motion statistics for fast, ultra-robust feature correspondence," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2828–2837.
[35] C. R. Qi, O. Litany, K. He, and L. Guibas, "Deep Hough voting for 3D object detection in point clouds," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2019, pp. 9277–9286.
[36] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "CCNet: Criss-cross attention for semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2019, pp. 603–610.
[37] S. Bai, P. Tang, P. H. S. Torr, and L. J. Latecki, "Re-ranking via metric fusion for object retrieval and person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 740–749.
[38] W. Hu, Y. Huang, F. Zhang, and R. Li, "Noise-tolerant paradigm for training face recognition CNNs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, p. 11.
[39] J. M. Murphy, J. Le Moigne, and D. J. Harding, "Automatic image registration of multimodal remotely sensed data with global shearlet features," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1685–1704, Mar. 2016.
[40] M. M. Alam, T. Howlader, and S. M. M. Rahman, "Entropy-based image registration method using the curvelet transform," *Signal, Image Video Process.*, vol. 8, no. 3, pp. 491–505, Mar. 2014.
[41] I. Zavorin and J. Le Moigne, "Use of multiresolution wavelet feature pyramids for automatic registration of multisensor imagery," *IEEE Trans. Image Process.*, vol. 14, no. 6, pp. 770–782, Jun. 2005.
[42] S. Sarfraz, V. Sharma, and R. Stiefelhagen, "Efficient parameter-free clustering using first neighbor relations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 8934–8943.
[43] H. Wang, T.-J. Chin, and D. Suter, "Simultaneously fitting and segmenting multiple-structure data with outliers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 6, pp. 1177–1192, Jun. 2012.
[44] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Commun. ACM*, vol. 18, no. 9, pp. 509–517, Sep. 1975.
[45] A. Bab-Hadiashar and D. Suter, "Robust segmentation of visual data using ranked unbiased scale estimate," *Robotica*, vol. 17, no. 6, pp. 649–660, Nov. 1999.
[46] X. Li and Z. Hu, "Rejecting mismatches by correspondence function," *Int. J. Comput. Vis.*, vol. 89, no. 1, pp. 1–17, Aug. 2010.
[47] M. Gesto-Diaz, F. Tombari, D. Gonzalez-Aguilera, L. Lopez-Fernandez, and P. Rodriguez-Gonzalvez, "Feature matching evaluation for multimodal correspondence," *ISPRS J. Photogramm. Remote Sens.*, vol. 129, pp. 179–188, Jul. 2017.

Guobao Xiao (Member, IEEE) received the B.S. degree in information and computing science from Fujian Normal University, Fuzhou, China, in 2013, and the Ph.D. degree in computer science and technology from Xiamen University, Xiamen, China, in 2016.

From 2016 to 2018, he was a Postdoctoral Fellow with the School of Aerospace Engineering, Xiamen University. He is a Professor with Minjiang University, Fuzhou. He has published over 30 articles in the international journals and conferences including IEEE TPAMI/TIP/TITS/TIE, IJCV, PR, ICCV, ECCV, AAAI, etc. His research interests include machine learning, computer vision, and pattern recognition.

Dr. Xiao has been received the Best Ph.D. Thesis in Fujian Province and the Best Ph.D. Thesis Award in China Society of Image and Graphics (a total of ten winners in China). He also served on the Program Committee (PC) of CVPR, ICCV, ECCV, AAAI, WACV, etc. He was the General Chair for the IEEE BDCLOUD 2019.

Huan Luo received the B.Sc. degree in software engineering from Nanchang University, Nanchang, China, in 2009, and the Ph.D. degree in computer science from Xiamen University, Xiamen, China, in 2017.

He is a Faculty Member with the College of Mathematics and Computer Science, Fuzhou University, Fuzhou, China. His research interests include point clouds processing, computer vision, and machine learning.

Kun Zeng (Member, IEEE) received the Ph.D. degree from the Department of Computer Science, Xiamen University, Xiamen, China, in 2015.

He was a Postdoctoral Fellow with the Department of Electronic Science, Xiamen University, from 2016 to 2019. He is a Lecturer with Minjiang University, Fuzhou, China. His research interests include image processing, machine learning, and medical image reconstruction.

Leyi Wei (Member, IEEE) received the Ph.D. degree in computer science from Xiamen University, Xiamen, China.

He is a Professor with the School of Software, Shandong University, Jinan, China. His research interests include machine learning and bioinformatics. He has 60+ peer-reviewed articles published on some top-tier journals such as the IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS, IEEE TRANSACTIONS ON NANOBIOSCIENCE, *Briefings in Bioinformatics*, and *Bioinformatics*, etc. He is now serving as a Section Editor of *Current Bioinformatics* and he also serves on the Program Committee (PC) of IEEE International Conference on Bioinformatics and Biomedicine (BIBM).

Jiayi Ma (Member, IEEE) received the B.S. degree in information and computing science and the Ph.D. degree in control science and engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2014, respectively.

From 2012 to 2013, he was an Exchange Student with the Department of Statistics, University of California at Los Angeles, Los Angeles, CA, USA. He is a Professor with the Electronic Information School, Wuhan University. He has authored or coauthored more than 150 refereed journal and conference papers, including the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (TPAMI)/TRANSACTIONS ON IMAGE PROCESSING (TIP), International Journal of Computer Vision IJCV, CVPR, ICCV, ECCV, etc. His research interests include computer vision, machine learning, and pattern recognition. He has been identified in the 2020 and 2019 Highly Cited Researcher lists from the Web of Science Group.

Dr. Ma is an Area Editor of *Information Fusion*, an Editorial Board Member of *Neurocomputing* and *Entropy*, and a Guest Editor of *Remote Sensing*.