Progressive Motion Coherence for Remote Sensing Image Matching

Yizhang Liu, Brian Nlong Zhao, Shengjie Zhao, *Senior Member, IEEE*, and Lin Zhang, *Senior Member, IEEE*

Abstract—**In this article, we present a feature-based remote sensing (RS) image matching method termed progressive motion coherence (PMC). We formulate the matching problem into a mathematical model and derive a closed-form solution. The objective function is only based on two novel coherence constraints, namely, efficient neighborhood element coherence and relative order-aware motion coherence, and hence, it is general enough and can be applied to RS image matching with different image types and degradations. The efficient neighborhood element coherence uses the Jaccard distance to measure the dissimilarity of two neighborhoods, which are lists composed of *k* nearest neighbors of feature points. To prevent overpenalization on the outliers, we combine it with an exponential function, which is simple yet efficient. The relative order-aware motion coherence is an alternative to motion smoothness, which is based on the observation that the relative order of neighboring matches for inliers in a small region can be well preserved, while for outliers, the relative order changes greatly. The above two coherences are robust to large rotation changes and low ratio inliers. Extensive experiments on five RS image datasets compared with seven state of the arts demonstrate that our PMC is more efficient and robust than the competitors.**

Index Terms—**Image matching, motion coherence, relative order aware, remote sensing (RS).**

I. Introduction

REMOTE sensing (RS) image matching is a fundamental prerequisite for many RS tasks, such as image stitching [1], [2], change detection [3], mapping sciences [4], [5], and image fusion [6], [7]. Image matching aims to align a pair of RS images captured from the same scene but at different times, from different perspectives, or by different sensors. The technologies developed for RS image matching can be roughly subdivided into two categories, namely, intensity- and feature-based methods [8], [9], [10]. The intensity-based methods utilize the initial pixel intensities in the overlapped area of two RS images with a particular similarity metric to find the matching information. These approaches solely consider the pixel intensities, which will be problematic when the image pairs suffer from illumination changes and repetitive structures. The above two challenges are often the cases in the RS image matching problem. Furthermore, the computational complexity is related to the number of pixels, which is not suitable for real-time RS image matching tasks. By contrast, feature-based methods focus on the extracted salient or abstract features of the images (e.g., scale invariant feature transform (SIFT) [11] and learned invariant feature transform (LIFT) [12]) and match them according to a specific similarity metric together with the geometric constraint as a postprocessing procedure. In addition, the discriminative feature points can be regarded as representative of the images, and hence, they are more general than pixel intensities. Since the extracted features are robust to noise, illumination changes, translation/rotation/scale changes, and the number of feature points is much less than the number of pixels, the major research trend in RS image matching has been the feature-based techniques.

The feature-based pipeline mainly includes four steps, namely, feature extraction, feature description, feature matching, and outlier removal. In this article, we focus on the outlier removal problem, i.e., seeking reliable matches from the initial putative match set and rejecting the outliers, where the putative match set is constructed by matching those feature points (from different images) with the most similar feature descriptors. Since the RS image pairs to be matched may be obtained by different sensors and they may also undergo various degradations, such as severe noise, ground relief variation, small overlap, repetitive structures, viewpoint/illumination/scale/rotation changes, and nonrigid distortion, the descriptors are not discriminative enough, and the initial putative match set will contain a lot of outliers. Thus, it is significant to develop a robust and efficient outlier removal method for RS image matching.

In recent years, many outlier removal methods have been proposed, and they have achieved promising results. Nevertheless, there are still some challenges to be solved. First, parametric model-based methods need to assume a specific transformation between two images in advance so as to estimate the parameters of the model iteratively. In fact, in real-world scenes, the transformation between the image pairs is usually unknown, and sometimes, it is not even possible to model with parameters. Thus, parametric model-based

Manuscript received 9 April 2022; revised 5 June 2022 and 2 August 2022; accepted 17 August 2022. Date of publication 8 September 2022; date of current version 11 October 2022. This work was supported in part by the National Key Research and Development Project under Grant 2019YFB2102300, in part by the National Natural Science Foundation of China under Grant 61936014, in part by the Shanghai Municipal Science and Technology Major Project under Grant 2021SHZDZX0100, and in part by the Fundamental Research Funds for the Central Universities. *(Corresponding author: Shengjie Zhao.)*

Yizhang Liu, Shengjie Zhao, and Lin Zhang are with the School of Software Engineering, Tongji University, Shanghai 201804, China (e-mail: lyz8023lyp@gmail.com; shengjiezhao@tongji.edu.cn; cslinzhang@tongji.edu.cn).

Brian Nlong Zhao is with the Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089 USA (e-mail: briannlongzhao@gmail.com).

Digital Object Identifier 10.1109/TGRS.2022.3205059

methods fail to perform well in the case of complex and nonrigid transformations. Second, iterative-based methods cannot guarantee the optimal solution when the initial putative match set contains a large proportion of outliers, and they are extremely time-consuming. Third, motion coherence-based methods are based on the slow-and-smooth assumption that correct correspondences are smooth in a small region. Actually, when the image pairs suffer from large rotation changes and the feature points are sparse in the local region, even the correct matches are hard to satisfy the smoothness constraint. In the aforementioned cases, the performance of motion coherence-based methods is degraded severely.

To address the challenges described above, we propose a method termed progressive motion coherence (PMC) for RS image matching. According to the motion coherence theory [13], if two feature points are close, they probably belong to the same object and thus tend to move together. In other words, for an inlier, there are many matches (e.g., inliers) in its local neighborhood. Conversely, for an outlier, there are fewer or no matches in its local neighborhood. Therefore, the number of neighboring matches with respect to a match can be used to measure the probability that a match is an inlier or outlier. Besides, although the assumption of smoothness constraint in the coherence-based method may not hold (e.g., in the case of large rotation changes), we find that the relative order of neighboring matches for inliers in a small region can be well preserved, while for outliers, the relative order changes greatly. Combining the above two coherences, we can efficiently distinguish outliers from inliers in linear time.

In summary, this work makes the following contributions. First, we formulate the outlier removal as a mathematical model and derive a closed form. The objective function is only based on two novel coherence constraints, and hence, it is general and can be applied to RS image matching with different image types and degradations. Second, we improve existing neighborhood element coherence, which is linear and sensitive to the ratio of inliers. The proposed efficient one is nonlinear, and its curve has "long tail," which can prevent overpenalization on the outliers. Third, to the best of our knowledge, we are the first to introduce the invariance of the relative order of neighboring matches for inliers in a small region to the motion coherence-based framework, which is more general and robust to outliers than existing ones. Experiments validate our method's superiority over the comparable state of the arts.

The rest of this article is organized as follows. Section II introduces the related works. Section III presents the details of our method. Section IV discusses the experimental results with respect to feature matching and image registration tasks. Section V provides the conclusion.

II. Related Work

To solve the aforementioned challenges, a number of outlier removal methods have been proposed, which can be roughly divided into traditional (nonlearning) methods and learning-based methods. Here, we briefly overview the related work.

The random sample consensus (RANSAC) [14] and its variants [15], [16], [17], [18], [19], [20], [21] are the most classic methods for the outlier removal problem. These methods attempt to find the outlier-free subset to estimate the parameters of the predefined models by iterative resampling. Generally, these methods can achieve satisfactory results when the following conditions hold: 1) the distribution of inliers can be explained by a parametric model and the outliers do not fit this model and 2) the ratio of inliers is relatively high. However, when the image pairs suffer from nonrigid transformations and the ratio of inliers is extremely low (e.g., $\leq 20\%$), these resampling-based methods fail to achieve good performance.

The nonparametric interpolation methods include identifying correspondence function (ICF) [22] and vector field consensus (VFC) [23] and its variants [24], [25], [26]. The ICF aims to find a correspondence function that optimally maps a point in one image to its corresponding point in the other. Those matches that are inconsistent with the correspondence function have been rejected. VFC solves the matching problem by interpolating a slow-and-smooth motion field, and its variants impose manifold regularization on the motion field. These methods can deal with nonrigid transformations and tolerate some outliers in the initial sets. Nevertheless, they suffer from cubic complexity, which is not suitable for real-time tasks.

The motion coherence-based methods, such as locality preserving matching (LPM) [27], grid-based motion statistics (GMS) [28], RFMSCAN [29], LAF [30], and MTOPK [31], are commonly based on the fact that inliers are consistent with each other in a small region, while outliers are randomly distributed. The LPM determines the correctness of each putative match by counting the number of neighboring matches. The GMS uses a predefined grid to accelerate the matching process. The RFMSCAN aims to adaptively cluster the putative matches into several motion-consistent clusters. The LAF regards outliers as noise and uses a linear adaptive filtering technique to filter out the outliers. The MTOPK measures the difference of ranking lists with respect to two matched feature points to reject outliers. These methods do not define any parametric model, and hence, they are general enough to be applied to fields with different image types and degradations.

The learning-based methods, such as LMR [32] and its variant [33], LFGC [34], NMNet [35], OANet [36], LMCNet [37], and MS2DGNet [38], make use of the powerful feature representative ability of the neural network to train a network for outlier removal. The LMR formulates the motion coherence of inliers into features to be trained. The LFGC is the first attempt that uses deep learning architecture to achieve outlier removal. The NMNet conquers the dilemma that LFGC cannot deal with multiple consistencies. The OANet, LMCNet, and MS2DGNet improve the design of the network and achieve good performance on several public datasets. Nevertheless, a massive amount of data and a lot of computational power are the basis of good performance. In the field of RS, due to the difficulty of data acquisition, it is hard to find a large amount of data to train the network.

Technically, the LPM is the published work most relevant to our PMC. It converts the motion smoothness into a measurement based on the number of neighboring matches. Furthermore, it presents a term, namely, the consensus of neighborhood topology, which is based on the relationship of the length and angle between a match and its neighboring matches. The LPM can provide satisfactory results when the motion smoothness holds. However, in the case of large rotation changes and low ratio inliers, its performance degrades greatly. Our PMC improves its neighborhood element coherence and provides an alternative to motion smoothness, which can adapt well to the above two challenges.

III. Method

A. Problem Formulation

Given a pair of images $\mathbf{I}_1$ and $\mathbf{I}_2$, we can use traditionally handcrafted feature descriptors (e.g., SIFT [11]) or newer deep learning-based ones (e.g., Superpoint [39]) to construct a set of putative matches based on a specific matching strategy. Generally, due to the ambiguity of the feature descriptors, the putative set inevitably contains many outliers. Hence, a robust and reliable outlier removal method is indispensable. In the following, we will introduce the proposed method in detail.

Suppose that we have obtained a putative set containing N matches $S = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N}$, where $\mathbf{x}_i$ and $\mathbf{y}_i$ denote the spatial coordinates of feature points. According to the motion coherence theory [13], inliers are consistent with each other in a small region, which is in accordance with the intuitionistic observation. Recently, several motion coherence-based methods have shown promising results in mismatch removal [27], [28], [32], [40], [41]. The number of neighboring matches with respect to a match has been used as a significant property for judging the matching correctness. Nevertheless, the number of neighboring matches with respect to different inliers may also vary a lot. Therefore, it is hard to find a suitable threshold to separate outliers from inliers. To solve this dilemma, we formulate the motion coherence from a novel perspective.

B. Efficient Neighborhood Element Coherence

Formally, we define $N_{\mathbf{x}_i}^k$ as the k nearest neighbors of feature point $\mathbf{x}_i$ under the Euclidean distance in image $\mathbf{I}_1$

$$N_{\mathbf{x}_i}^k = \left\{\mathbf{x}_1^i, \mathbf{x}_2^i, \ldots, \mathbf{x}_k^i\right\}, |N_{\mathbf{x}_i}^k| = k \tag{1}$$

where $|\cdot|$ denotes the cardinality of a set and $N_{\mathbf{x}_i}^k$ is an unordered set. Similarly, we can also construct neighborhood for feature point $\mathbf{y}_i$ in image $\mathbf{I}_2$, namely, $N_{\mathbf{y}_i}^k$. Ideally, if the match $(\mathbf{x}_i, \mathbf{y}_i)$ is an inlier, the neighborhoods $N_{\mathbf{x}_i}^k$ and $N_{\mathbf{y}_i}^k$ will be totally identical (i.e., its neighboring matches are all inliers). However, in real-world scenes, there will be noise or outliers in $N_{\mathbf{x}_i}^k$ and $N_{\mathbf{y}_i}^k$ due to image degradations and complex transformations. Therefore, the dissimilarity between $N_{\mathbf{x}_i}^k$ and $N_{\mathbf{y}_i}^k$ can be calculated by the Jaccard distance as follows:

$$d_J\left(N_{\mathbf{x}_i}^k, N_{\mathbf{y}_i}^k\right) = 1 - \frac{|N_{\mathbf{x}_i}^k \cap N_{\mathbf{y}_i}^k|}{|N_{\mathbf{x}_i}^k \cup N_{\mathbf{y}_i}^k|}. \tag{2}$$

Obviously, if $N_{\mathbf{x}_i}^k$ and $N_{\mathbf{y}_i}^k$ are totally identical, $|N_{\mathbf{x}_i}^k \cap N_{\mathbf{y}_i}^k| = |N_{\mathbf{x}_i}^k \cup N_{\mathbf{y}_i}^k|$ leads to $d_J(N_{\mathbf{x}_i}^k, N_{\mathbf{y}_i}^k) = 0$. Conversely, if $N_{\mathbf{x}_i}^k$ and $N_{\mathbf{y}_i}^k$ are completely different, $|N_{\mathbf{x}_i}^k \cap N_{\mathbf{y}_i}^k| = 0$ leads to $d_J(N_{\mathbf{x}_i}^k, N_{\mathbf{y}_i}^k) = 1$. Although the above distance can capture the dissimilarity of neighborhood between feature point $\mathbf{x}_i$ and $\mathbf{y}_i$, there still remains one problem. To simplify the statement, we denote $|N_{\mathbf{x}_i}^k \cap N_{\mathbf{y}_i}^k|$ as n_i $(0 \le n_i \le k)$, and $|N_{\mathbf{x}_i}^k \cup N_{\mathbf{y}_i}^k|$ is equal to $2k - n_i$. Thus, we can rewrite the Jaccard distance in (2) as

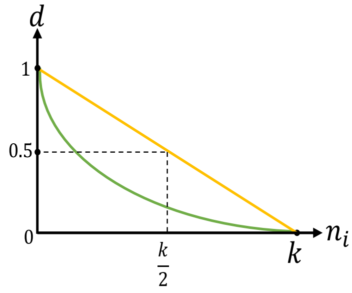


Fig. 1. Illustration of the efficient (marked in green) and original (marked in yellow) neighborhood element coherence.

$$d_J\left(N_{\mathbf{x}_i}^k, N_{\mathbf{y}_i}^k\right) = \frac{2k - 2n_i}{2k - n_i}. \tag{3}$$

It is easy to obtain that the first and second derivatives of $d_J(N_{\mathbf{x}_i}^k, N_{\mathbf{y}_i}^k)$ with respect to n_i are both less than zero. Thus, $d_J(N_{\mathbf{x}_i}^k, N_{\mathbf{y}_i}^k)$ is a decreasing and concave function. However, we hope that the value of the penalty function decreases from fast to slow as n_i increases, which means that the penalty function should be a decreasing and convex function. Fortunately, the exponential function $y = a^x$ $(0 < a < 1)$ can well meet this requirement. Therefore, we propose a simple yet efficient strategy to redefine $d_J(N_{\mathbf{x}_i}^k, N_{\mathbf{y}_i}^k)$ as

$$d_J\left(N_{\mathbf{x}_i}^k, N_{\mathbf{y}_i}^k\right) = \frac{2k - 2n_i}{2k - n_i} \cdot a^{n_i} \tag{4}$$

where a is a constant to control the degree of attenuation of the penalty curve.

Compared with the original neighborhood element coherence used in LPM [27] and LMR [32], our proposed efficient neighborhood element coherence is not sensitive to noise and outliers. As shown in Fig. 1, the relationship between the number of neighboring matches n_i with respect to the putative match $(\mathbf{x}_i, \mathbf{y}_i)$ and the distance d in LPM [27] and LMR [32] is linear, and even if $n_i = (k/2)$ holds, we have $d = 0.5$, which is still large. When the initial putative match set contains many outliers, the value of distance d for inliers is easily less than 0.5, which leads to a misjudgment about this situation. Conversely, the curve of our proposed one is nonlinear, which has "long tail" and can prevent overpenalization on the outliers.

C. Relative Order-Aware Motion Coherence

It is noticeable that the Jaccard distance aims to exploit the neighborhood element coherence and it ignores the relative order of the neighbors, which will be problematic in some cases. We show an example in Fig. 2. For a putative match $(\mathbf{x}_i, \mathbf{y}_i)$, we first extract the k nearest neighbors for $\mathbf{x}_i$ and $\mathbf{y}_i$,

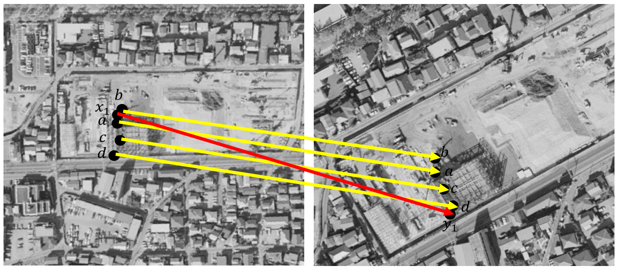


Fig. 2. Sample fails to identify outliers only using neighborhood element coherence. The arrows in red and yellow denote the identified match and its neighboring matches, respectively (red denotes outliers and yellow denotes inliers).

respectively. In this case, k and n_i are both equal to 4. According to (4), it is easy to obtain that $d_J(N_{\mathbf{x}_i}^k, N_{\mathbf{y}_i}^k) = 0$, which means zero cost. Obviously, $(\mathbf{x}_i, \mathbf{y}_i)$ is an outlier, which has been misjudged.

To conquer this problem, we propose a relative order-aware motion coherence, which mainly exploits the invariance of the relative order of inliers in a small region. Specifically, similar to [31], we denote $\sigma(\mathbf{x}_i)$ and $\sigma(\mathbf{y}_i)$ as the ranking lists, consisting of the k nearest neighbors of the feature point $\mathbf{x}_i$ and $\mathbf{y}_i$, respectively. Then, we define the overlapped ranking list with respect to $\mathbf{x}_i$ and $\mathbf{y}_i$ as

$$\hat{\sigma}(\mathbf{x}_i) = \left(\mathbf{x}_j | \mathbf{x}_j \in \sigma(\mathbf{x}_i) \cap \sigma(\mathbf{y}_i)\right) \tag{5}$$

$$\hat{\sigma}(\mathbf{y}_i) = \left(\mathbf{y}_j | \mathbf{y}_j \in \sigma(\mathbf{x}_i) \cap \sigma(\mathbf{y}_i)\right). \tag{6}$$

Obviously, we can find that $\hat{\sigma}(\mathbf{x}_i)$ and $\hat{\sigma}(\mathbf{y}_i)$ have the same elements but different relative orders according to (5) and (6). Inspired by the Levenshtein distance, the relative order-aware motion coherence between $\hat{\sigma}(\mathbf{x}_i)$ and $\hat{\sigma}(\mathbf{y}_i)$ can be defined as

$$d_S\left(\hat{\sigma}(\mathbf{x}_i), \hat{\sigma}(\mathbf{y}_i)\right) = \begin{cases} |\hat{\sigma}(\mathbf{x}_i)| & \text{if } |\hat{\sigma}(\mathbf{y}_i)| = 0 \\ |\hat{\sigma}(\mathbf{y}_i)| & \text{if } |\hat{\sigma}(\mathbf{x}_i)| = 0 \\ d_S\left(t(\hat{\sigma}(\mathbf{x}_i)), t(\hat{\sigma}(\mathbf{y}_i))\right) & \text{if } \hat{\sigma}(\mathbf{x}_i)[0] = \hat{\sigma}(\mathbf{y}_i)[0] \\ \varphi\left(\hat{\sigma}(\mathbf{x}_i), \hat{\sigma}(\mathbf{y}_i)\right) & \text{otherwise} \end{cases} \tag{7}$$

with $\varphi(\hat{\sigma}(\mathbf{x}_i), \hat{\sigma}(\mathbf{y}_i))$ defined as

$$\varphi\left(\hat{\sigma}(\mathbf{x}_i), \hat{\sigma}(\mathbf{y}_i)\right) = 1 + \min \begin{cases} \left(d_S\left(t(\hat{\sigma}(\mathbf{x}_i)), \hat{\sigma}(\mathbf{y}_i)\right) - 1\right) \\ d_S\left(\hat{\sigma}(\mathbf{x}_i), t(\hat{\sigma}(\mathbf{y}_i))\right) \\ d_S\left(t(\hat{\sigma}(\mathbf{x}_i)), t(\hat{\sigma}(\mathbf{y}_i))\right) \end{cases} \tag{8}$$

where $t(\hat{\sigma}(\mathbf{x}_i))$ is a list of all but the first element of $\hat{\sigma}(\mathbf{x}_i)$ and $\hat{\sigma}(\mathbf{x}_i)[n]$ is the nth element of the list $\hat{\sigma}(\mathbf{x}_i)$, starting with element 0. In information theory, the Levenshtein distance between two lists is the minimum number of single-element edits (e.g., deletions, insertions, and substitutions) required to change one list into the other. It is noticing that the first element in the minimum of (8) corresponds to deletion (from $\hat{\sigma}(\mathbf{x}_i)$ to $\hat{\sigma}(\mathbf{y}_i)$). Since $\hat{\sigma}(\mathbf{x}_i)$ and $\hat{\sigma}(\mathbf{y}_i)$ have the same elements, to make the modified $\hat{\sigma}(\mathbf{x}_i)$ equal to $\hat{\sigma}(\mathbf{y}_i)$, the deleted elements must be inserted at the specified position of $\hat{\sigma}(\mathbf{x}_i)$. Therefore, to avoid double counting the number of operations, the first element in the minimum of (8) should be $d_S(t(\hat{\sigma}(\mathbf{x}_i)), \hat{\sigma}(\mathbf{y}_i)) - 1$ instead of $d_S(t(\hat{\sigma}(\mathbf{x}_i)), \hat{\sigma}(\mathbf{y}_i))$.

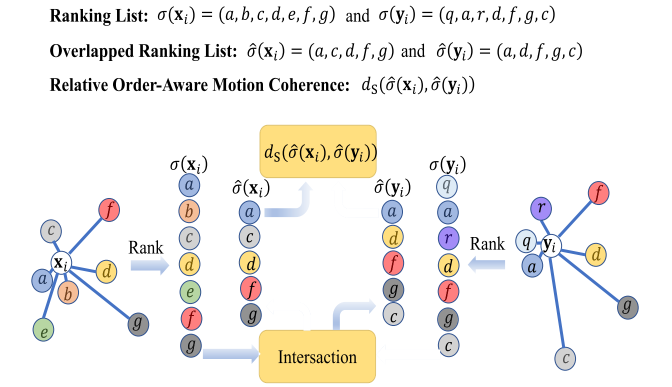

Ranking List: $\sigma(\mathbf{x}_i) = (a,b,c,d,e,f,g)$ and $\sigma(\mathbf{y}_i) = (q,a,r,d,f,g,c)$
Overlapped Ranking List: $\hat{\sigma}(\mathbf{x}_i) = (a,c,d,f,g)$ and $\hat{\sigma}(\mathbf{y}_i) = (a,d,f,g,c)$
Relative Order-Aware Motion Coherence: $d_S(\hat{\sigma}(\mathbf{x}_i),\hat{\sigma}(\mathbf{y}_i))$
$d_S(\hat{\sigma}(\mathbf{x}_i),\hat{\sigma}(\mathbf{y}_i))$
$\sigma(\mathbf{x}_i)$
$\hat{\sigma}(\mathbf{x}_i)$
$\hat{\sigma}(\mathbf{y}_i)$
$\sigma(\mathbf{y}_i)$
Rank
Rank
$\mathbf{x}_i$
$\mathbf{y}_i$
a b c d e f g
a c d f g
a d f g c
q a r d f g c
Intersaction

Fig. 3. Illustration of the relative order-aware motion coherence.

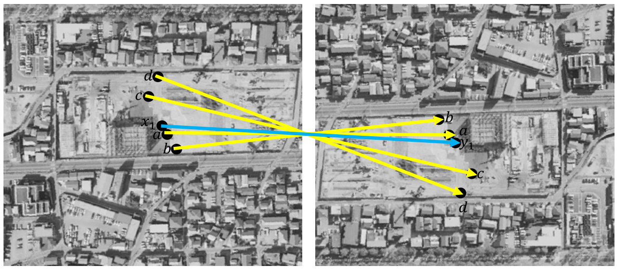


Fig. 4. Sample fails to identify inliers using the difference in length and angle between the putative match $(\mathbf{x}_i, \mathbf{y}_i)$ and its neighboring matches. The arrows in blue and yellow denote the match to be identified and its neighboring matches, respectively (blue and yellow both denote inliers).

If two lists have more elements with consistent relative order, the distance will be smaller and vice versa. To normalize $d_S(\hat{\sigma}(\mathbf{x}_i), \hat{\sigma}(\mathbf{y}_i))$, we multiply it with $1/|\hat{\sigma}(\mathbf{x}_i)|$, converting the value into $[0, 1)$.

We show an example in Fig. 3 to revisit the whole process of the relative order-aware motion coherence. First, we construct k nearest neighbors (e.g., $k = 7$) for $\mathbf{x}_i$ and $\mathbf{y}_i$ and obtain the ranking lists $\sigma(\mathbf{x}_i)$ and $\sigma(\mathbf{y}_i)$. Then, we can obtain the overlapped ranking lists $\hat{\sigma}(\mathbf{x}_i)$ and $\hat{\sigma}(\mathbf{y}_i)$ according to (5) and (6). Note that $\hat{\sigma}(\mathbf{x}_i)$ and $\hat{\sigma}(\mathbf{y}_i)$ have the same elements but different relative orders. We calculate their distance $d_S(\hat{\sigma}(\mathbf{x}_i), \hat{\sigma}(\mathbf{y}_i))$ by using (7) and (8). In this example, we just do one operation, i.e., deleting the element c from $\hat{\sigma}(\mathbf{x}_i)$ and $\hat{\sigma}(\mathbf{y}_i)$, which will make the rest of $\hat{\sigma}(\mathbf{x}_i)$ and $\hat{\sigma}(\mathbf{y}_i)$ to be identical.

Therefore, the value of $d_S(\hat{\sigma}(\mathbf{x}_i), \hat{\sigma}(\mathbf{y}_i))$ is 0.2 in this case, which is small and make it easy for the putative match $(\mathbf{x}_i, \mathbf{y}_i)$ to be regarded as an inlier with a fixed threshold.

By contrast, recently published [27] and its variants [32], [42], [43] solve the problem in Fig. 2 by comparing the difference in length and angle between the putative match $(\mathbf{x}_i, \mathbf{y}_i)$ and its neighboring matches. However, this strategy is seriously sensitive to large rotation changes. As shown in Fig. 4, the match $(\mathbf{x}_i, \mathbf{y}_i)$ is not consistent with its neighboring matches both in length and angle. Therefore, it is easy to be regarded as an outlier. On the contrary, according to our relative order-aware motion coherence, the overlapped ranking lists $\hat{\sigma}(\mathbf{x}_i)$ and $\hat{\sigma}(\mathbf{y}_i)$ are completely identical both in elements and their relative orders. Thus, $d_S(\hat{\sigma}(\mathbf{x}_i), \hat{\sigma}(\mathbf{y}_i))$ is equal to zero, which means that the relative order of neighboring elements for inliers has been well preserved with zero cost. Accordingly, match $(\mathbf{x}_i, \mathbf{y}_i)$ can be easily regarded as an inlier by our proposed strategy.

D. Objective Function and Solution

Based on the above two coherences, we denote $\mathcal{I}$ as the unknown inlier set and formulate the outlier removal problem into a mathematical model with the objective function as

$$\mathcal{I}^* = \arg\min_{\mathcal{I}} C(\mathcal{I}; S, \lambda) \tag{9}$$

where C is the cost function

$$\begin{aligned} C(\mathcal{I}; S, \lambda) = &\sum_{i\in\mathcal{I}} \left(d_J\left(N_{\mathbf{x}_i}^k, N_{\mathbf{y}_i}^k\right) + d_S\left(\hat{\sigma}(\mathbf{x}_i), \hat{\sigma}\left(\mathbf{y}_i\right)\right)\right) \\ &+ \lambda(N - |\mathcal{I}|) \end{aligned} \tag{10}$$

where $d_J(N_{\mathbf{x}_i}^k, N_{\mathbf{y}_i}^k)$ is the efficient neighborhood element coherence and $d_S(\hat{\sigma}(\mathbf{x}_i), \hat{\sigma}(\mathbf{y}_i))$ is the relative order-aware motion coherence. The first term of the cost function penalizes any match that violates the above two coherence conditions and the second is a regularization term that discourages outliers, with the parameter $\lambda > 0$ controlling the tradeoff between the two terms. To indicate the correctness of each putative match, we introduce an $N \times 1$ binary vector $\mathbf{p} = [p_1, p_2, \ldots, p_n]$, where $p_i \in \{0, 1\}$. Specifically, $p_i = 1$ points to an inlier and $p_i = 0$ points to an outlier. Therefore, the cost function can be rewritten as

$$\begin{aligned} C(\mathbf{p}; S, \lambda) = &\sum_{i=1}^{N} p_i \left(d_J\left(N_{\mathbf{x}_i}^k, N_{\mathbf{y}_i}^k\right) + d_S\left(\hat{\sigma}(\mathbf{x}_i), \hat{\sigma}\left(\mathbf{y}_i\right)\right)\right) \\ &+ \lambda\left(N - \sum_{i=1}^{N} p_i\right). \end{aligned} \tag{11}$$

Since the distribution and ratio of inliers are different for different image pairs, the optimal value of k may vary, rendering a fixed value of k unsuitable for general feature matching task. Therefore, considering using multiscale $\mathbf{K} = \{k_m\}_{m=1}^M$ nearest neighbors for the neighborhood construction, the cost function turns out to be

$$\begin{aligned} C(\mathbf{p}; S, \lambda) = &\sum_{i=1}^{N} \frac{p_i}{M} \sum_{m=1}^{M} \left(d_J\left(N_{\mathbf{x}_i}^{k_m}, N_{\mathbf{y}_i}^{k_m}\right) + d_S\left(\hat{\sigma}(\mathbf{x}_i), \hat{\sigma}\left(\mathbf{y}_i\right)\right)\right) \\ &+ \lambda\left(N - \sum_{i=1}^{N} p_i\right) \end{aligned} \tag{12}$$

where $1/M$ is used to normalize the contribution of each scale of neighborhood. With the cost function defined above, the outlier removal problem has been converted into an optimization problem. We reorganize the objective function (12) by merging the terms related to p_i and obtain

$$C(\mathbf{p}; S, \lambda) = \sum_{i=1}^{N} p_i(c_i - \lambda) + \lambda N \tag{13}$$

where

$$c_i = \frac{1}{M} \sum_{m=1}^{M} \left(d_J\left(N_{\mathbf{x}_i}^{k_m}, N_{\mathbf{y}_i}^{k_m}\right) + d_S\left(\hat{\sigma}(\mathbf{x}_i), \hat{\sigma}\left(\mathbf{y}_i\right)\right)\right). \tag{14}$$

Algorithm 1 Pseudocode of the PMC Algorithm

```
Input: putative set $S = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$, parameters $\mathbf{K}$, $a$, $\lambda$,
        $MaxIter$.
Output: optimal inlier set $\mathcal{I}^*$.
Initialize $j = 1$.
Construct neighborhood $\{N_{\mathbf{x}_i}^{k_m}, N_{\mathbf{y}_i}^{k_m}\}_{m=1,i=1}^{M,N}$ based on $S$.
Calculate $c_i$ using Eq. (17).
Obtain a subset with high ratio inliers $\mathcal{I}_j$.
repeat
    Construct neighborhood $\{N_{\mathbf{x}_i}, N_{\mathbf{y}_i}\}_{i=1}^N$ based on $\mathcal{I}_j$.
    Calculate $c_i$ using Eq. (17).
    $j = j + 1$.
    Obtain a subset with high ratio inliers $\mathcal{I}_j$.
until $j \geq MaxIter$;
Calculate $c_i$ using Eq. (14).
Determine the optimal $\mathcal{I}^*$ using Eq. (15) and (16).
```

Observing (13) carefully, we can find that those matches with $c_i > \lambda$ will increase the cost function. Conversely, the matches with $c_i < \lambda$ will decrease the cost function. Therefore, the correctness of each match can be determined by a simple strategy

$$p_i = \begin{cases} 1, & c_i \leq \lambda \\ 0, & c_i > \lambda \end{cases}, \quad i = 1, \ldots, N. \tag{15}$$

Hence, the optimal inlier set $\mathcal{I}^*$ can be determined by

$$\mathcal{I}^* = \{(\mathbf{x}_i, \mathbf{y}_i) | p_i = 1, i = 1, \ldots, N\}. \tag{16}$$

E. Implementation Details

The general pipeline of our PMC is summarized in Algorithm 1. Here, we introduce the implementation details of the PMC. Neighborhood construction for each putative match is absolutely significant. However, in the case of low ratio inliers, the neighborhood construction is unreliable. Thus, we take a coarse-to-fine strategy to deal with this problem. Specifically, we first only consider using the efficient neighborhood element coherence to construct the cost function, and c_i in (14) becomes

$$c_i = \frac{1}{M} \sum_{m=1}^{M} d_J\left(N_{\mathbf{x}_i}^{k_m}, N_{\mathbf{y}_i}^{k_m}\right) \tag{17}$$

which is simple yet efficient to reject many outliers of the putative set (the neighborhood construction is based on the whole putative set S at this time). With $\mathbf{K} = [8, 10, 12]$, $a = 0.85$, and $\lambda = 0.8$, we can obtain a subset with high ratio inliers according to (15), denoted as $\mathcal{I}_1$. After that, for each putative match, the neighborhood construction is based on $\mathcal{I}_1$ instead of the whole putative set S. We repeat the above processes two times with $\lambda = 0.5$ and 0.3 in the next two iterations and obtain $\mathcal{I}_2$ and $\mathcal{I}_3$. Note that, in the third iteration, the neighborhood construction is based on $\mathcal{I}_2$, which contains more inliers and has higher ratio inliers than $\mathcal{I}_1$. Now, $\mathcal{I}_3$ is a subset with higher ratio inliers. Thus, we construct neighborhood for each putative match based on $\mathcal{I}_3$ and construct the complete

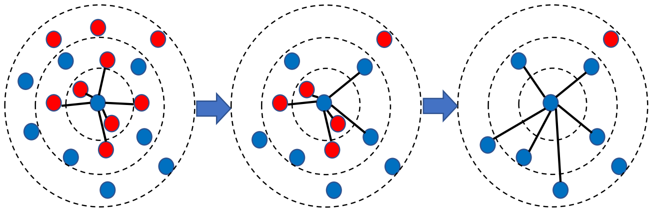


Fig. 5. Illustration of the coarse-to-fine strategy. The blue dots and red dots represent inliers and outliers, respectively.

cost function as in (12). With multiscale $\mathbf{K} = [18, 20, 22]$, $a = 0.85$, and $\lambda = 0.57$, the optimal inlier set $\mathcal{I}^*$ can be determined by (16). To better illustrate the effectiveness of the coarse-to-fine strategy, we show the process in Fig. 5. In this case, we construct six nearest neighbors for an inlier. In the beginning, there are many outliers in the neighborhood. As the iteration progresses, the constructed neighborhood based on the subset with a higher ratio of inliers is more reliable, which contains more inliers. After about two or three iterations, PMC can achieve satisfactory matching results.

IV. Experimental Results

In this section, we compare our proposed PMC on feature matching and image registration tasks with seven representative methods, such as RANSAC [14], ICF [22], MTOPK [31], LPM [27], GMS [28], RFMSCAN [29], and LAF [30]. In particular, RANSAC is a classical resampling-based method, ICF is a nonparametric interpolation-based method, MTOPK [31], LPM and GMS are motion consistency-based methods, RFMSCAN is a clustering-based method, and LAF is a filtering-based method. The parameters of the competitors are the same as the original articles. The SIFT descriptor and K-D tree provided by the open-source VLFeat toolbox [44] are used to construct initial putative matches and the k nearest neighbors for each feature point, respectively. All experiments are conducted on a laptop with Windows 10 operating systems, Intel 1.60-GHz, 16-GB RAM, and MATLAB code.

A. Datasets and Evaluation Criterion

The details of the five RS image datasets are given as follows.

1) *UAV:* This dataset comprises 43 image pairs, which were taken by an unmanned aerial vehicle, also known as UAV, flying over a patch of farmland. The image pairs can be used in the field of agricultural automatic monitoring, and they mainly suffer from projective distortions.
2) *SAR:* This dataset comprises 14 image pairs, which were taken by a satellite's synthetic aperture radars, also known as SAR. The image pairs can be used in the field of positioning and navigation, and they mainly undergo severe noise and affine distortions.
3) *PAN:* This dataset comprises 36 image pairs, which are all panchromatic aerial photographs. The image pairs can be used in the field of change detection, and they mainly involve repetitive patterns, projective, or affine distortions.
4) *CIAP:* This dataset comprises 54 image pairs, which are all color infrared aerial photographs with small overlapped areas. The image pairs can be used in the field of image mosaic, and they mainly include rigid transformation.
5) *UCD [45]:* This dataset comprises 20 image pairs, which are multispectral images taken from Sentinel-2 satellites between 2015 and 2018. The image pairs can be used in the field of change detection and they mainly suffer from low resolution, repetitive patterns, ground relief variations, and large viewpoint changes.

The UAV, SAR, PAN, and CIAP datasets are provided in [31], including the image pairs, initial putative matches (generated by SIFT descriptor), and the ground truth. The ground truth is the index of inliers, which is obtained by manually checking the correctness of each match in the initial putative matches. Note that there are duplicate matches in the given putative sets, which will severely affect the performance evaluation of the competitors, especially for those neighborhood-based methods. Therefore, we choose to keep only one match of the duplicate matches. The UCD dataset was originally designed for urban change detection and the ground truth is not suitable for our evaluation. We use the image pairs and the SIFT descriptor to construct initial putative matches and check the correctness of each putative match manually. Since we use the nearest neighbor matching strategy, the initial inlier ratio of the putative set is relatively low, which is problematic for some iterative methods. Furthermore, to better demonstrate that our proposed PMC is robust to large rotation changes, we set the rotation angle of one of the two images to be matched from 0 to 2π.

The evaluation criteria for feature matching and image registration are given as follows.

1) *Feature Matching:* The matching performance can be measured by precision, recall, and F-score, which are defined as

$$\text{Precision} = \frac{\#\text{identified correct matches}}{\#\text{preserved matches}} \tag{18}$$

$$\text{Recall} = \frac{\#\text{identified correct matches}}{\#\text{total correct matches}} \tag{19}$$

$$\text{F-score} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision+Recall}} \tag{20}$$

where # denotes the size of a set and F-score is a comprehensive evaluation of the matching performance. The larger the value of F-score is, the better the matching performance is. The identified correct matches are the true inliers retained by the method. The preserved matches are the matches that the method regards as correct ones. The total correct matches are all inliers of the image pairs.
2) *Image Registration:* The registration performance can be measured by root-mean-square error (RMSE), maximum error (MAE), and median error (MEE), which are

TABLE I

AVERAGE F-SCORE FOR OUR METHOD UNDER DIFFERENT PARAMETERS **K** ON THE UCD DATASET. THE BEST RESULT IS HIGHLIGHTED IN BOLD

K	[14,16,18]	[16,18,20]	[18,20,22]	[20,22,24]	[24,26,28]
AF	0.9249	0.9302	**0.9363**	0.9275	0.9223

defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(r_i - \mathcal{F}(s_i))^2} \tag{21}$$

$$\mathrm{MAE} = \max\left\{\sqrt{(r_i - \mathcal{F}(s_i))^2}\right\}_{i=1}^{M} \tag{22}$$

$$\mathrm{MEE} = \mathrm{median}\left\{\sqrt{(r_i - \mathcal{F}(s_i))^2}\right\}_{i=1}^{M} \tag{23}$$

where r_i and s_i denote the reference and sensed images' landmarks, respectively. $\mathcal{F}$ is the estimated mapping from the sensed image to the reference image. M is the number of selected landmarks. max(·)and median(·) return the maximum and median value of a set, respectively. Similar to [30], we choose TPS [46] as the estimated mapping since it has smooth functional mapping nature. Specifically, the identified inliers are used to the estimated mapping function $\mathcal{F}$. Then, for each pixel in the sensed image, its corresponding coordinate in the reference image can be calculated by $\mathcal{F}$, together with a bicubic interpolation method calculating the intensity at that coordinate.

B. Parameters Setting

In our method, there are three parameters: **K**, a, and λ. Parameter **K** determines the size of neighborhood of feature points. To investigate the effect of parameters on the performance of the proposed method, we choose the more challenging UCD dataset for evaluation. From Table I, we can find that our PMC achieves the best performance at $\mathbf{K} = [18, 20, 22]$. Actually, when **K** is in an appropriate range, the performance of PMC remains stable and acceptable. A small value of **K** will result in insufficient and unreliable neighborhood information statistics, while a large value of **K** will bring computational burden. Therefore, we choose $\mathbf{K} = [18, 20, 22]$ for the neighborhood construction. Parameter a is a constant to control the degree of attenuation of the penalty curve. As shown in Table II, PMC achieves the best performance at $a = 0.85$, which can prevent overpenalization on the outliers. Parameter λ can be regarded as the margin between inliers and outliers. A small value of λ can obtain a subset with high ratio inliers. Meanwhile, a large number of inliers are incorrectly discarded. A large value of λ will find more inliers but sacrifice the precision. From Fig. 6, we can find that PMC achieves the best performance at $\lambda = 0.57$. According to the aforementioned analysis, in the first and second iterations, $\mathbf{K} = [8, 10, 12]$ and $\lambda = 0.3, 0.5$ are for the consideration of efficiency.

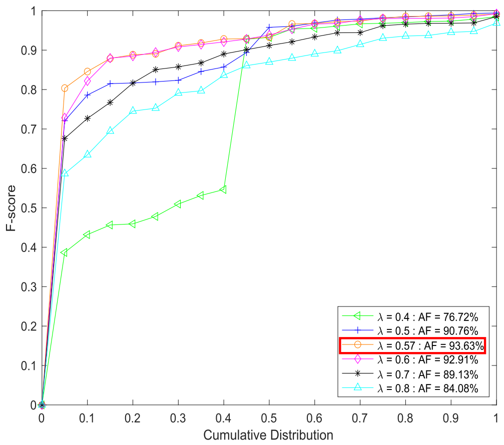


Fig. 6. Average F-score for our method under different parameters λ on the UCD dataset. The best result is indicated by a red box.

TABLE II

AVERAGE F-SCORE FOR OUR METHOD UNDER DIFFERENT PARAMETERS a ON THE UCD DATASET. THE BEST RESULT IS HIGHLIGHTED IN BOLD

a	0.75	0.8	0.85	0.9	0.95
AF	0.9041	0.9209	**0.9363**	0.9324	0.9021

C. Results on Feature Matching

1) Qualitative Analysis: We first present the feature matching results of our proposed PMC on ten representative image pairs with different challenges in Fig. 7. The transformations and degradations of the selected ten image pairs are shown in Table III. The match numbers with respect to the ten image pairs are 784, 724, 832, 1814, 1537, 1495, 2011, 2115, 1219, and 1331 and the corresponding initial inlier ratios are 44.0%, 49.7%, 70.2%, 40.5%, 28.9%, 54.8%, 14.1%, 17.0%, 32.4%, and 29.0%, respectively. The performance statistics (precision, recall, and F-score) of our PMC on these image pairs are (99.13%, 99.13%, 99.13%), (99.16%, 98.89%, 99.03%), (98.97%, 98.63%, 98.80%), (98.52%, 99.59%, 99.05%), (99.55%, 98.65%, 99.10%), (100.0%, 98.66%, 99.32%), (100.0%, 100.0%, 100.0%), (100.0%, 100.0%, 100.0%), (97.98%, 98.23%, 98.10%), and (95.44%, 99.78%, 96.54%). We can see that our proposed PMC is able to identify the vast majority of inliers and only a few matches are misjudged, even if in the case of low inlier ratio (e.g., CIAP1 and CIAP2) and large rotation changes (e.g., UCD1 and UCD2). Besides, the tests on different types of RS image pairs and different degradations also demonstrate our effectiveness and robustness.

2) Quantitative Comparison: Next, we compare our PMC with seven state of the arts (i.e., RANSAC [14], ICF [22], MTOPK [31], LPM [27], GMS [28], RFMSCAN [29], and LAF [30]) on the aforementioned five datasets. The details of match number and initial inlier ratio with respect to the above five datasets are shown in Fig. 8. The average putative match numbers are 525.67, 854.07, 1034.28, 768.42, and 1166.45 and the initial inlier ratios are 73.78%, 76.93%, 72.58%, 65.98%, and 26.20%, respectively. We report the precision, recall,

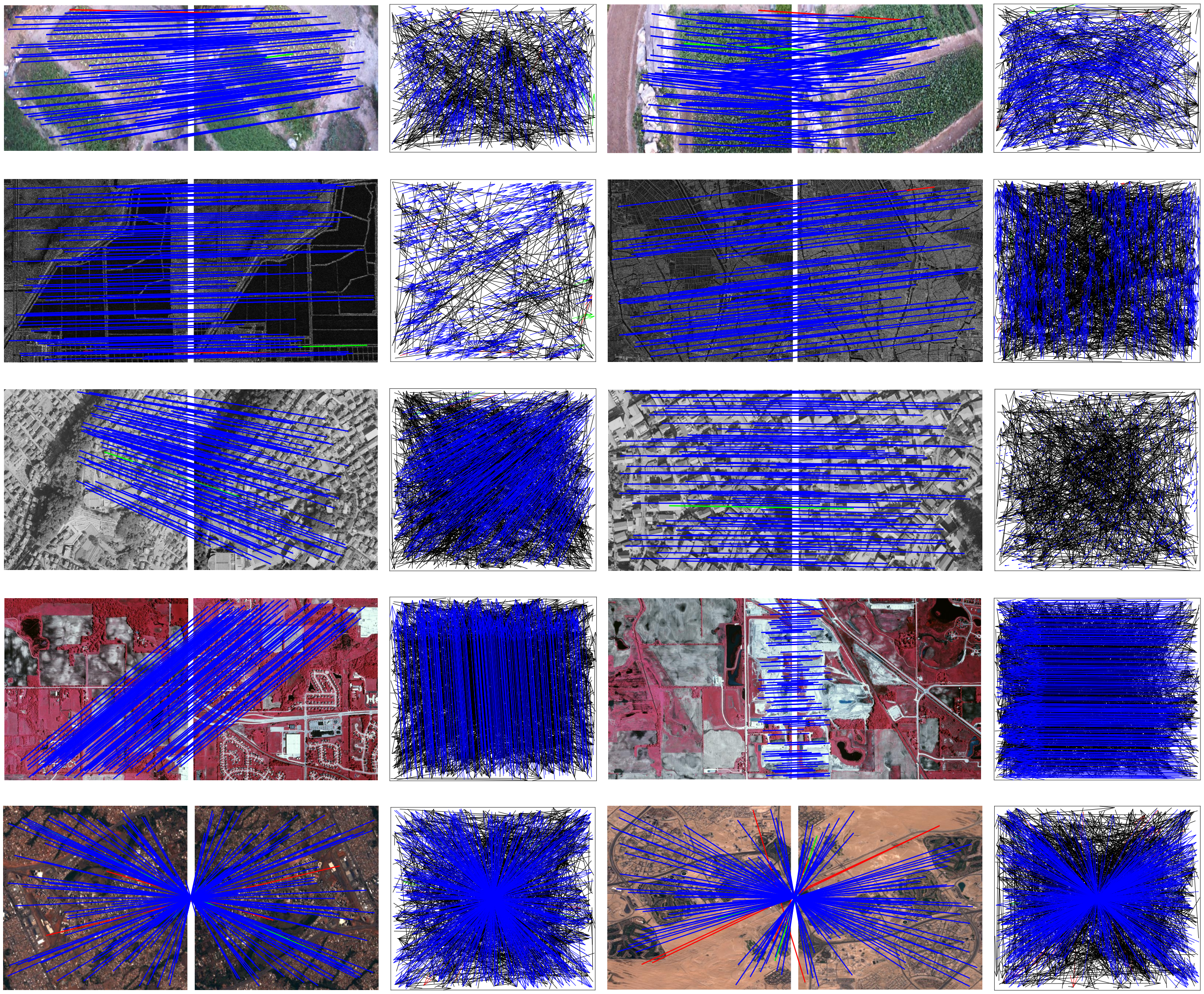


Fig. 7. Feature matching results of our proposed PMC on ten representative RS image pairs (UAV1, UAV2, SAR1, SAR2, PAN1, PAN2, CIAP1, CIAP2, UCD1, and UCD2). Each row includes the feature matching results of two image pairs (from the same dataset) and their corresponding motion vector fields. The head and tail of each arrow in the motion vector field indicate the positions of feature points in the two images (blue: true positive, black: true negative, green: false negative, and red: false positive). For visibility, we randomly select at most 100 correspondences in each image pair and the true negatives are not shown.

TABLE III

Transformations and Degradations of the Selected Ten Image Pairs

	Rigid	Projective	Affine	Ground relief variation	Severe noise	Small overlap	Repetitive structure	Low ratio inlier
UAV1		✓					✓	
UAV2		✓					✓	
SAR1			✓		✓			
SAR2			✓		✓			
PAN1		✓		✓			✓	✓
PAN2			✓	✓			✓	
CIAP1	✓					✓		✓
CIAP2	✓					✓		✓
UCD1			✓	✓	✓		✓	✓
UCD2			✓	✓	✓		✓	✓

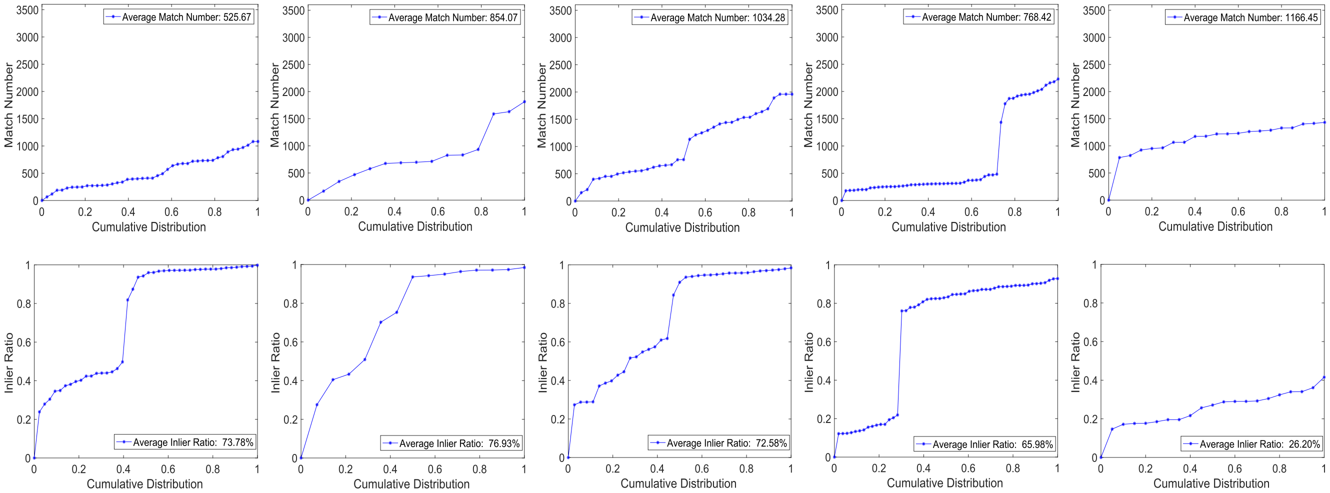


Fig. 8. Information statistics with respect to five datasets: (from Left to Right) UAV, SAR, PAN, CIAP, and UCD. The first row shows the match number of each dataset, and the second row shows the initial inlier ratio of each dataset. A point on the curve with coordinate (x, y) indicates that there are $(100*x)\%$ of image pairs whose match number/initial inlier ratio does not exceed y.

TABLE IV

Feature Matching Performance Statistics of Eight Competitors With Respect to the Five RS Datasets. A Pair of (Precision, Recall, and F-Score) Is Used for Evaluation. The Best Results Are Highlighted in Bold

	UAV	SAR	PAN	CIAP	UCD
RANSAC [14]	(0.9646,0.9740,0.9688)	(0.9596,**0.9975**,0.9778)	(0.9678,0.9948,0.9809)	(0.9352,0.9579,0.9449)	(0.8349,0.8923,0.8572)
ICF [22]	(0.9870,0.5736,0.7378)	(**0.9972**,0.3490,0.4847)	(0.9919,0.3374,0.4549)	(0.9049,0.4255,0.4655)	(**0.9186**,0.8758,0.8731)
MTOPK [31]	(0.9866,0.9913,0.9889)	(0.9838,0.9894,0.9862)	(0.9953,0.9932,0.9942)	(0.9948,0.9918,0.9932)	(0.8145,0.9660,0.8781)
LPM [27]	(0.9901,0.9863,0.9881)	(0.9883,0.9798,0.9839)	(**0.9959**,0.9790,0.9871)	(0.9745,0.9956,0.9846)	(0.8706,0.6310,0.6491)
GMS [28]	(0.9857,0.7683,0.8244)	(0.9812,0.8878,0.9250)	(0.9823,0.9417,0.9503)	(0.9960,0.9850,0.9903)	(0.4532,0.9460,0.5324)
RFMSCAN [29]	(0.9516,0.9866,0.9677)	(0.9679,0.9970,0.9820)	(0.9776,**0.9993**,0.9881)	(**0.9963**,**0.9991**,**0.9977**)	(0.4573,0.7597,0.5416)
LAF [30]	(**0.9920**,0.9831,0.9875)	(0.9893,0.9950,**0.9921**)	(0.9946,0.9944,**0.9945**)	(0.9932,0.9947,0.9939)	(0.8554,0.7008,0.7594)
PMC	(0.9901,**0.9917**,**0.9909**)	(0.9871,0.9945,0.9907)	(0.9938,0.9932,0.9935)	(0.9952,0.9990,0.9970)	(0.9100,**0.9672**,**0.9363**)

F-score, and run time of all competitors with respect to the cumulative distribution in Fig. 9.

From Fig. 9 and Table IV, we can see that RANSAC, MTOPK, LPM, RFMSCAN, LAF, and our PMC obtain satisfactory matching results on the UAV, SAR, PAN, and CIAP datasets since the initial inlier ratio of these datasets is relatively high and the smooth assumption of the motion vectors holds. ICF achieves almost perfect precision in UAV, SAR, and PAN datasets but not simultaneously for recall, which is mainly due to its parameter sensitivity and unstable spatial constraint. Similarly, GMS obtains satisfactory precision on UAV, SAR, PAN, and CIAP datasets but not simultaneously for recall. The reason is that the GMS is sensitive to the size of grid and the matches along the edges of grid will be easily discarded as outliers. MTOPK achieves competitive matching results on the UCD dataset since it is not sensitive to rotation changes. It is noticed that LPM, RFMSCAN, GMS, and LAF have bad performance on the UCD dataset. On the one hand, the low initial inlier ratio will make the neighborhood information unreliable. On the other hand, the formulations of these methods are mainly based on the smooth assumption of the motion vectors, which does not hold when the image pair suffers from large rotation changes, as shown in the last row of Fig. 7. RANSAC and ICF are rotation-invariant, and hence, they achieve good performance.

It is worth pointing out that, although our PMC does not achieve the best matching results on SAR, PAN, and CIAP datasets, the performance is still attractive and very close to the best one. Especially for UCD, PMC obtains the best and satisfactory matching results compared with other methods. For run time, LPM is the fastest one among the competitors, while GMS achieves a similar speed as LPM. RANSAC and ICF are relatively slow, especially in the case of low ratio inliers. Our PMC is moderately fast, which is mainly because of the recursive implementation of the relative order-aware motion coherence.

D. Results on Image Registration

1) Qualitative Analysis: We first give the image registration results of all competitors on five image pairs, UAV3, SAR3, PAN3, CIAP3, and UCD3 in Fig. 10. They undergo projective transformation, severe noise, repetitive structure, small overlap, and low ratio inliers. The first row of Fig. 10 shows the original image pairs, where the left image is the reference image and the right one is the sensed image. The second to the last rows are the registration results of all competitors. For each group, the left image is the checkboard result and the right one is the warped sensed image.

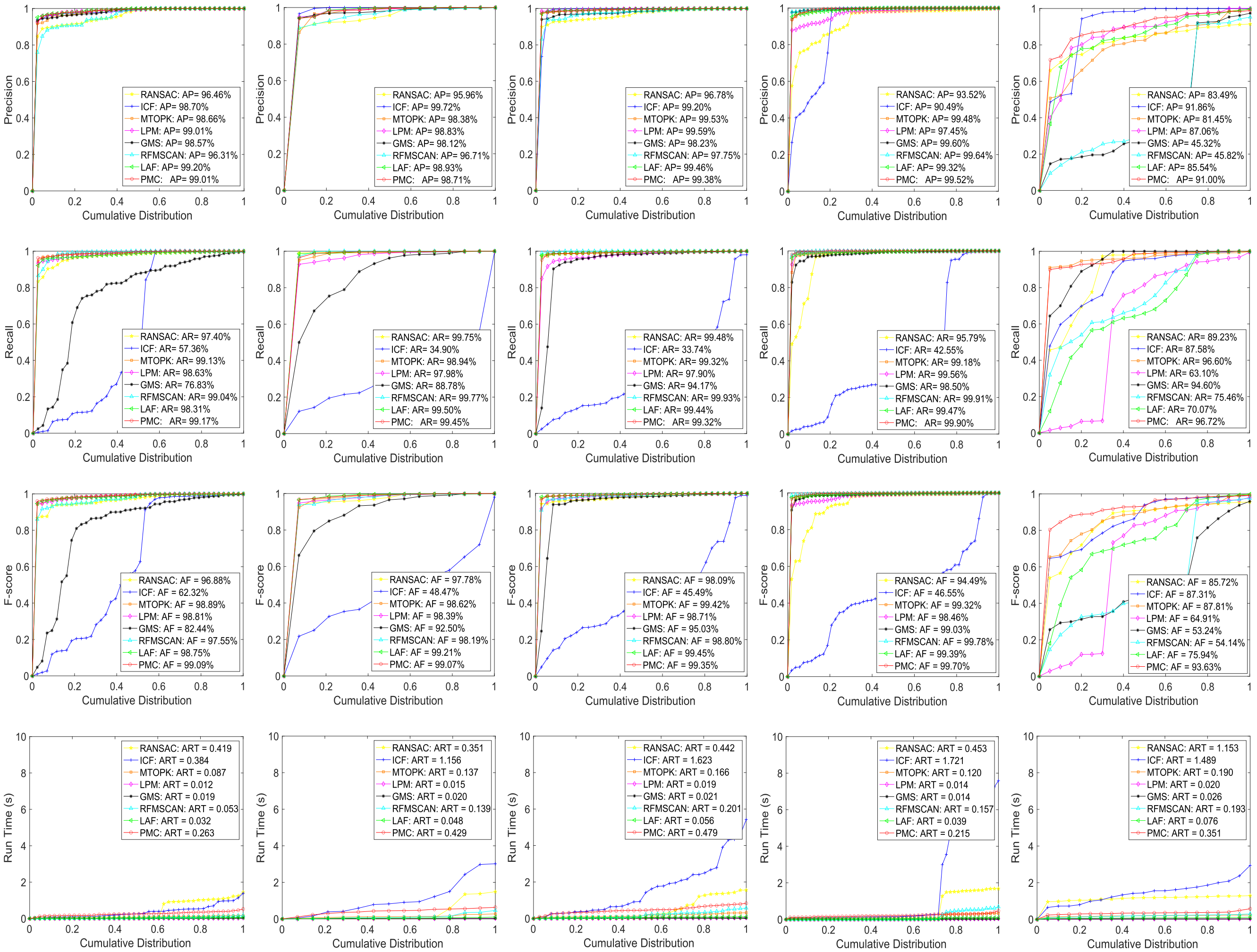


Fig. 9. Quantitative feature matching performance comparisons of RANSAC [14], ICF [22], MTOPK [31], LPM [27], GMS [28], RFMSCAN [29], LAF [30], and our proposed PMC on five datasets. (From Left to Right) UAV, SAR, PAN, CIAP, and UCD. (From Top to Bottom) Precision, recall, F-score, and run time statistics with respect to the cumulative distribution. AP: average precision. AR: average recall. AF: average F-score. ART: average run time. A point on the curve with coordinate (x, y) indicates that there are (100*x)% of image pairs whose precision/recall/F-score/run time does not exceed y.

RANSAC shows strange registration results since the matches used to recover the spatial mapping contain a lot of outliers. Besides, RANSAC can only estimate a parametric model, which is a global model and cannot deal with local deformations and nonrigid transformations. ICF can achieve good performance in many cases except for some small regions, which is mainly because of the high precision but low recall, leading to no supporting matches in some regions. LPM performs badly on CIAP3, which suffers from small overlap. The inliers preserved by LPM in the small region are not enough to estimate an accurate transformation. The registration results of MTOPK, GMS, RFMSCAN, and LAF all have some deformations. Although the motion coherences they used do not rely on any specific parametric model, the slack constraints make the preserved matches contain some outliers, especially when the initial inlier ratio is low. Our PMC uses a coarse-to-fine strategy and combines with two effective motion coherences that can preserve many inliers and only keep a few outliers, which makes the estimated transformation more accurate.

TABLE V

Quantitative Image Registration Performance Statistics of RANSAC [14], ICF [22], MTOPK [31], LPM [27], GMS [28], RFMSCAN [29], LAF [30], and Our Proposed PMC With Respect to the Aforementioned RS Datasets. The Average and Standard Deviation of RMSE, MAE, and MEE Are Used for Evaluation. The Best Results Are Highlighted in Bold

Method	RMSE	MAE	MEE
RANSAC [14]	5.992(±18.34)	11.59(±**43.33**)	4.665(±16.83)
ICF [22]	14.55(±52.20)	37.49(±133.8)	16.27(±65.27)
MTOPK [31]	24.08(±72.80)	54.78(±169.1)	30.67(±95.29)
LPM [27]	3.002(±13.65)	12.14(±55.94)	1.289(±9.255)
GMS [28]	28.38(±88.48)	64.45(±199.0)	37.57(±122.2)
RFMSCAN [29]	11.77(±41.02)	39.27(±132.7)	5.890(±38.99)
LAF [30]	5.408(±20.16)	25.29(±101.9)	1.523(±10.11)
PMC	**1.176**(±**9.974**)	**6.426**(±62.94)	**0.188**(±**0.890**)

2) Quantitative Comparison: Next, we test the image registration of all competitors on the aforementioned five datasets in a quantitative way. Similar to [30], we randomly choose 20 pairs of landmarks $\{r_i, s_i\}_{i=1}^{20}$ for the calculation of RMSE,

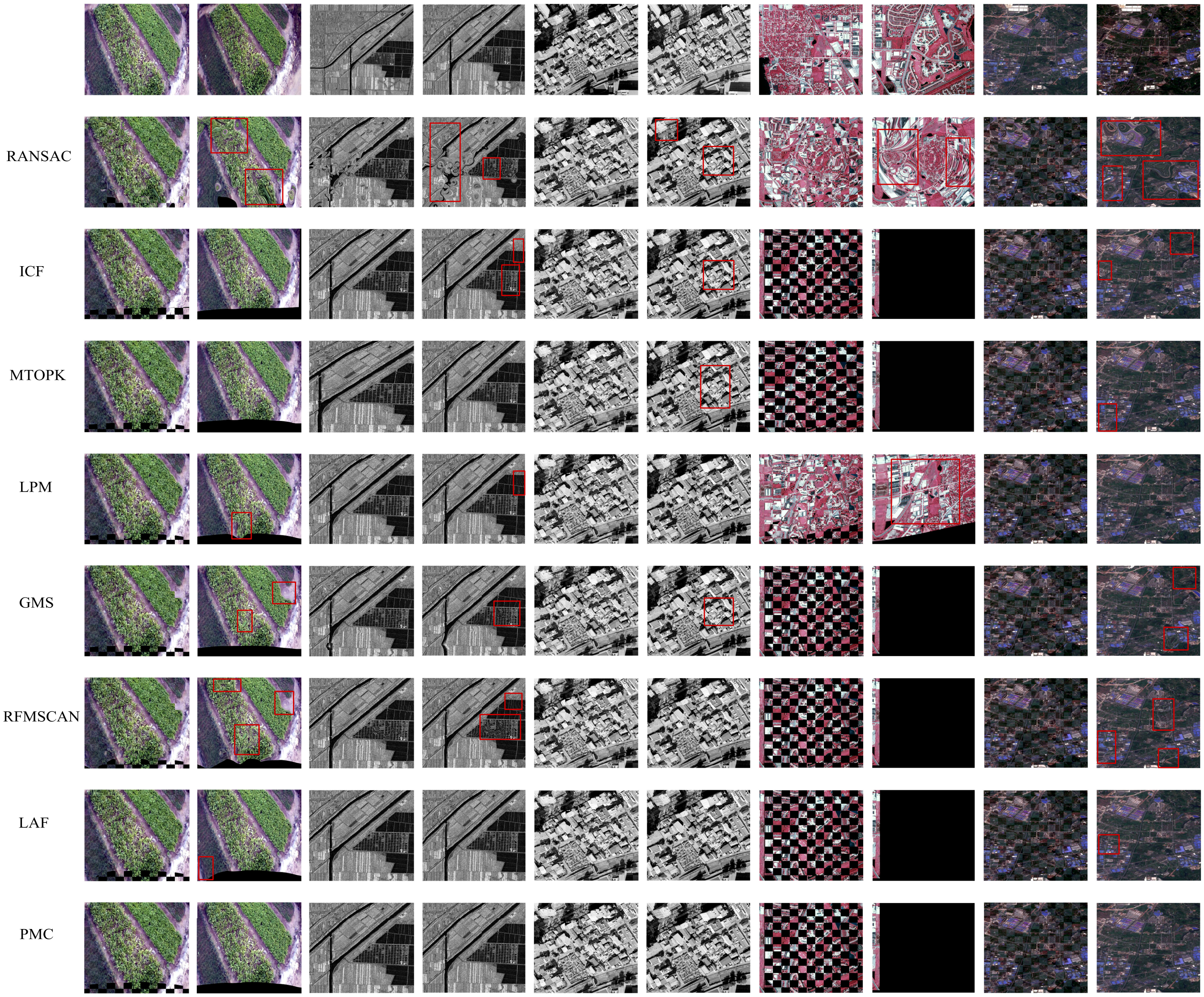


Fig. 10. Image registration results of our proposed PMC and other competitors on five representative RS image pairs (i.e., UAV3, SAR3, PAN3, CIAP3, and UCD3). The first row shows the input images (for each group, Left: reference image and Right: sensed image). The second to the last rows are the registration results of all competitors (for each group, Left: checkboard result and Right: warped sensed image). The obvious distortions of registration results are marked in red boxes.

MAE, and MEE. The average and standard deviation of RMSE, MAE, and MEE are reported in Table V and the best results are highlighted in bold. GMS obtains the worst performance for the three metrics, which is in accordance with the feature matching results. MTOPK also performs badly in the image registration task. Although ICF achieves high precision in feature matching tasks, the low recall makes the number of inliers for estimating accurate transformation not enough. As for RFMSCAN, the limited similarity measurement inevitably clusters inliers and some outliers together. RANSAC achieves stable registration performance because of its global geometrical constraint. LPM and LAF can obtain competitive performance since they can obtain sufficient inliers to estimate the transformation. Our PMC achieves the best results except for the standard deviation of MAE. Therefore, it can be inferred that PMC is robust to the image registration of different image types and degradations.

V. Conclusion

We present a simple yet efficient method for RS image matching. We formulate the matching problem into a mathematical model and derive a closed-form solution. The objective function is constructed based on two novel coherence constraints. With the interaction of these two coherence constraints, the cost distribution with regard to inliers is close to 0, while for outliers, the cost distribution is close to 1. Therefore, we can use a predefined threshold to separate inliers from outliers. Experimental results on five RS image datasets demonstrate the effectiveness and robustness of our PMC. However, the relative order-aware motion coherence is recursive, which increases the complexity of this algorithm. Though our PMC is not the fastest among the competitors, it can also accomplish outlier removal for 1000 putative matches in a few milliseconds. In short, our proposed method achieves a good tradeoff between effectiveness and efficiency.

References

[1] Y. Yuan, F. Fang, and G. Zhang, "Superpixel-based seamless image stitching for UAV images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1565–1576, Feb. 2021.

[2] H. Zhou *et al.*, "Feature matching for remote sensing image registration via manifold regularization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4564–4574, 2020.

[3] M. Zhang and W. Shi, "A feature difference convolutional neural network-based change detection method," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7232–7246, Oct. 2020.

[4] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM: A versatile and accurate monocular SLAM system," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1147–1163, Oct. 2015.

[5] H. Hu, Z. Qiao, M. Cheng, Z. Liu, and H. Wang, "DASGIL: Domain adaptation for semantic and geometric-aware image-based localization," *IEEE Trans. Image Process.*, vol. 30, pp. 1342–1353, 2021.

[6] Y. Zhou, A. Rangarajan, and P. D. Gader, "An integrated approach to registration and fusion of hyperspectral and multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3020–3033, May 2020.

[7] Y. Xiang, R. Tao, and H. You, "OS-PC: Combining feature representation and 3-D phase correlation for subpixel optical and SAR image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 9, pp. 6451–6466, Mar. 2020.

[8] W. Ma, J. Zhang, Y. Wu, L. Jiao, H. Zhu, and W. Zhao, "A novel two-step registration method for remote sensing images based on deep and local features," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4834–4843, Jul. 2019.

[9] Y. Jin *et al.*, "Image matching across wide baselines: From paper to practice," *Int. J. Comput. Vis.*, vol. 129, pp. 517–547, Oct. 2021.

[10] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, "Image matching from handcrafted to deep features: A survey," *Int. J. Comput. Vis.*, vol. 129, no. 1, pp. 23–79, 2021.

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

[12] K. M. Yi, E. Trulls, V. Lepetit, and P. Fua, "LIFT: Learned invariant feature transform," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2016, pp. 467–483.

[13] A. L. Yuille and N. M. Grzywacz, "A mathematical analysis of the motion coherence theory," *Int. J. Comput. Vis.*, vol. 3, no. 2, pp. 155–175, 1989.

[14] 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.

[15] 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, Apr. 2000.

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

[17] E. Brachmann *et al.*, "Dsac-differentiable ransac for camera localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 6684–6692.

[18] O. Chum, T. Werner, and J. Matas, "Two-view geometry estimation unaffected by a dominant plane," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 772–779.

[19] D. Barath and J. Matas, "Graph-cut RANSAC," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6733–6741.

[20] T. Lai, H. Fujita, C. Yang, Q. Li, and R. Chen, "Robust model fitting based on greedy search and specified inlier threshold," *IEEE Trans. Ind. Electron.*, vol. 66, no. 10, pp. 7956–7966, Oct. 2019.

[21] T. Lai *et al.*, "Efficient robust model fitting for multistructure data using global greedy search," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 3294–3306, Jul. 2020.

[22] X. Li and Z. Hu, "Rejecting mismatches by correspondence function," *Int. J. Comput. Vis.*, vol. 89, no. 1, pp. 1–17, 2010.

[23] 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.

[24] G. Wang and Y. Chen, "SCM: Spatially coherent matching with Gaussian field learning for nonrigid point set registration," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 203–213, Jan. 2021.

[25] J. Ma, J. Wu, J. Zhao, J. Jiang, H. Zhou, and Q. Z. Sheng, "Nonrigid point set registration with robust transformation learning under manifold regularization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 12, pp. 3584–3597, Dec. 2019.

[26] C. Yang *et al.*, "Non-rigid point set registration via adaptive weighted objective function," *IEEE Access*, vol. 6, pp. 75947–75960, 2018.

[27] 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.

[28] J. Bian *et al.*, "GMS: Grid-based motion statistics for fast, ultra-robust feature correspondence," *Int. J. Comput. Vis.*, vol. 128, no. 6, pp. 1580–1593, Jun. 2020.

[29] 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.

[30] X. Jiang *et al.*, "Robust feature matching for remote sensing image registration via linear adaptive filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1577–1591, Feb. 2021.

[31] 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.

[32] 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.

[33] Y. Li, Q. Huang, Y. Liu, Y. Huang, and X. Sun, "Efficient properties-based learning for mismatch removal," *IEEE Access*, vol. 7, pp. 149612–149622, 2019.

[34] K. M. Yi, E. Trulls, Y. Ono, V. Lepetit, M. Salzmann, and P. Fua, "Learning to find good correspondences," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2666–2674.

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

[36] J. Zhang *et al.*, "Learning two-view correspondences and geometry using order-aware network," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 5845–5854.

[37] Y. Liu, L. Liu, C. Lin, Z. Dong, and W. Wang, "Learnable motion coherence for correspondence pruning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 3236–3245.

[38] L. Dai *et al.*, "MS2DG-Net: Progressive correspondence learning via multiple sparse semantics dynamic graph," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2022, pp. 8973–8982.

[39] D. DeTone, T. Malisiewicz, and A. Rabinovich, "Superpoint: Self-supervised interest point detection and description," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Aug. 2018, pp. 224–236.

[40] Y. Liu *et al.*, "Motion consistency-based correspondence growing for remote sensing image matching," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2021, doi: 10.1109/LGRS.2020.3048258.

[41] Y. Liu *et al.*, "Robust feature matching via advanced neighborhood topology consensus," *Neurocomputing*, vol. 421, pp. 273–284, Jan. 2021.

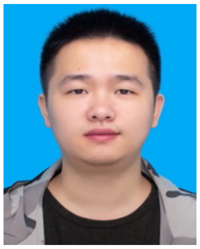
[42] G. Xiao, J. Ma, S. Wang, and C. Chen, "Deterministic model fitting by local-neighbor preservation and global-residual optimization," *IEEE Trans. Image Process.*, vol. 29, pp. 8988–9001, 2020.

[43] Y. Liu, B. N. Zhao, and S. Zhao, "Rectified neighborhood construction for robust feature matching with heavy outliers," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Jun. 2022, Art. no. 6515005, doi: 10.1109/LGRS.2022.3181754.

[44] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," in *Proc. Int. Conf. Multimedia*, 2010, pp. 1469–1472.

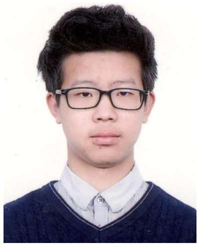
[45] R. C. Daudt, B. L. Saux, A. Boulch, and Y. Gousseau, "Urban change detection for multispectral earth observation using convolutional neural networks," in *Proc. Int. Geosci. Remote Sens. Symp.*, Jul. 2018, pp. 2115–2118.

[46] J. Ma, J. Zhao, Y. Zhou, and J. Tian, "Mismatch removal via coherent spatial mapping," in *Proc. 19th IEEE Int. Conf. Image Process.*, Sep. 2012, pp. 1–4.



Yizhang Liu received the B.S. degree in electronic and information engineering and the master's degree in computer science and technology from Fujian Agriculture and Forestry University, Fuzhou, China, in 2013 and 2017, respectively. He is currently pursuing the D.Eng. degree with the School of Software Engineering, Tongji University, Shanghai, China.

His research interests include computer vision and image matching.



Brian Nlong Zhao is currently pursuing the bachelor's degree in computer engineering and computer science and applied mathematics with the University of Southern California, Los Angeles, CA, USA.

His research interests are machine learning and computer vision.

Shengjie Zhao (Senior Member, IEEE) received the B.S. degree in electrical engineering from the University of Science and Technology of China, Hefei, China, in 1988, the M.S. degree in electrical and computer engineering from the China Aerospace Institute, Beijing, China, in 1991, and the Ph.D. degree in electrical and computer engineering from Texas A&M University, College Station, TX, USA, in 2004.

He is currently the Dean of the College of Software Engineering and a Professor with the College of Software Engineering and the College of Electronics and Information Engineering, Tongji University, Shanghai, China. In previous postings, he conducted research at Lucent Technologies, Whippany, NJ, USA, and the China Aerospace Science and Industry Corporation, Beijing. He is a fellow of the Thousand Talents Program of China and an Academician of the International Eurasian Academy of Sciences. His research interests include artificial intelligence, big data, wireless communications, image processing, and signal processing.

Lin Zhang (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees from the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China, in 2003 and 2006, respectively, and the Ph.D. degree from the Department of Computing, The Hong Kong Polytechnic University, Hong Kong, in 2011.

From March 2011 to August 2011, he was a Research Associate with the Department of Computing, The Hong Kong Polytechnic University. In August 2011, he joined the School of Software Engineering, Tongji University, Shanghai, where he is currently a Full Professor. His research interests include environment perception of intelligent vehicles, pattern recognition, computer vision, and perceptual image/video quality assessment.

Dr. Zhang serves as an Associate Editor for IEEE ROBOTICS AND AUTOMATION LETTERS and *Journal of Visual Communication and Image Representation*. He was awarded as a Young Scholar of Changjiang Scholars Program, Ministry of Education, China.