Multiscale Locality and Rank Preservation for Robust Feature Matching of Remote Sensing Images

Xingyu Jiang, Junjun Jiang, Aoxiang Fan, Zhongyuan Wang, and Jiayi Ma

***Abstract*— As a fundamental and important task in many applications of remote sensing and photogrammetry, feature matching tries to seek correspondences between the two feature sets extracted from an image pair of the same object or scene. This paper focuses on eliminating mismatches from a set of putative feature correspondences constructed according to the similarity of existing well-designed feature descriptors. Considering the stable local topological relationship of the potential true correspondences, we propose a simple yet efficient method named multiscale Top K Rank Preservation (mTopKRP) for robust feature matching. To this end, we first search the K-nearest neighbors of each feature point and generate a ranking list accordingly. Then we design a metric based on the weighted Spearman's footrule distance to describe the similarity of two ranking lists specifically for the matching problem. We build a mathematical optimization model and derive its closed-form solution, enabling our method to establish reliable correspondences in linearithmic time complexity, which requires only tens of milliseconds to handle over 1000 putative matches. We also introduce a multiscale strategy for neighborhood construction, which increases the robustness of our method and can deal with different types of degradation, even when the image pair suffers from a large scale change, rotation, nonrigid deformation, or a large number of mismatches. Extensive experiments on several representative remote sensing image data sets demonstrate the superiority of our method over state of the art.**

***Index Terms*— Feature matching, local structure, multiscale, nonrigid, ranking list.**

I. Introduction

FEATURE matching, which aims to establish reliable feature point (e.g., pixel coordinate) correspondences from two images of the same object or scene, is a fundamental and crucial problem for many vision-based tasks, especially for the applications of remote sensing and photogrammetry, including image registration and fusion [1]–[3], panoramic image mosaic [4], [5], 3-D reconstruction [6], [7], object identification, and tracking [8], [9]. These matching-based tasks are typically accomplished based on a robust and efficient feature matching method to obtain as many correct matches as possible while keeping the mismatches to a minimum.

Feature matching is a combinatorial nature problem, which causes highly computation complexity in non-Pareto criterion complex optimization. Specifically, matching N points to another N points may lead to a total of $N!$ permutations [10]. To address this issue, the common idea of many existing methods is to solve the feature matching problem in a two-stage manner, which first constructs a set of putative matches based on the similarity of the local patch descriptors associated with the distinct feature points, such as the local pixels' gradients, intensity values, or other descriptions of extracted distinct and stable extremum points [11]–[13]. However, in addition to some correct correspondences (i.e., inliers), there are a large number of false matches (i.e., outliers) in the putative set as well due to the ambiguities of local descriptors, particularly when the image pairs suffer from the low-quality, occlusion, or repetitive patterns. Therefore, in the next step, it is critical to filter out the mismatches from the constructed putative match set using additional constraints.

Many existing mismatch removal methods are usually based on a geometrical constraint, which regards the putative match set as a mapping from inputs to outputs by satisfying a predefined geometrical transformation model [14]. However, due to the ground relief variations, imaging viewpoint changes, captured at low-altitude or by a fisheye (FE) camera, remote sensing images often involve local distortions which result in complex spatial relationships. Therefore, the transformation model can vary with respect to different data and is usually unknown beforehand, especially if it is nonrigid. In this situation, using a predefined geometrical model will probably lead to inferior matching accuracy. Another obstacle of existing methods is the large computational complexity caused by complex nonrigid transformation models, which is not applicable for dealing with large-scale or real-time tasks, for example, unmanned aerial vehicle (UAV)-based tasks including 3-D land surface reconstruction, large-area image mosaic, moving object detection, and tracking [4], [6], [15].

To address the above-mentioned challenges, we propose a robust method for remote sensing image matching based on the similarity between the two local structures, which are composed by the K-nearest neighbors (K-NN) of the two matched feature points, respectively. In particular, the relationship

Manuscript received December 13, 2018; revised January 30, 2019; accepted March 16, 2019. Date of publication April 17, 2019; date of current version August 27, 2019. This work was supported by the National Natural Science Foundation of China under Grant 61773295 and Grant 61671332. *(Corresponding author: Jiayi Ma.)*

X. Jiang, A. Fan, and J. Ma are with the Electronic Information School, Wuhan University, Wuhan 430072, China (e-mail: jiangx.y@whu.edu.cn; fanaoxiang@whu.edu.cn; jyma2010@gmail.com).

J. Jiang is with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China (e-mail: junjun0595@163.com).

Z. Wang is with the National Engineering Research Center for Multimedia Software, School of Computer, Wuhan University, Wuhan 430072, China (e-mail: wzy_hope@163.com).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2019.2906183

among the neighboring feature points in a remote sensing image is usually stable and only changed slightly, even when the image undergoes scale change, rotation, or nonrigid transformation [1]. Based on this observation, we develop a measurement to describe the similarity of two local structures and build a mathematical optimization model accordingly to identify the false matches. Our model is general which does not require a predefined transformation model, and it is also quite efficient with a closed-form solution. This work is an extension of our previous works in [16] and [17]. Compared to our previous works, in this paper, we redefine the ranking list distance measurement based on multiscale neighborhoods to preserve the local topological structure more strictly and generally, which makes our method more robust to outliers and different types of degradation.

In summary, our major contribution is twofold. On the one hand, a simple yet efficient method is proposed to address the robust feature matching problem. Compared with existing approaches, our method removes mismatches merely by preserving the local topological structure of the potential inliers, which is independent of any predefined global image transformation with a specific parametric or nonparametric model, and hence, it is more general and robust to different kinds of image transformations. On the other hand, we introduce a general mathematical optimization model for the mismatch removal problem and derive its closed-form solution. Based on this model, any criteria for describing the latent difference between inliers and outliers can be integrated into our formulation. Moreover, the closed-form solution enables us to establish the reliable correspondences in linearithmic time complexity, which requires only tens of milliseconds to handle over 1000 putative matches.

The remainder of this paper is organized as follows. Section II describes the background material and related works. In Section III, we introduce our multiscale Top K Rank Preservation (mTopKRP) algorithm in detail, including the measurement of K ranking similarity, a general objective function and its closed-form solution. Section IV provides the qualitative and quantitative evaluations of our method in comparison with several state-of-the-art methods on different types of remote sensing image data sets. Finally, some concluding remarks are summarized in Section V.

II. Related Works

In general, the problem of feature matching can be solved in a two-stage manner. In the first stage, salient features, i.e., control points and their descriptors, are extracted from each image, and a set of putative point correspondences is constructed by using a similarity constraint of descriptors. Classic methods addressing this stage include scale-invariant feature transform (SIFT) [11], speeded up robust features (SURF) [12], and oriented FAST and rotated BRIEF (ORB) [13], which have been proven to be both efficient and effective. SIFT detects feature points in Gaussian-scale space and uses gradient histogram to form descriptors, which is known to be scale, viewpoint, rotation, and illumination invariant. SURF improves SIFT by using the Hessian matrix for accurate feature detection and an integral image strategy for efficiency. Specifically, Li *et al.* [18] refined the descriptor of SIFT to overcome the difference in the gradient intensity and orientation between remote image pairs. In contrast, ORB adopts a different scheme for both feature detection and description. It achieves a much higher speed by using FAST detector [19] and BRIEF descriptor [20], while the scale invariance property is sacrificed. However, due to the ambiguities of descriptor matching, this putative correspondence set is typically contaminated by a large number of outliers. Therefore, the second stage is desired to remove the outliers with a global geometric constraint, resulting in an accurate feature matching result. In this paper, we assume that the putative correspondences are constructed, and then we focus on the mismatch removal problem. A variety of methods addressing this problem have been proposed in the literature, and here, we give a brief review of them.

The mismatch removal methods can be roughly divided into two categories, i.e., resampling and nonparametric fitting methods. Random sample consensus (RANSAC) [14] is one of the most representative methods in the literature, which tries to find the largest outlier-free correspondence set that conforms to a predefined parametric model by randomly resampling. Inspired by RANSAC, maximum likelihood estimation sample consensus (MLESAC) [21] and progressive sample consensus (PROSAC) [22] are proposed as effective variants. Although resampling methods have achieved great success for the feature matching problem, they also have some limitations. When the motion of image scene is nonrigid, which cannot be characterized by a parametric model, these methods will not be valid. To this end, the nonparametric fitting methods are introduced, such as vector field consensus (VFC) [23], identify correspondence function (ICF) [24], and robust L_2 estimator [25]. By using M-estimators or nonlinear regression techniques, these methods are able to conduct robust fitting of nonrigid motion field. The correspondences that are consistent with the recovered motion are identified as inliers. However, their performance will inevitably degrade in the presence of a large proportion of outliers.

Another strategy for feature matching is to directly establish correspondences between the two feature sets. These methods also involve two categories. Iterative closest point (ICP) [26] and coherent point drift (CPD) [27] are the representatives of the first category, which aims to estimate the transformation model between the two feature sets. In particular, with a predefined parametric or nonparametric model, these methods iterate between the two update processes, i.e., correspondence establishment and transformation estimation, until convergence. The feature correspondences and the transformation model then can be determined simultaneously. Gaussian mixture model-based registration (GMMREG) [28] provides a universal framework for point set registration, which aims to minimize a statistical discrepancy measurement between the two Gaussian mixture distribution of the two point sets. The second category is known as graph matching methods [29], [30]. Spectral matching (SM) [31], SM with affine constraint (SMAC) [32], and graph shift (GS) [33] are the representatives in this category. Generally, these methods

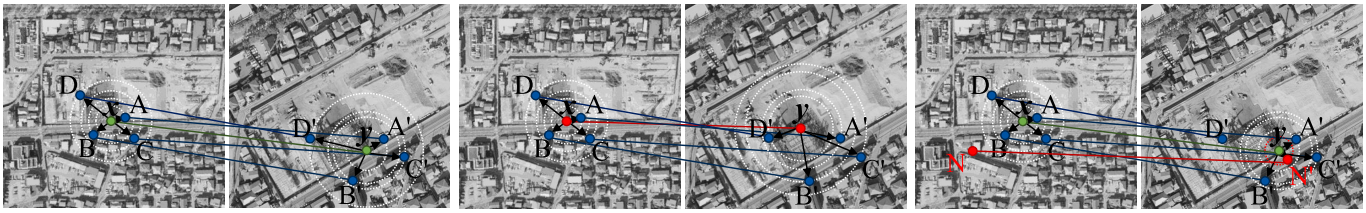


Fig. 1. Illustration of the local structure ranking list similarity measurement in our method. There are three examples shown above to judge whether the putative match $(\mathbf{x}, \mathbf{y})$ is correct or not. We search the K-NNs ($K = 4$) of feature points $\mathbf{x}$ and $\mathbf{y}$ from the putative point set. We can obtain the neighboring ranking lists of $\mathbf{x}$ and $\mathbf{y}$ for these three examples, which are $\{\sigma(\mathbf{x}) = (A, B, C, D), \sigma(\mathbf{y}) = (A', B', C', D')\}$, $\{\sigma(\mathbf{x}) = (A, B, C, D), \sigma(\mathbf{y}) = (D', A', B', C')\}$, and $\{\sigma(\mathbf{x}) = (A, B, C, D), \sigma(\mathbf{y}) = (N', A', B', C')\}$, respectively. Regarding the matched points like (A, A') as the same item, we can calculate the local structure difference or cost with locality preserving matching (LPM) [16], TopKRP [17], and the improved TopKRP in this paper (i.e., mTopKRP) for each example, and we denote them as $[C_{\text{LPM}}, C_{\text{TopKRP}}, C_{\text{mTopKRP}}]$ accordingly. The costs are [0, 0, 0] (left, true match), [0, 0.3452, 0.3452] (middle, false match), and [0.25, 0.4595, 0.25] (right, true match), respectively. The putative matches are denoted with full lines, where blue and green lines indicate true match, and red line indicates false match. The dotted circles indicate the Euclidean distance range centered on the object point $\mathbf{x}$ or $\mathbf{y}$.

do not rely on a specific transformation model. Instead, they formulate the feature matching problem as an integer quadratic programming (IQP) problem, which enforces the preservation of geometric structures between the two feature sets. The correspondences are determined by the optimal solution of it. However, methods of these two categories throw out the local image information of features (e.g., descriptors), and utilize only the spatial information. Therefore, their performance may be degraded to some extent, meanwhile, the time cost is high as well.

In the remote sensing community, the feature matching literature also includes a variety of methods [34]–[37]. For example, locally linear transforming (LLT) [1] is proposed to address the mismatch removal problem, which accommodates both parametric and nonparametric transformation models, and a novel local structure constraint is developed. Wen *et al.* [38] introduced a unified feature match criterion, which combines spatial consistency and feature similarity. In addition, a graph matching-based method, restricted spatial order constraint (RSOC) [39], is proposed to accomplish accurate point matching based on the spatial order constraints. In the recent past, Zhou *et al.* [40] used a probabilistic method with global and local regularization terms to differentiate true and false correspondences. Li *et al.* [5] proposed to use support-line voting to filter mismatches and affine-invariant ratios to subsequently refine the matching results and is able to discover more correct matches than initial matches.

Most recently, several novel methods were proposed to address the feature matching problem by using the piecewise-smoothness constrains and local structure consistence, such as LPM [16], [41], [42], coherence-based decision boundaries [7], grid-based motion statistics [43], and deep learning method [44], which have achieved promising performance in terms of both accuracy and efficiency. Specifically, the key observation of LPM is that even under severe deformation, the local topological elements are generally well preserved to some extent. To achieve scale and rotation invariant, LPM considers the spatial neighborhood relationship of feature points and develops a robust and surprisingly effective mismatch removal method.

Although our previously developed LPM [16] and TopKRP [17] methods preserve the local topological structures of two putatively matched feature points from two images, their differences lie in the definition of the topological similarity measurement between the two feature points. Specifically, LPM defines the local structure similarity between feature points by merely counting the intersection of K-NN. Obviously, LPM ignores the differences (i.e., specific position relationship) among the neighboring elements and cannot exploit the true topological structure. In other words, it is not strict enough for the constraint in LPM to preserve the local structure, e.g., as the first two examples shown in Fig. 1, where the costs generated by LPM on these two situations are both 0, while the putative match $(\mathbf{x}, \mathbf{y})$ in these two image pairs is correct and incorrect, respectively. By contrast, in the TopKRP [17], we present a more strict measuring criterion for local structure preservation by transforming the putative matched feature points from the feature space to the ranking list space, and the topological structure similarity of two feature points is then measured by comparing their ranking lists, which can identify inliers and outliers more accurately for the first two examples in Fig. 1.

III. Method

To establish reliable feature correspondences between the two remote sensing images, we first construct a set of putative matches by using the SIFT algorithm [11], and then the matching task boils down to remove false matches from the given putative set. Therefore, in the following, we will focus on the mismatch removal based on a K-NN ranking similarity for preserving the local topology structure.

A. Top K Rank Similarity Measurement

To measure the local elements' ranking similarity, we first determine the K-NN of each feature point. For each putative match $(\mathbf{x}, \mathbf{y})$, we can obtain two ranking lists of the feature points $\mathbf{x}$ and $\mathbf{y}$ denoted as $\sigma(\mathbf{x})$ and $\sigma(\mathbf{y})$, respectively. The difference between two ranking lists can be defined based on the weighted Spearman's footrule distance [45]. We denote the top K rank difference as $D_K(\sigma(\mathbf{x}), \sigma(\mathbf{y}))$[1] with the following

[1]The returned value of D_K is normalized within [0, 1] for the denominator in (1). $D_K = 0$ when the two lists are identical, and $D_K = 1$ when they are completely disjoint.

form:

$$D_K(\sigma(\mathbf{x}),\sigma(\mathbf{y})) = \frac{1}{\Phi_K}\sum_{t\in\mathcal{T}}\phi_t \tag{1}$$

where $\mathcal{T} = \sigma(\mathbf{x}) \cup \sigma(\mathbf{y})$ with the cardinality $|\mathcal{T}| \in [K, 2K]$, and Φ_K is the weighted Spearman's footrule measurement between the two full ranked lists $\sigma(\mathbf{x})$ and $\sigma(\mathbf{y})$, which is used for normalizing $\sum_{t\in\mathcal{T}}\phi_t$ into $[0, 1]$ with the following form:

$$\Phi_K = -2K + 2z\sum_{k=1}^{K} 1/k. \tag{2}$$

In (1), the ranking distance contribution of item t to Φ_K is

$$\phi_t = left\{\begin{cases}\frac{\|\sigma_t(\mathbf{x})-\sigma_t(\mathbf{y})\|_1}{\min\{\sigma_t(\mathbf{x}),\sigma_t(\mathbf{y})\}}, & t\in\sigma(\mathbf{x})\cap\sigma(\mathbf{y})\\ \frac{\|\sigma_t(\mathbf{x})-z\|_1}{\min\{\sigma_t(\mathbf{x}),z\}}, & t\in\sigma(\mathbf{x}), t\notin\sigma(\mathbf{y})\\ \frac{\|\sigma_t(\mathbf{y})-z\|_1}{\min\{\sigma_t(\mathbf{y}),z\}}, & t\notin\sigma(\mathbf{x}), t\in\sigma(\mathbf{y})\end{cases}\right. \tag{3}$$

where $\sigma_t(\mathbf{x})$ and $\sigma_t(\mathbf{y})$ are the rankings of element t in the lists $\sigma(\mathbf{x})$ and $\sigma(\mathbf{y})$, respectively, and z is defined based on

$$\begin{cases}\sum_{t\in\mathcal{T}}\phi_t = \frac{1}{2}\Phi_K, & \forall t:\sigma_t(\mathbf{x})+\sigma_t(\mathbf{y}) = K+1\\ \sum_{t\in\mathcal{T}}\phi_t = \Phi_K, & |\mathcal{T}| = 2K,\quad \sigma(\mathbf{x})\cap\sigma(\mathbf{y}) = \emptyset\end{cases}. \tag{4}$$

Therefore, the form of z can be written as

$$z = \frac{K - 4\lfloor K/2\rfloor + 2(K+1)\sum_{k=1}^{\lfloor K/2\rfloor} 1/k}{\sum_{k=1}^{K} 1/k} \tag{5}$$

where the operator $\lfloor\cdot\rfloor$ rounds the element to the nearest integer not greater than that element. Next, we prove $z \geq K$, so that with $\sigma_t(\cdot) \leq K$, we can convert ϕ_t in (3) as

$$\phi_t = \begin{cases}\frac{\|\sigma_t(\mathbf{x})-\sigma_t(\mathbf{y})\|_1}{\min\{\sigma_t(\mathbf{x}),\sigma_t(\mathbf{y})\}}, & t\in\sigma(\mathbf{x})\cap\sigma(\mathbf{y})\\ \frac{z}{\sigma_t(\mathbf{x})}-1, & t\in\sigma(\mathbf{x}), t\notin\sigma(\mathbf{y})\\ \frac{z}{\sigma_t(\mathbf{y})}-1, & t\notin\sigma(\mathbf{x}), t\in\sigma(\mathbf{y})\end{cases}. \tag{6}$$

Proof:

$$z - K = \frac{K + 2(K+1)\sum_{k=1}^{\lfloor K/2\rfloor}\frac{1}{k} - 4\lfloor K/2\rfloor - K\sum_{k=1}^{K}\frac{1}{k}}{\sum_{k=1}^{K}\frac{1}{k}} \tag{7}$$

Let the numerator of (7) be denoted as τ. When $K = 1$ we can obtain $\tau = 0$. When $K \geq 2$ and K is an even number, we set $K = 2N$ with N being a positive integer and obtain

$$\begin{aligned}\tau &= 2N + 2(2N+1)\sum_{k=1}^{N}\frac{1}{k} - 4N - 2N\sum_{k=1}^{2N}\frac{1}{k}\\ &= (4N+2)\sum_{k=1}^{N}\frac{1}{k} - 2N\left(\sum_{k=1}^{2N}\frac{1}{k}+1\right)\\ &= (2N+2)\sum_{k=1}^{N}\frac{1}{k} - 2N\left(\sum_{k=N+1}^{2N}\frac{1}{k}+1\right)\\ &= \left(2N+2N\sum_{k=2}^{N}\frac{1}{k}+2\sum_{k=1}^{N}\frac{1}{k}\right) - \left(2N\sum_{k=N+1}^{2N-1}\frac{1}{k}+2N+1\right)\\ &= 2N\sum_{k=2}^{N}\left(\frac{1}{k}-\frac{1}{N+k-1}\right)+2\sum_{k=1}^{N}\frac{1}{k}-1\end{aligned} \tag{8}$$

where $N \geq 1$, and hence, we have $1/k \geq 1/(N+k-1)$ and $2\sum_{k=1}^{N} 1/k > 1$. Therefore, $\tau > 0$ holds. Similarly, when K is an odd number, we set $K = 2N+1$, and can obtain $\tau > 0$ as well. Therefore, for any positive integer K, $z \geq K$ holds.

We can find from the definition of ϕ_t in (6) that ϕ_t is in inverse proportion to its position $\sigma_t(\mathbf{x})$ or $\sigma_t(\mathbf{y})$ in the ranking list even when item $t \notin \sigma(\mathbf{x})\cap\sigma(\mathbf{y})$. That is to say, when the item t is an outlier which is close to the object (i.e., $\mathbf{x}$ or $\mathbf{y}$) such as the third example shown in Fig. 1, the false match (N, N') gets near the front of the neighboring ranking list of $\mathbf{y}$ (e.g., $\sigma_{N'}(\mathbf{y}) = 1$), thus the cost contributed by item N' is dominated in the ranking list distance measurement, which may return a large D_K even when the object match $(\mathbf{x}, \mathbf{y})$ is correct. For instance, in Fig. 1, the ranking list distance D_K calculated by (6) is 0.4595 in the right image pair, while for the false match in the middle situation, the distance calculated by (6) is 0.3452, which leads to a misjudging about these two situations. Therefore, the ranking list similarity measurement D_K is sensitive to the outliers if we directly use the weighted Spearman's footrule measurement in the feature matching task.

In fact, the K-NN ranking list may be constructed in the presence of a large number of mismatches or outliers, which often occurs in feature matching, and the outlier equiprobably locates in any position in a ranking list. To address this issue, we redefine the criterion of calculating ϕ_t as

$$\phi_t = \begin{cases}ll\frac{\|\sigma'_t(\mathbf{x})-\sigma'_t(\mathbf{y})\|_1}{\min\{\sigma'_t(\mathbf{x}),\sigma'_t(\mathbf{y})\}}, & t\in\sigma(\mathbf{x})\cap\sigma(\mathbf{y})\\ \frac{1}{2K}\Phi_K, & \text{otherwise}\end{cases} \tag{9}$$

where $\sigma'_t(\mathbf{x})$ and $\sigma'_t(\mathbf{y})$ are the reranking of element t in new lists $\sigma'(\mathbf{x})$ and $\sigma'(\mathbf{y})$, respectively, only based on the common elements $\{t|t \in \sigma(\mathbf{x})\cap\sigma(\mathbf{y})\}$, e.g., in the right example of Fig. 1, $\sigma'(\mathbf{x}) = (A, B, C)$ and $\sigma'(\mathbf{y}) = (A', B', C')$. In this way, the new criterion for neighboring rank distance measurement is more robust to outliers, and D_K in these three examples of Fig. 1 are 0, 0.3452, and 0.25, respectively, which can distinguish the inlier and outlier more correctly.

B. Problem Formulation

Given a pair of remote sensing images I and I', suppose we have obtained a set of N putative matches $\mathcal{S} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N}$ extracted from the image pair above, where the corresponding feature points $\mathbf{x}_i$ and $\mathbf{y}_i$ are the pixel coordinates in I and I', respectively. $\mathcal{I}$ denote the unknown inlier set. To preserve the local structure of feature points, the optimal solution is

$$\mathcal{I}^* = \arg\min_{\mathcal{I}} C(\mathcal{I}; \mathcal{S}, \lambda) \tag{10}$$

with the cost function C defined as

$$C(\mathcal{I}; \mathcal{S}, \lambda) = \sum_{i\in\mathcal{I}} D_K(\sigma(\mathbf{x}_i), \sigma(\mathbf{y}_i)) + \lambda(N - |\mathcal{I}|) \tag{11}$$

where $\sigma(\mathbf{x}_i)$ and $\sigma(\mathbf{y}_i)$ denote the top K ranking lists of $\mathbf{x}_i$ and $\mathbf{y}_i$, respectively, $D_K(\sigma(\mathbf{x}_i), \sigma(\mathbf{y}_i))$ measures the difference of the top K ranking lists between $\mathbf{x}_i$ and $\mathbf{y}_i$, and $|\cdot|$ denotes the cardinality of a set. In this cost function, the first

term penalizes any match not preserving the local neighborhoods' ranking similarity, the second term is used to discourage the outliers, and the positive parameter λ controls the tradeoff between the two terms. Ideally, the optimal solution should achieve zero penalty, i.e., the first term of C should be zero. That is to say, it tries to obtain the maximum inlier number and keep the cost value to minimum.

We introduce an $N \times 1$ binary vector $\mathbf{p}$ to associate the putative match set, where $p_i = \{0, 1\}$ indicates whether the ith putative match is correct or not. Therefore, the cost function in (11) can be written as

$$C(\mathbf{p}; \mathcal{S}, \lambda) = \sum_{i=1}^{N} p_i D_K(\sigma(\mathbf{x}_i), \sigma(\mathbf{y}_i)) + \lambda \left(N - \sum_{i=1}^{N} p_i \right). \tag{12}$$

1) Multiscale Cost: In the formulation above, we have constructed the K-NN for each feature point to obtain its top K ranking list, and measured the ranking similarity of each putative match with D_K in (1). However, the optimal value of K may change with respect to different image data, due to that the distribution and the proportion of outliers in the putative set are vary with image domain. In this case, it is not suitable to address the general feature matching task by fixing the value of K. To address this issue, we develop a multiscale strategy, which measures the local K neighborhoods' ranking similarity and calculates the cost function under different scales of K. Therefore, we define a set of different K with $\mathbf{K} = \{K_m\}_{m=1}^{M}$, and the ranking list similarity of ith putative match with respect to K_m nearest neighborhoods can be denoted as $D_{K_m}(\sigma(\mathbf{x}_i), \sigma(\mathbf{y}_i))$. Accordingly, the multiscale cost function about (12) can be written as

$$C(\mathbf{p}\mathcal{S}, \lambda) = \sum_{i=1}^{N} \frac{p_i}{M} \sum_{m=1}^{M} D_{K_m}(\sigma(\mathbf{x}_i), \sigma(\mathbf{y}_i)) + \lambda \left(N - \sum_{i=1}^{N} p_i \right). \tag{13}$$

We reorganize its form by merging the terms related to p_i and obtain

$$C(\mathbf{p}; \mathcal{S}, \lambda) = \sum_{i=1}^{N} p_i (c_i - \lambda) + \lambda N \tag{14}$$

where

$$c_i = \frac{1}{M} \sum_{m=1}^{M} D_{K_m}(\sigma(\mathbf{x}_i), \sigma(\mathbf{y}_i)). \tag{15}$$

C. Solution

Given a putative match set, once the K-NNs of all feature points are constructed, all the cost values of $\{c_i\}_{i=1}^{N}$ can be calculated beforehand. Therefore, to determine the value of p_i in (14), we can easily observe that any putative match with a cost smaller than the parameter λ will lead to a negative term and decreases the objective cost function, thus we prefer to set the value of p_i to 1, and vice versa. That is to say, the optimal solution of $\mathbf{p}$ that minimizes the cost function in (14) can be determined simply by the following criterion:

$$p_i = \begin{cases} 1, & c_i \le \lambda \\ 0, & c_i > \lambda. \end{cases} \tag{16}$$

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

$$\mathcal{I}^* = \{i | p_i = 1, \quad i = 1, 2, \ldots, N\}. \tag{17}$$

Obviously, we can see from (16) that the parameter λ is used for judging the correctness of putative matches. However, the ranking list $\{(\sigma(\mathbf{x}_i), \sigma(\mathbf{y}_i))\}_{i=1}^{N}$ of local neighborhoods is constructed based on the whole putative set, usually including a large number of outliers, which may cause the incorrect measurement of the ranking list similarity, and optimally, it will be more desirable if the ranking list is constructed based on only the inlier set $\mathcal{I}$, so that we can separate the inliers and outliers based on c_i and parameter λ correctly. However, the inlier set is what we need to solve in our matching task and cannot be known in advance. To address this dilemma, we utilize an iterative strategy to determine the optimal $\mathcal{I}$ by seeking an approximation $\mathcal{I}_{\text{iter}}$ in each iteration and use it for neighborhood construction.

To verify how well the new measuring criterion and the iterative strategy work, we randomly select in total 30 remote sensing image pairs with different types of transformations involving rigid, rotation, scale change, nonrigid deformation such as FE images, and so on. The average initial inlier number and inlier percentage of putative matches obtained by SIFT on the whole test data are 1148.2% and 56.28%, respectively. In addition, the F-score is used for evaluating the matching performance, which is defined as F-score=2 $\times$ Precision $\times$ Recall/(Precision+Recall), where the Precision is defined as the ratio of the identified correct match number and the preserved match number, and the Recall is defined as the ratio of identified correct match number and the correct match number contained in the putative set. The F-score curves with respect to different values of parameter λ are summarized in Fig. 2, where we have reported the results of LPM [16], TopKRP [17], and mTopKRP. Clearly, our mTopKRP greatly promotes the matching performance. With a proper value of λ (e.g., 0.8), we obtain the best average F-score (AF) 96.48%, where the average precision (AP) and recall are about 94.76% and 98.41%, respectively.

Even the K-NN ranking lists are constructed based on the whole putative sets, which involve quite a number of false matches, the top K rank preservation strategy can work well and generates a correspondence set, which can filter out most of the outliers and simultaneously keep most of the inliers. Thus, the generated correspondence set can be a good approximation of the true inlier set, and we denote it as $\mathcal{I}_1$. To further improve the performance, we subsequently use $\mathcal{I}_1$ to construct the K-NN ranking list for each putative match in $\mathcal{S}$, and solve the optimal $\mathcal{I}^*$ in an iterative manner, i.e., the jth inlier set $\mathcal{I}_j$ can be generated based on the ranking list constructed with the $(j-1)$th inlier set $\mathcal{I}_{j-1}$, which can be written as

$$\mathcal{I}_j = \arg\min_{\mathcal{I}} C(\mathcal{I}; \mathcal{I}_{j-1}, \mathcal{S}, \lambda) \tag{18}$$

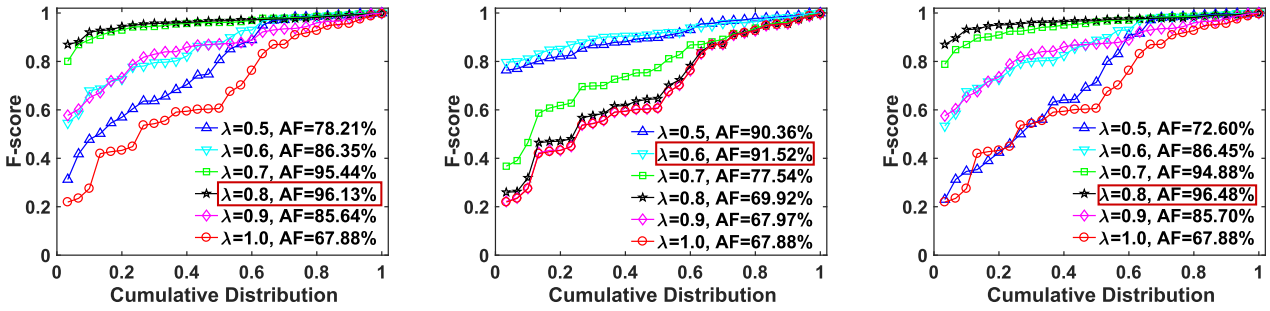


Fig. 2. F-score with respect to the cumulative distribution by using the whole feature set to construct local neighboring structure on 30 remote-sensing image pairs. The average inlier number and average inlier rate of the putative matches are 1148.2% and 56.28%. (Left) Results of LPM [16], where we normalize the cost by dividing it with $2K$, e.g., $K = 15$. (Middle) Results of TopKRP [17] with $K = 15$. (Right) Results of mTopKRP using a new similarity measuring criterion and multiscale neighborhood representation, e.g., $\mathbf{K} = [13, 15, 17]$. A point on the curve with coordinate (x, y) denotes that there are $(100 * x)\%$ percent of image pairs, which have F-score not more than y. The best average F-score and its threshold λ are marked with red box in the legend.

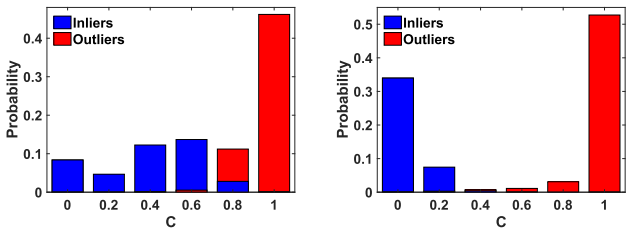


Fig. 3. Distributions of the cost c_i in (15) by using the whole feature set $\mathcal{I}_0 = \mathcal{S}$ (left) and by using $\mathcal{I}_1$ (right) to construct neighboring structure.

where $j = 1, 2, \ldots, MaxIter$, $\mathcal{I}_0 = \mathcal{S}$, and the optimal inlier set $\mathcal{I}^*$ can be approximated as $\mathcal{I}_{\text{MaxIter}}$. Fig. 3 reports the distributions of the cost c_i in (15) by using $\mathcal{S}$ and $\mathcal{I}_1$ to construct neighboring structure. Clearly, we can see from the results that the margin between inlier and outlier has been distinctly enlarged by using the refined set $\mathcal{I}_1$ for neighborhood construction. With the parameter $\lambda = 0.35$, the AP, recall, and F-score on the 30 testing image pairs can be largely increased from $(94.76\%, 98.41\%, 96.48\%)$ to $(98.42\%, 99.27\%, 98.83\%)$. As the iteration proceeds, the AP, recall, and F-score can be further increased to $(98.70\%, 99.42\%, 99.05\%)$ after convergence. In this paper, considering the time complexity, we set $MaxIter = 3$, which works well and can generate satisfying matching performance on remote sensing images.

Since our method aims to preserve the similarity of top K ranking lists under the multiscale KNN, we name it as mTopKRP and summarize the whole procedure in Algorithm 1.

D. Computational Complexity

To construct the ranking lists $\{\sigma(\mathbf{x}_i), \sigma(\mathbf{y}_i)\}_{i=1}^N$ based on K-NN for all the feature points in the putative set, the time complexity scales like $O((K + N)\log N)$ by using K-D tree [46]. Thus, the time cost of Line 3 in Algorithm 1 is close to $O((K_M + N)\log N)$ for the reason that the small scale K-NN can be directly obtained from the larger one. According to (6), the least and most time costs of calculating $\{c_i\}_{i=1}^N$ in Line 4 are $O(\sum_{m=1}^M K_m N)$ and $O(\sum_{m=1}^M 2K_m N)$, respectively. Moreover, determining $\mathbf{p}$ and inlier set $\mathcal{I}$ in Line 6 just costs $O(N)$ time. Therefore, the total time complexity

Algorithm 1: mTopKRP Algorithm

Input: Putative set $\mathcal{S} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$, params. $\mathbf{K}$, λ, $MaxIter$

Output: Inlier set $\mathcal{I}^*$

1 Initialize $j = 0$, $\mathcal{I}_0 = S$;
2 **Iteration:**
3 Construct ranking lists based on $\mathcal{I}_j$ with multiscale K-NN $\mathbf{K}$;
4 Calculate cost $\{c_i\}_{i=1}^N$ according to (1) and (15);
5 $j = j + 1$;
6 Determine $\mathcal{I}_j$ using (16) and (17);
7 **Until:** $j \geq MaxIter$
8 $\mathcal{I}^* = \mathcal{I}_j$.

TABLE I
AF AND AVERAGE RUN TIME (ART, UNIT: MS) OF OUR MTOPKRP UNDER DIFFERENT SETTINGS OF PARAMETER $\mathbf{K}$ ON THE 30 REMOTE-SENSING IMAGE PAIRS IN FIG. 2

$\mathbf{K}$	[3 5 7]	[8 10 12]	[13 15 17]	[18 20 22]	[23 25 27]
AF	0.9040	0.9571	0.9648	0.9674	0.9678
ART	57.5	58.6	59.7	64.4	66.3

of our mTopKRP in one iteration is less than $O((K_M + N)\log N + (1 + 2\sum_{m=1}^M K_m)N)$. The space complexity of our method is $O(K_M N)$ due to the memory requirement for storing the top K ranking list. Generally, $\sum_{m=1}^M K_m \ll N$ and our algorithm can converge in just a few iterations. That is to say, the time and space complexity of our mTopKRP can be simply written as $O(N \log N)$ and $O(N)$, respectively. This is significant for addressing real-time or large-scale tasks, especially in the case of matching high-resolution remote sensing images.

E. Implementation Details

There are three parameters in our method: $\mathbf{K}$, λ, and $MaxIter$. Parameter $\mathbf{K}$ determines the number of nearest neighborhoods for multiscale K-NN ranking lists construction. To seek the optimal value of $\mathbf{K}$, we test different settings on the 30 remote sensing image pairs shown in Fig. 2, and report the AFs and run time in Table I. From the results, we see that

the F-score increases as $\mathbf{K}$ increases, while the run time also increases at the same time. To achieve a tradeoff between the accuracy and efficiency, we set $\mathbf{K} = [13, 15, 17]$ as its default setting. Parameter λ controls the threshold for judging the correctness of putative matches. Parameter $MaxIter$ indicates the max iteration in our method. Obviously, a small value of λ will increase the precision and simultaneously decrease the recall and vice versa. A large value of $MaxIter$ will improve the performance slightly but requires more time cost. In our evaluation, based on the experiments and analysis aforementioned, we set the default values of these parameters as $MaxIter = 3$, $\mathbf{K} = [13, 15, 17]$, and $\lambda = 0.8, 0.35, 0.35$ in the three iterations, respectively.

IV. Experimental Results

In this section, we test the performance of our proposed method on different kinds of remote sensing data sets and compare it with other feature matching methods. In particular, both rigid and nonrigid remote sensing image data sets are selected for evaluation, which are of different imaging scenarios. Six state-of-the-art methods are chosen for comparison, namely RANSAC [14], ICF [24], GS [33], LLT [1], and LPM [16]. The parameters are set according to the original papers and fixed throughout our experiments. For LLT, we select the adaptive model for each data set. The open source VLFeat toolbox [47] is employed for SIFT detector and descriptor as well as K-NN searching with K-D tree. The experiments are conducted on a desktop with 3.4-GHz Intel Core CPU, 8-GB memory, and MATLAB R2016b code.

A. Data Sets

To evaluate the performance of our method, we use five remote sensing image data sets as follows.

1) UAV: The data set consists of 35 pairs of color images, which are of resolution 600×337 and captured by a UAV over a piece of farmland. Accurate feature matching of these images is particularly required in the field of automatic crop monitoring. Generally, the images suffer from projective distortions due to the unstable imaging condition.

2) SAR: The data set contains 34 image pairs corrupted with strong noise. For each image pair, the two images are separately obtained by synthetic-aperture radars (SARs) on a satellite and on an UAV, respectively. Feature matching for such image pairs is critical in the positioning and navigating problem, where the UAV images are required to match the corresponding stored satellite images to accurately estimate the current position. The image pairs can be modeled with similarity or rigid transformation in most cases.

3) PAN: The data set consists of 31 pairs of panchromatic (PAN) aerial photographs captured by a frame camera at different times. Viewpoint changes often exist in these image pairs, resulting in affine or projective distortions. Feature matching for these images typically arises in change detection. The images are of two sizes of 561×518 and 600×700.

4) CIAP: The data set consists of 40 pairs of color infrared aerial photographs (CIAP), with the size of 700×700. We note that these images are already orthorectified, and hence, the transformation model is just rigid. However, the overlap areas are quite small. Feature matching for these images is important in the image mosaic problem.

5) FE: The data set consists of 30 pairs of images captured from four scenes with an FE camera [48].[2] Most of the image pairs suffer from viewpoint changes, and severe nonrigid deformations are also involved. These images are used for nonparametric image matching evaluation.

To establish the ground truth, i.e., determine the true correspondence set, we have made a benchmark before conducting any experiments, to ensure objectivity; specifically, each putative correspondence in each image pair is checked manually.

B. Qualitative Results

To demonstrate the effectiveness of our proposed feature matching method, we first give the results on some typical image pairs in Fig. 4. From top to bottom, the five rows represent the aforementioned five data sets, namely, UAV, SAR, PAN, CIAP, and FE, and each contains two examples. For each example, the left plot presents the intuitive result on the image pair, and the right plot presents the motion field of the correspondences. For visibility, only 100 randomly selected correspondences of true positive, false negative, and false positive are shown in the left plot for visuality. The right motion field presents the result for all correspondences, denoted by the arrows.

The selected image pairs are challenging for the feature matching task for various reasons. Specifically, the first and the third rows suffer from projective distortions, the second row suffers from severe noise, the fourth row suffers from small overlap areas, and the last row suffers from nonrigid deformations. As aforementioned, the SIFT algorithm is adopted to extract features and a large number of correspondences are constructed together with many outliers. The number of initial correspondences is 574, 852, 539, 1084, 504, 628, 305, 388, 503, and 644 for the ten image pairs, with the inlier rate being 42.49%, 41.40%, 36.05%, 42.99%, 26.88%, 27.26%, 13.28%, 15.76%, 50.30%, and 15.45%, respectively. By using our proposed method to filter out the mismatches, we can obtain the precision, recall and F-score statistics being (99.65%, 99.30%, 0.9948), (99.81%, 99.63%, 0.9972), (99.36%, 99.82%, 0.9959), (99.18%, 99.65%, 0.9941), (99.60%, 100.0%, 0.9980), (100.0%, 99.84%, 0.9992), (100.0%, 100.0%, 1.000), (100.0%, 100.0%, 1.000), (97.24%, 98.21%, 0.9773), and (98.77%, 99.53%, 0.9915), respectively. Clearly, our method successfully identifies most of the true correspondences, and only a few are wrongly classified. These results prove the generality and robustness of our method, which can handle different scenarios even in the presence of a large percentage of outliers.

C. Quantitative Results

To test the performance of feature matching methods in different scenarios, we divide the five data sets into three groups. SAR and CIAP are classified as a rigid data set,

[2] http://www.ti.uni-bielefeld.de/html/people/ahoffmann/outdoordb.html

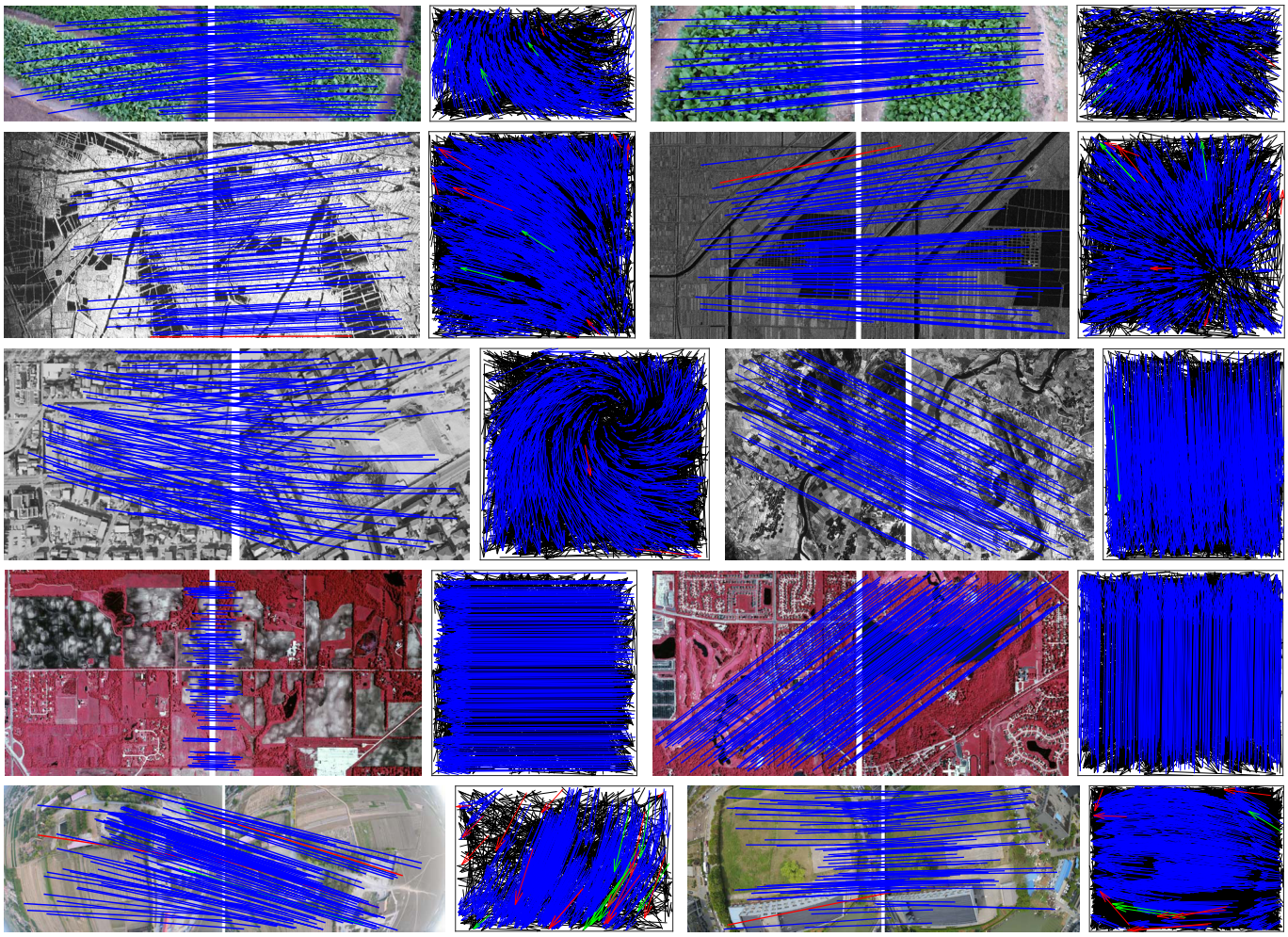


Fig. 4. Feature matching results of our mTopKRP on ten representative remote sensing image pairs. (Top to bottom and left to right) UAV1, UAV2, SAR1, SAR2, PAN1, PAN2, CIAP1, CIAP2, FE1, and FE2. The ratio of inliers in the ten image pairs is 42.49%, 41.40%, 36.05%, 42.99%, 26.88%, 27.26%, 13.28%, 15.76%, 50.30%, and 15.45%, respectively. The head and tail of each arrow in the motion field correspond to the positions of feature points in the two images (blue blue = true positive, black = true negative, green green = false negative, red red = false positive). For visibility, in the image pairs, at most 100 randomly selected matches are presented, and the true negatives are not shown. Best viewed in color.

UAV and PAN are classified as projective data set, and FE for nonrigid data set. The average putative match number of these three data sets is 939.43, 888.94, and 463.00, respectively. The quantitative experiments are conducted on these three groups, and the results are reported separately. The cumulative distribution of initial inlier ratios on the three data sets is provided in the first row of Fig. 5. We see that for the rigid data set, the inlier ratio is generally high, with a few image pairs being challenging because of low inlier ratio. For the projective data set, the inlier ratio distribution is polarized, the easy ones and the challenging ones are balanced. For the nonrigid data set, the image pairs maintain a relatively high inlier ratio, however, this data set is still challenging because of its nonrigid nature.

The statistic results on the three data sets, such as precision, recall, F-score, and run time, are summarized in Fig. 4. From left to right, each column presents the results of a rigid data set, projective data set, and nonrigid data set, respectively. For the rigid data set, we see that all methods obtain good results. This is due to the simple transformation model between the image pairs. ICF has relatively low precision, as it is designed for nonrigid feature matching, and the spatial constraint is relaxed. Although GS and RANSAC achieve great accuracy, our method shows its ability to preserve more correct matches, leading to better recalls. LLT also has comparative results, however, it needs a prior for the data, where the model has to be manually set. The results also reveal that our method has significant improvements over LPM. The recalls are close for the two methods, however, with our improved interpretation of local structure, our method obtains much better precisions than LPM. The performances are clearly characterized with the summary statistic, e.g., F-score. We can observe that our method is the best, with obvious advantages. For the projective data set, the situation has a slight change. ICF shows its weakness to preserve correct matches, resulting in the poor recall statistic. Except for ICF, all methods show great performance. Our method remains to be the best due to the robustness to outliers. In terms of both precision and recall, our method consistently obtains excellent results, while the other methods are less robust and fail in some cases. For the nonrigid

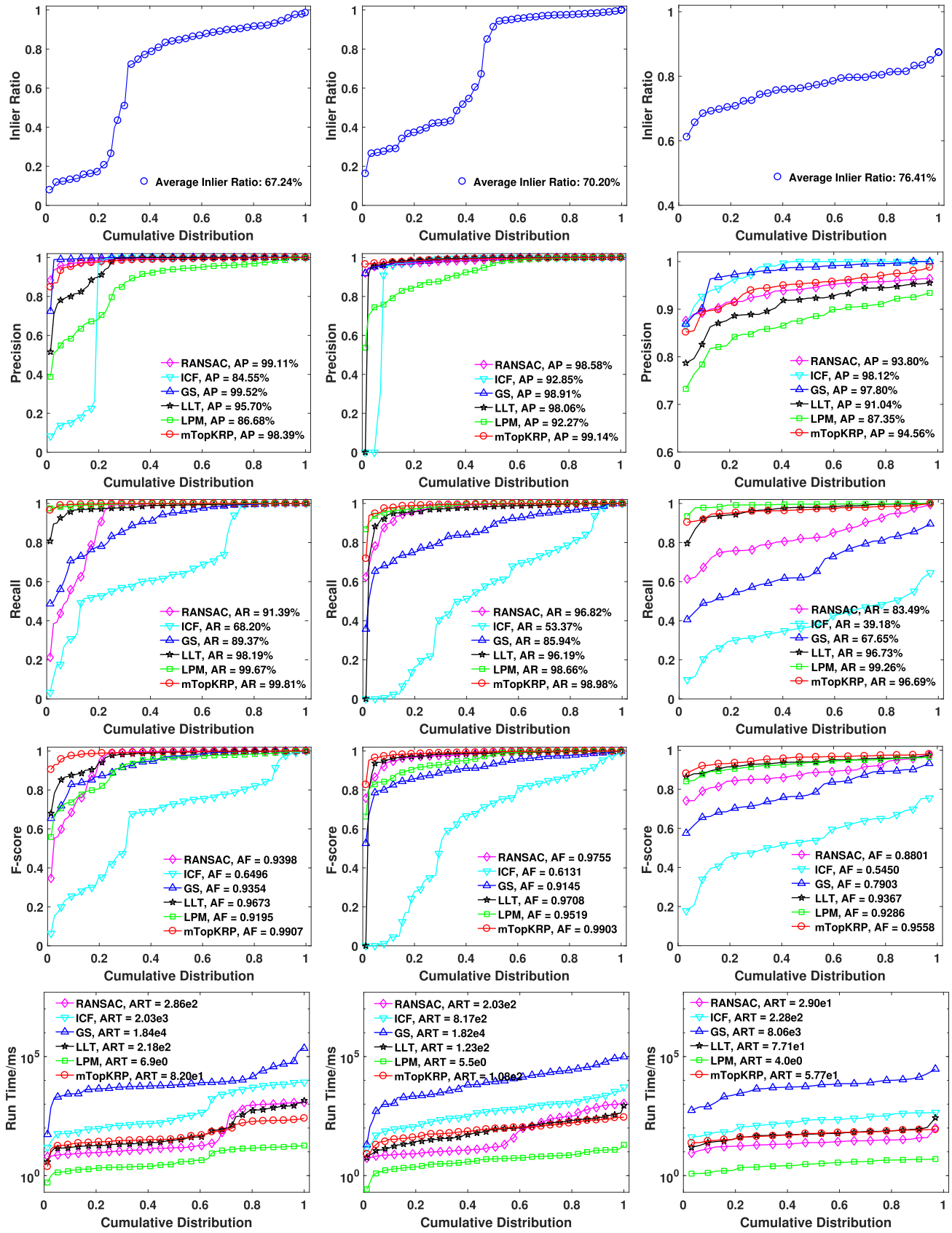


Fig. 5. Quantitative comparisons of RANSAC [14], ICF [24], GS [33], LLT [1], LPM [16], and our mTopKRP on five image sets which are divided into three groups according to their transformation models. (Left to right) Rigid (SAR, CIAP), projective (UAV, PAN), and nonrigid (FE). (Top to bottom) Initial inlier ratio, precision, recall, F-score, and run time with respect to the cumulative distribution. A point on the curve with coordinate (x, y) denotes that there are $(100 * x)\%$ percent of image pairs, which have the performance value (i.e., inlier ratio, precision, recall, F-score, or run time) not more than y. The AP, average recall (AR), AF, and ART are reported in the legend.

data set, we can observe that LPM, LLT, and mTopKRP have the best performance. This is due to that GS and RANSAC are known to be sensitive to nonrigid deformation, and ICF always has low recalls. Compared to LPM and LLT, our method clearly obtains the best performance, which demonstrates its effectiveness. The time costs of different methods are also given in Fig. 5. Our method has a relatively low complexity, as illustrated by the last half of the cumulative distribution, which is only inferior to LPM. This result meets our expectation, as our method is based on the efficient LPM method. In summary, our method is privileged in terms of efficiency.

V. Conclusion

In this paper, we have proposed a simple yet efficient mismatch removal method for robust feature matching of remote sensing images named mTopKRP, which is based on the stable neighboring topological relationship of feature correspondences between the two images of the same object or scene. In mTopKRP, we use the K-NN ranking lists to denote the topology of feature points to be matched. The weighted Spearman footrule distance is then redefined and improved to measure the similarity between two top K ranking lists of an image pair. Meanwhile, we formulate the feature matching task as an iterative optimization problem and introduce a closed-form solution, which can establish the reliable correspondences in linearithmic time complexity. The qualitative and quantitative results on various remote sensing images have demonstrated the generality, robustness, and the superior performance of our method for handling various matching tasks in remote sensing and photogrammetry over the state of the art. Moreover, its low time cost, only requiring tens of milliseconds to handle over 1000 putative matches, is significant for addressing real-time or large-scale tasks.

References

[1] 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.

[2] J. Ma, W. Yu, P. Liang, C. Li, and J. Jiang, "FusionGAN: A generative adversarial network for infrared and visible image fusion," *Inf. Fusion*, vol. 48, pp. 11–26, Aug. 2019.

[3] J. Ma, Y. Ma, and C. Li, "Infrared and visible image fusion methods and applications: A survey," *Inf. Fusion*, vol. 45, pp. 153–178, Jan. 2019.

[4] R. Xie, M. Xia, J. Yao, and L. Li, "Guided color consistency optimization for image mosaicking," *ISPRS J. Photogram. Remote Sens.*, vol. 135, pp. 43–59, Jan. 2018.

[5] J. Li, Q. Hu, M. Ai, and R. Zhong, "Robust feature matching via support-line voting and affine-invariant ratios," *ISPRS J. Photogramm. Remote Sens.*, vol. 132, pp. 61–76, Oct. 2017.

[6] Y. Furukawa and J. Ponce, "Accurate, dense, and Robust multiview stereopsis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 8, pp. 1362–1376, Aug. 2010.

[7] 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.

[8] S. Nebiker, N. Lack, and M. Deuber, "Building change detection from historical aerial photographs using dense image matching and object-based image analysis," *Remote Sens.*, vol. 6, no. 9, pp. 8310–8336, 2014.

[9] G. Cheng and J. Han, "A survey on object detection in optical remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 117, pp. 11–28, Jul. 2016.

[10] C. Wang, L. Wang, and L. Liu, "Progressive mode-seeking on graphs for sparse feature matching," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2014, pp. 788–802.

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

[12] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up Robust features (SURF)," *Comput. Vis. Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[13] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2564–2571.

[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] G. R. Rodríguez-Canosa, S. Thomas, J. D. Cerro, A. Barrientos, and B. MacDonald, "A real-time method to detect and track moving objects (DATMO) from unmanned aerial vehicles (UAVS) using a single camera," *Remote Sens.*, vol. 4, no. 4, pp. 1090–1111, 2012.

[16] J. Ma, J. Zhao, H. Guo, J. Jiang, H. Zhou, and X. Gao, "Locality preserving matching," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 4492–4498.

[17] J. Jiang, Q. Ma, T. Lu, Z. Wang, and J. Ma, "Feature matching based on top k rank similarity," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2018, pp. 2316–2320.

[18] Q. Li, G. Wang, J. Liu, and S. Chen, "Robust scale-invariant feature matching for remote sensing image registration," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 2, pp. 287–291, Apr. 2009.

[19] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proc. Eur. Conf. Comput. Vis.*, May 2006, pp. 430–443.

[20] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2010, pp. 778–792.

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

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

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

[25] J. Ma, W. Qiu, J. Zhao, Y. Ma, A. L. Yuille, and Z. Tu, "Robust L_2E estimation of transformation for non-rigid registration," *IEEE Trans. Signal Process.*, vol. 63, no. 5, pp. 1115–1129, Mar. 2015.

[26] P. J. Besl and N. D. McKay, "Method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992.

[27] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2262–2275, Dec. 2010.

[28] B. Jian and B. C. Vemuri, "Robust point set registration using Gaussian mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1633–1645, Aug. 2011.

[29] J. Yan, C. Li, Y. Li, and G. Cao, "Adaptive discrete hypergraph matching," *IEEE Trans. Cybern.*, vol. 48, no. 2, pp. 765–779, Feb. 2018.

[30] J. Yan, M. Cho, H. Zha, X. Yang, and S. M. Chu, "Multi-graph matching via affinity optimization with graduated consistency regularization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 6, pp. 1228–1242, Jun. 2016.

[31] M. Leordeanu and M. Hebert, "A spectral technique for correspondence problems using pairwise constraints," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2005, pp. 1482–1489.

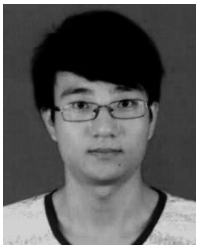
[32] T. Cour, P. Srinivasan, and J. Shi, "Balanced graph matching," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 313–320.

[33] 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.

[34] F. Song, M. Li, Y. Yang, K. Yang, X. Gao, and T. Dan, "Small UAV based multi-viewpoint image registration for monitoring cultivated land changes in mountainous terrain," *Int. J. Remote Sens.*, vol. 39, no. 21, pp. 7201–7224, 2018.

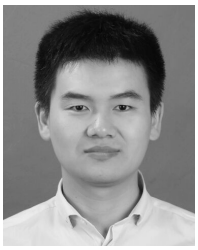
[35] Z. Wei *et al.*, "A small UAV based multi-temporal image registration for dynamic agricultural terrace monitoring," *Remote Sens.*, vol. 9, no. 9, p. 904, 2017.

[36] S. Zhang, K. Yang, Y. Yang, and Y. Luo, "Nonrigid image registration for low-altitude SUAV images with large viewpoint changes," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 4, pp. 592–596, Apr. 2018.
[37] K. Yang, A. Pan, Y. Yang, S. Zhang, S. H. Ong, and H. Tang, "Remote sensing image registration using multiple image features," *Remote Sens.*, vol. 9, no. 6, p. 581, 2017.
[38] G.-J. Wen, J.-J. Lv, and W.-X. Yu, "A high-performance feature-matching method for image registration by combining spatial and similarity information," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 4, pp. 1266–1277, Apr. 2008.
[39] Z. Liu, J. An, and Y. Jing, "A simple and Robust feature point matching algorithm based on restricted spatial order constraints for aerial image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 514–527, Feb. 2012.
[40] H. Zhou, J. Ma, C. Yang, S. Sun, R. Liu, and J. Zhao, "Nonrigid feature matching for remote sensing images via probabilistic inference with global and local regularizations," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 374–378, Mar. 2016.
[41] 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. doi: 10.1007/s11263-018-1117-z.
[42] 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. doi: 10.1109/TGRS.2018.2820040.
[43] 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.
[44] 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.
[45] R. Fagin, R. Kumar, and D. Sivakumar, "Comparing top k lists," *SIAM J. Discrete Math.*, vol. 17, no. 1, pp. 134–160, 2003.
[46] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Commun. ACM*, vol. 18, no. 9, pp. 509–517, 1975.
[47] A. Vedaldi and B. Fulkerson, "Vlfeat: An open and portable library of computer vision algorithms," in *Proc. 18th ACM Int. Conf. Multimedia*, Oct. 2010, pp. 1469–1472.
[48] M. Horst and R. Möller, "Visual place recognition for autonomous mobile robots," *Robotics*, vol. 6, no. 2, p. 9, 2017.



Xingyu Jiang received the B.E. degree from the Department of Mechanical and Electronic Engineering, Huazhong Agricultural University, Wuhan, China, in 2017. He is currently pursuing the Ph.D. degree with the Electronic Information School, Wuhan University, Wuhan, China.

His research interests include computer vision, machine learning, and pattern recognition.



Junjun Jiang received the B.S. degree from the Department of Mathematics, Huaqiao University, Quanzhou, China, in 2009, and the Ph.D. degree from the School of Computer, Wuhan University, Wuhan, China, in 2014.

From 2015 to 2018, he was an Associate Professor with the China University of Geosciences, Wuhan. Since 2016, he has been a Project Researcher with the National Institute of Informatics, Tokyo, Japan. He is currently a Professor with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China. His research interests include image processing and computer vision.

Dr. Jiang was a recipient of the Finalist of the World's First 10K Best Paper Award at ICME 2017, the Best Student Paper Runner-up Award at MMM 2015, the 2016 China Computer Federation (CCF) Outstanding Doctoral Dissertation Award, and the 2015 ACM Wuhan Doctoral Dissertation Award.

Aoxiang Fan received the B.S. degree from the Electronic Information School, Wuhan University, Wuhan, China, in 2018, where he is currently pursuing the master's degree with the Multi-Spectral Vision Processing Lab.

His research interests include computer vision and pattern recognition.

Zhongyuan Wang received the Ph.D. degree in communication and information system from Wuhan University, Wuhan, China, in 2008.

He is currently a Professor with the School of Computer, Wuhan University. He is also directing three projects funded by the National Natural Science Foundation Program of China.

Jiayi Ma 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. From 2014 to 2015, he was a Post-Doctoral Researcher with the Electronic Information School, Wuhan University, Wuhan, where he is currently an Associate Professor. He has authored or co-authored more than 100 refereed journal and conference papers, including IEEE TPAMI/TIP/TSP/TNNLS/TGRS/TCYB/TMM/TCSVT, IJCV, CVPR, IJCAI, AAAI, ICRA, IROS, and ACM MM. His research interests include computer vision, machine learning, and pattern recognition.

Dr. Ma was a recipient of the Natural Science Award of Hubei Province (first class), the CAAI (Chinese Association for Artificial Intelligence) Excellent Doctoral Dissertation Award (a total of eight winners in China), and the CAA (Chinese Association of Automation) Excellent Doctoral Dissertation Award (a total of ten winners in China). He has been an Associate Editor of *Neurocomputing* since 2018 and IEEE ACCESS since 2017. He is a Guest Editor of *Remote Sensing*.