Robust Feature Matching for Remote Sensing Image Registration via Linear Adaptive Filtering

Xingyu Jiang, *Member, IEEE*, Jiayi Ma, *Member, IEEE*, Aoxiang Fan, Haiping Xu, Geng Lin, Tao Lu, *Member, IEEE*, and Xin Tian, *Member, IEEE*

***Abstract*— As a fundamental and critical task in feature-based remote sensing image registration, feature matching refers to establishing reliable point correspondences from two images of the same scene. In this article, we propose a simple yet efficient method termed linear adaptive filtering (LAF) for both rigid and nonrigid feature matching of remote sensing images and apply it to the image registration task. Our algorithm starts with establishing putative feature correspondences based on local descriptors and then focuses on removing outliers using geometrical consistency priori together with filtering and denoising theory. Specifically, we first grid the correspondence space into several nonoverlapping cells and calculate a typical motion vector for each one. Subsequently, we remove false matches by checking the consistency between each putative match and the typical motion vector in the corresponding cell, which is achieved by a Gaussian kernel convolution operation. By refining the typical motion vector in an iterative manner, we further introduce a progressive strategy based on the coarse-to-fine theory to promote the matching accuracy gradually. In addition, an adaptive parameter setting strategy and posterior probability estimation based on the expectation–maximization algorithm enhance the robustness of our method to different data. Most importantly, our method is quite efficient where the gridding strategy enables it to achieve linear time complexity. Consequently, some sparse point-based tasks may inspire from our method when they are achieved by deep learning techniques. Extensive feature matching and image registration experiments on several remote sensing data sets demonstrate the superiority of our approach over the state of the art.**

***Index Terms*— Adaptive, convolution, feature matching, filtering, outlier, registration.**

I. Introduction

IMAGE registration, which aims to geometrically warp the sensed image into the spatial coordinate system of the reference image and align their common area pixel-to-pixel, is a fundamental and challenging problem in remote sensing and photography community [1]. The images to be registered are usually taken from the same scene and captured at multitemporal, from multiviewpoints or multimodalities. Many remote sensing tasks, such as image mosaic, image fusion, change detection, and map updating, are performed on well-registered images, leading to an urgent requirement for efficient and robust registration methods [2]–[7].

During the last decades, a growing amount and diversity of methods are proposed for remote sensing image registration, particularly when the deep learning techniques are widely used in recent years. These methods can be roughly classified into three categories, saying area-, feature-, and learning-based methods [1], [8]. Area-based methods register two images by using the similarity measurement of the original pixel intensity or information after domain transforming. This is implemented using sliding windows of predefined size or even entire images, without attempting to detect any salient objects. Feature-based methods start with detecting the sparse and salient features from two images and then establish reliable correspondences under similarity of local image descriptors and/or spacial geometrical constraints. As for learning-based methods, due to the strong ability in deep feature acquisition and nonlinear expression, applying deep learning techniques for image information representation, similarity measurement, and parameters regression has received considerable attention recently [8], [9].

According to the basic idea, area-based methods can achieve better performance when the images have few prominent details, where the distinctive information is provided by pixel intensities rather than local shapes and structures. However, they badly suffer from the high computational complexity, image distortion, and intensity changes. These handicaps are typically introduced, for instance, by noise, varying illumination, and imaging from different sensors. By contrast, since the features, such as points, lines, and salient regions [10]–[12], can be seen as a simplistic representation of an image, feature-based methods are generally more efficient and robust to complex image distortions. Therefore, salient features have been widely used in remote sensing tasks. In addition, points can be regarded as the basic form of other features, and hence, they are more general and easy to extract and define. The learning-based methods are not so well studied, particularly for sparse feature matching. The existing methods can merely handle the large overlapped image pairs within slight rotation, scaling, and nonrigid deformation. However, they have shown great potential in the image registration task. In this article,

Manuscript received September 23, 2019; revised April 2, 2020 and May 8, 2020; accepted June 6, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61773295 and Grant 61971315 and in part by the Natural Science Fund of Hubei Province under Grant 2019CFA037. *(Corresponding authors: Jiayi Ma; Xin Tian.)*

Xingyu Jiang, Jiayi Ma, Aoxiang Fan, and Xin Tian are with Electronic Information School, Wuhan University, Wuhan 430072, China (e-mail: jiangx.y@whu.edu.cn; jyma2010@gmail.com; fanaoxiang@whu.edu.cn; xin.tian@whu.edu.cn).

Haiping Xu and Geng Lin are with the College of Mathematics and Data Science, Minjiang University, Fuzhou 350108, China (e-mail: haiping@mju.edu.cn; lingeng413@163.com).

Tao Lu is with the School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan 430073, China (e-mail: lutxyl@gmail.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.2020.3001089

we mainly focus on point feature-based methods for remote sensing image registration. Specifically, we first establish accurate feature point correspondences and then estimate a predefined transformation between two images accordingly. Subsequently, the sensed image is aligned with the reference image by using an appropriate interpolation method [1].

Feature-based methods typically desire a robust and efficient matching strategy to establish correct correspondences between two feature point sets. This stimulates various methods for better performance in efficiency and accuracy in the past decades. Nevertheless, there are still several challenges to develop a general and efficient matching technique for remote sensing image registration. First, an efficient technique is in urgent need in large-scale remote sensing tasks, for the reason that matching N points to another N points would create the computational cost of $O(N!)$ due to its combinatorial nature [13], [14]. Very often, thousands of feature points are extracted and to be corresponded for large or high-resolution remote sensing images, leading to a significant burden on the existing matching methods. Second, a more complex nonrigid transformation modeled in high dimension is required for an accurate alignment. This is because some inevitable local distortions caused by ground surface fluctuation and imaging viewpoint variations are usually contained in remote sensing images, which severely restricted their matchability if merely using a simple transformation (such as rigid or affine) model. Third, the putative match set inevitably involves a large number of false matches due to the only use of local descriptors, which is even worse for the complex nature of remote sensing data, such as unavoidable noise, occlusions, repeated structures, and so on. Therefore, a robust mismatch removal approach is required to seek as many correct matches as possible while keeping the mismatch to a minimum.

To address the abovementioned challenges, in this article, we propose a simple yet efficient approach for remote sensing image registration namely linear adaptive filtering (LAF), which can handle both rigid and nonrigid transformations within linear time and space complexity. In particular, our algorithm starts with gridding the putative correspondence space and calculating an average motion vector for each cell, which can convert the sparse points into convolutable Euclidean data. Then, the Gaussian kernel convolution operation is utilized to enhance the connection among the neighboring cells and obtain a typical motion vector for each cell. Finally, the outliers are removed by checking the consistency between each putative match and the corresponding typical motion vector through a threshold. To improve the matching accuracy, we introduce a progressive matching strategy to iteratively refine the typical motion vectors. In addition, an adaptive parameter setting strategy and posterior probability estimation based on the expectation–maximization (EM) algorithm enable our method to be robust to different data. Furthermore, extensive feature matching and image registration experiments with qualitative and quantitative result analyses have shown a significant superiority of our method over state-of-the-art competitors.

Our method has the following three advantages. First, the proposed method does not require a predefined transformation model as many existing methods do, which is more general and can handle both rigid and nonrigid image deformations. Second, the nonorder points (i.e., non-Euclidean data) are converted into a convolutable matrix (i.e., Euclidean data) [15], and hence, we can handle the outliers with a convolution operation, which provides a guide to address the feature matching problem and other sparse point-based tasks using deep learning techniques in the future. Third, the gridding strategy enables our method to achieve linear time and space complexity and fulfill the matching problem in dozens of milliseconds even the putative set contains thousands of matches. This is beneficial for addressing large-scale and real-time remote sensing tasks.

This article is an extension of our previous work in [16], and the primary new contributions include the following three aspects. First, we introduce an efficient strategy for adaptive parameter setting and propose the posterior probability estimation under a maximum-likelihood framework, which can improve the robustness of our method. Second, we improve the strategy of convolution, which avoids the calculation of density in our previous method and thus can save the computation cost. Third, we generalize the proposed method to address the remote sensing image registration problem, with the thin plate spline (TPS) [17] being chosen for transformation estimation.

The remainder of this article is organized as follows. Section II describes the necessary background material and related work. In Section III, we present our LAF algorithm for remote sensing image registration. Section IV illustrates the matching and registration performance of our method in comparison with other approaches on different types of remote sensing images, followed by some concluding remarks in Section V.

II. Related Works

Registration between two or more images is a critical and fundamental process and has been widely used in various fields, including computer vision [14], [18]–[21], pattern recognition [22]–[24], image analysis [4], [25], security [5], [26], and especially in the field of remote sensing [3], [26]–[28]. Comprehensive and exhaustive reviews about image registration are summarized in [1] and [30]–[32]. According to the abovementioned in Section I, image registration is generally divided into three categories, i.e., area-, feature-, and learning-based. In the following, we will give a brief introduction of these three major types of methods, particularly in the remote sensing community.

A. Area-Based Methods

Area-based methods typically register two images based on directly matching image intensities or domain transformed information in a sliding window of predefined size even the entire image. These methods can be broadly classified into three types: correlation-like methods, domain transformation methods, and mutual information (MI) methods [1], [3].

As a classical representative in area-based methods, correlation-like methods correspond two images by maximizing the similarities of two sliding windows [32]. In remote

sensing applications, the maxima correlation of wavelet features has been developed for automatic registration [33]. However, the method of this type may suffer a lot from the serious image deformations (only be successfully applied when slight rotation and scaling occur), the windows containing a smooth area without any prominent details, as well as its huge computing burden. Domain transformed methods tend to align two images based on converting the original images into another domain, such as phase correlation based on Fourier shift theorem [34]–[38] and Walsh transform-based methods [39], [40]. The applications of such methods to remote sensing are depicted in [41]. Such kinds of methods are more robust against the correlated and frequency-dependent noise and nonuniform, time-varying illumination disturbances. However, they have some limitations in the case of image pairs with significantly different spectral contents and small overlap area. Finally, deriving from the information theory, the MI is a measurement of statistical dependence between two images and works with the entire image [42], e.g., nonrigid image registration using MI together with B-splines [43] and conditional MI [44]. Therefore, MI is particularly suitable for registration of remote sensing images captured from different modalities [45]–[47]. However, the disadvantage of MI refers to the difficulty of determining the global maximum of the entire searching space, which inevitably reduces its robustness.

B. Feature-Based Methods

In order to address the challenges in efficiency and robustness of image registration, feature-based methods have been extensively studied [1], [48]. These methods commonly start with extracting stable physical features (usually are interest points or key points) and establishing preliminary correspondences through the similarity of descriptors, then remove the false matches from putative correspondence sets using extra geometrical constraints, simultaneously estimate the transformation, and align overlapped area of two images.

In the first stage, i.e., constructing putative correspondence sets based on similarity of local descriptors, traditional methods are well-known as scale-invariant feature transform (SIFT) [49], speeded up robust features (SURFs) [50], and oriented FAST and rotated BRIEF (ORB) [51]. These classical feature matches have been proven to be both efficient and robust and widely used in various fields. In terms of the unique nature of remote sensing images such as intensity changes and multimodal, Dai and Khorram [52] proposed a feature-based method using improved chain-code representation combined with invariant moments, and Li *et al.* [53] addressed the multimodal image matching problem using radiation-invariant feature transform. In addition, other SIFT improvements [54], [55] or multiple features [56] are also used for remote sensing images. Although there have been various approaches for putative feature correspondence construction, the use of only local appearance information will unavoidably result in a large number of false matches. The problem becomes more severe when images undergo serious nonrigid deformation, extreme viewpoint changes, low quality, and/or repeated contents. Therefore, in the second stage, a robust and efficient mismatch elimination method is needed to detect and remove the outliers.

In order to eliminate mismatches and preserve the true matches, numerous methods have been proposed during the past decades. The most classic methods for this task may be resampling-based methods, such as random sample consensus (RANSAC) [57] and its variants [58], [59]. The common idea of these methods is to find the smallest consistent inlier set to fit a given geometric model following a hypothesize-and-verify strategy. However, resampling-based methods suffer a lot and even fail in the case that the two images undergo nonrigid or other complex deformations. From this point, several nonparametric techniques have been developed, such as identifying correspondence function (ICF) [60] and other nonparameter modeling methods with high-dimensional representation [61], [62]. Methods of this type typically tend to estimate the nonrigid model and remove the outliers simultaneously, which have shown promising matching results in both rigid and nonrigid image deformations. However, the optimal solution will be challenging to determine due to the vanishing smooth priori and the large search space when putative match sets are contaminated by heavy outliers and/or image pairs contain large discontinuous motion, for instance, image pairs captured from wide baseline and images containing multitargets with different motion attributes.

In addition, graph matching-based methods are widely studied as well, which usually formulate the matching problem as a quadratic assignment problem to seek the maximum inlier set with subgraph isomorphism theory [63]–[66]. Some representatives methods of this type include graph shift (GS) [67], graduated consistency regularization [68], and so on. Nevertheless, it is extremely slow for graph matching methods because of their high computational costs, leading to the low applicability for a large-scale matching problem. Recently, several approaches based on locality or piecewise consistency assumption are proposed for fast matching, such as grid-based motion statistics [69], locality preserving matching [14], [70], feature matching using spatial clustering with heavy outliers [24], and learning-based methods [71], [72] (will be introduced in Section II-C). These methods are quite efficient with low computation complexity but cannot work well when the putative set involves a large number of outliers and/or inliers are distributed dispersedly.

As for the applications in remote sensing, there have also been a variety of feature matching methods. For example, a general framework for both rigid and nonrigid registrations is widely studied, including locally linear transforming (LLT) [6], multiscale locality and rank preserving method namely mTopKRP [73], and a guided strategy with the high rate but less number of inliers to obtain more reliable feature correspondences [74]. In addition, Wen *et al.* [75] introduced a unified feature matching criterion by combining spatial consistency and feature similarity. Li *et al.* [76] used support-line voting to remove false matches and subsequently refine the matching results with affine-invariant ratios. Other strategies also include graph matching-based methods [77].

C. Learning-Based Methods

Learning-based approaches are actually covered in area- and feature-based methods. They can be regarded as a direct replacement of traditional methods in information extraction and representation, as well as similarity measurement. Broadly speaking, three strategies are prevalent in current works: 1) training a convolutional neural network (CNN) model to estimate a similarity measure for two images to drive an iterative optimization strategy [78]; 2) to directly predict transformation parameters using deep regression networks [79], [80]; and 3) learning to substitute one or more processes of traditional feature-based methods such as SIFT, e.g., key points detection and description as well as similarity metric measure and matching [81]–[83], moreover establishing sparse point matching from raw images in an end-to-end manner [84]. Details are well surveyed in studies [8], [85].

Learning strategy is widely used in remote sensing images as well and mostly registers directly on image patch pairs with deep features or metric learning. For example, Wang *et al.* [86] proposed an end-to-end architecture to learn directly between patch pairs and their matching labels for later registration. Yang *et al.* [87] proposed a CNN feature-based multitemporal remote sensing image registration method by learning for multiscale feature descriptors and gradually increasing the selection of inliers to improve the registration performance. Another strategy is to register two images using both handcrafted and deep features [88], [89]. This type of learning-based method is more robust to noisy and textureless images. However, they still suffer from the serious deformation and nonrigid transformation, due to the lack of generalization of model and the insufficient training samples.

In addition, learning-based mismatch removal methods have been developed gradually in recent years. Yi *et al.* [71] made a first attempt to introduce a learning-based technique termed learning to find good correspondences (LFGCs). It aims to train a network from a set of sparse putative matches together with the image intrinsics under the rigid geometrical transformation constraints, label the test correspondences as inliers or outliers, and output the camera motion simultaneously. However, LFGC may sacrifice many true correspondences to estimate the motion parameters and fail to handle general matching problems, such as deformable and nonrigid image matching. To this end, Ma *et al.* [72] proposed a general framework to learn a two-class classifier for mismatch removal namely LMR. This method can generate promising matching performance with linearithmic time complexity on arbitrary data, but it may preserve bizarre and obvious false matches due to its limited match representation. Generally speaking, it is quite difficult to apply the CNNs well onto sparse point sets for the points classification, mismatch removal, or corresponding. The major reason is that point data, also called non-Euclidean data [15] as their unordered and dispersed nature, are challenging to operate and extract the spatial relationships between two or more points (e.g., neighboring elements, relative positions, length, and angle information among multipoints) using a deep convolutional technique.

III. Method

This section describes an efficient feature matching method for registering two remote sensing images of the same or similar scenes. To this end, we start by constructing a set of putative matches with the similarity of feature descriptors such as SIFT [49]. Then, the matching task boils down to rejecting the false matches from the given putative set using extra geometrical constraints and smooth priori. Subsequently, by using the preserved reliable feature correspondences, the transformation between the two given images can be estimated accordingly.

A. Problem Formulation

Given the sensed image I and reference image I' of the same scene or object to be registered, suppose that we have obtained a set of N putative matches $\mathcal{S} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N}$, where $\mathbf{x}_i = (u, v)^{\mathrm{T}}$ and $\mathbf{y}_i = (u', v')^{\mathrm{T}}$ are the pixel coordinates (i.e., extracted feature points) of I and I', respectively. Let $\mathcal{F}$ indicate the transformation or mapping function from $\mathbf{x}$ to $\mathbf{y}$, and then, for any true match $(\mathbf{x}_i, \mathbf{y}_i)$, we have $\mathbf{y}_i = \mathcal{F}(\mathbf{x}_i)$. However, the putative set $\mathcal{S}$ is inevitably contaminated by some unknown noise and outliers, which strongly encourage us to remove the outliers and establish accurate correspondences. To this end, a general assumption is that the noise on inliers is isotropic Gaussian with a zero mean and a covariance matrix $\sigma^2\mathbf{I}$, i.e., $\mathbf{y}_i - \mathcal{F}(\mathbf{x}_i) \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$, where $\mathbf{0}$ is a 2-D vector of zeros and $\mathbf{I}$ is a 2×2 identity matrix, and the outlier is random uniform distribution of $1/a$ [6], [58], where a is the area of output in reference image. Thus, the mixture model can be formulated as the following form:

$$p(\mathbf{y}_i|\mathbf{x}_i, \theta) = \frac{\gamma}{2\pi\sigma^2} e^{-\frac{\|\mathbf{y}_i - \mathcal{F}(\mathbf{x}_i)\|^2}{2\sigma^2}} + \frac{1-\gamma}{a} \tag{1}$$

where parameter set $\theta = \{\mathcal{F}, \sigma^2, \gamma\}$ with γ being the mixing coefficient. Let $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)^{\mathrm{T}}$ and $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_N)^{\mathrm{T}}$ be the $N \times 2$ matrix representing the extracted two feature sets, respectively. By taking the assumption that the data satisfy independent and identically distributed (i.i.d.), the likelihood function can be written as

$$p(\mathbf{Y}|\mathbf{X}, \theta) = \prod_{i=1}^{N} p(\mathbf{y}_i, |\mathbf{x}_i, \theta). \tag{2}$$

Then, the maximum-likelihood estimation of parameter set can be converted as the following minimization form:

$$\mathcal{E}(\theta) = -\ln p(\mathbf{Y}|\mathbf{X}, \theta) = -\sum_{i=1}^{N} \ln p(\mathbf{y}_i|\mathbf{x}_i, \theta). \tag{3}$$

B. LAF

From the formulation aforementioned, we can find that it is extremely difficult to directly seek the global minimum of the negative log-likelihood function (3) to obtain the optimal parameter set θ. The estimation of transformation $\mathcal{F}$ may require the cubic computation complexity, and it is usually calculated repeatedly in an iteration way [6], [61], leading to

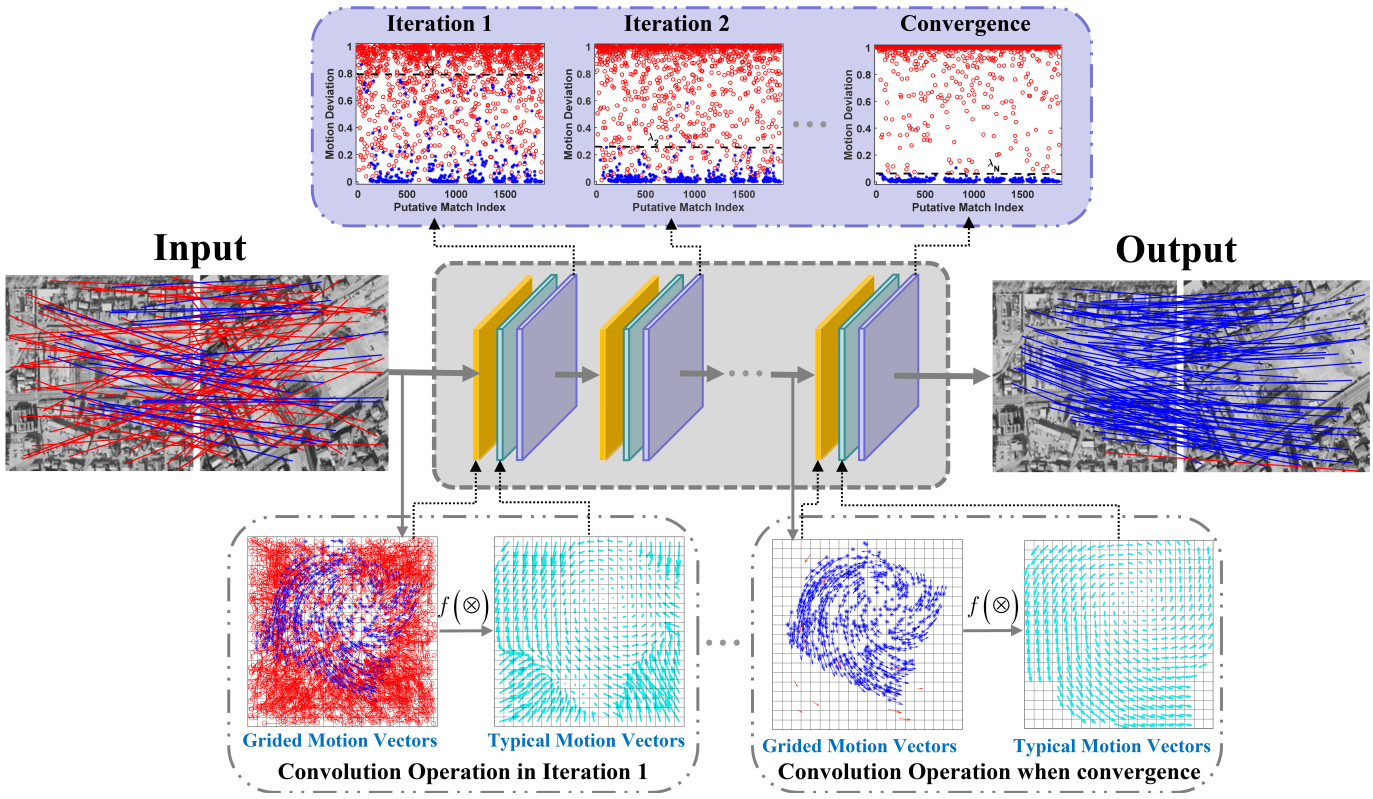

Fig. 1. Proposed progressive filtering framework for robust feature matching. Second row: main body. There are three steps in each iteration indicated with cascaded colored plates, saying match space gridding (yellow), kernel convolution to generate typical motion field (cyan), and motion consistency checking or motion deviation calculation (purple). Putative matches with the deviation under a given threshold λ will be preserved as the input in the next iteration, which is just used for typical motion field generation, and the motion deviations are still calculated based on the entire putative set. First row: enlarged separability of motion consistency between inliers and outliers using iterative filtering strategy. Third row: gridding and convolution results in the first and last iterations. For a better visibility, at most 100 randomly selected matches are presented in the input and output image pairs, and motion vectors are displayed with a quarter of their actual length in the third row. Blue: inlier. Red: outlier.

a huge computation burden in actual applications. Moreover, severe outliers and serious deformation can invalidate the whole process due to the difficulty in finding the optimal solution. To this end, based on the smooth priori of potential correct matches, we reformulate the mixture model in (1) and the target function in (3) as an approximate form without a predefined and the heavy computation of $\mathcal{F}$, hence reducing it to linear complexity.

1) Problem Approximation: According to image denoising, when given a noisy image, the common idea is to consider the pixels in a local area (determined by the convolution kernel size) comprehensively, such as mean or median operation, to update and obtain the true pixel intensity and filter the Gaussian or salt noise. Similarly, by transforming the putative match set $\mathcal{S}$ into $\mathcal{S}' = \{(\mathbf{x}_i, \mathbf{m}_i)\}_{i=1}^{N}$ with $\mathbf{m}_i = \mathbf{y}_i - \mathbf{x}_i$ denoting the motion vector of match $(\mathbf{x}_i, \mathbf{y}_i)$, then the potential true matches tend to be regular and smooth, i.e., geometrical consistency[1] [14], [69]. In this way, it is feasible to calculate the average motion vector on the potential true matches in a small region and reject the false matches by checking the deviation between each putative motion vector and the average motion vector based on the consistency.

[1]The geometrical consistency denotes that the correct matches should have a similar motion behavior, as least in local neighborhoods, whereas the false matches are usually randomly distributed, seen the motion behavior in Fig. 1 or 3 for example.

To this end, we divide each dimension of feature points $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^{N}$ into n_c nonoverlapping parts equally and obtain $G = n_c \times n_c$ cells. Accordingly, the putative set $\mathcal{S}'$ can be divided into G parts with $\mathcal{X} = \{\mathcal{C}_{j,k}\}_{j,k=1}^{n_c}$, which is shown as the gridded putative motion vectors in Fig. 1 ($n_c = 20$). Also, we denote $\overline{\mathbf{M}}$ as the average motion matrix, where $\overline{\mathbf{M}}_{j,k}$ is the average motion vector in the (j,k)th cell determined by

$$\overline{\mathbf{M}}_{j,k} = \begin{cases} \dfrac{1}{|\mathcal{C}_{j,k}|} \displaystyle\sum_{i|\mathbf{x}_i \in \mathcal{C}_{j,k}} \mathbf{m}_i, & \text{if } |\mathcal{C}_{j,k}| > 0 \\ \mathbf{0}, & \text{if } |\mathcal{C}_{j,k}| = 0. \end{cases} \tag{4}$$

Next, we convert the problem formulation into the following approximation form:

$$\begin{aligned} \mathbf{y}_i - \mathcal{F}(\mathbf{x}_i) &= (\mathbf{y}_i - \mathbf{x}_i) - (\mathcal{F}(\mathbf{x}_i) - \mathbf{x}_i) \\ &= \mathbf{m}_i - (\mathcal{F}(\mathbf{x}_i) - \mathbf{x}_i) \end{aligned} \tag{5}$$

where $\mathcal{F}(\mathbf{x}_i) - \mathbf{x}_i \approx \overline{\mathbf{M}}_{j,k}, \forall i, \mathbf{x}_i \in \mathcal{C}_{j,k}$ holds, particularly when $\overline{\mathbf{M}}$ is calculated by only inliers and n_c is large. Thus, we can obtain $\mathbf{y}_i - \mathcal{F}(\mathbf{x}_i) \approx \mathbf{m}_i - \overline{\mathbf{M}}_{j,k}, \mathbf{x}_i \in \mathcal{C}_{j,k}$. From this point, with $\theta = \{\sigma^2, \gamma\}$, we can approximate the mixture model as follows:

$$p(\mathbf{y}_i|\mathbf{x}_i, \theta) \approx \frac{\gamma}{2\pi\sigma^2} e^{-\frac{\|\mathbf{m}_i - \overline{\mathbf{M}}_{j,k}\|^2}{2\sigma^2}} + \frac{1-\gamma}{a} \quad \forall i, \mathbf{x}_i \in \mathcal{C}_{j,k}. \tag{6}$$

By using (6) and (3), we can find the optimal solution of θ^*; in particular, there is no need to estimate the transformation

$\mathcal{F}$ in this process. Therefore, the problem is converted into the estimation of average motion matrix $\overline{\mathbf{M}}$ that is ideally calculated only based on potential true matches. To this end and following the filtering strategy, we define the motion deviation as $\{\mathbf{e}_i = \mathbf{m}_i - \overline{\mathbf{M}}_{j,k}, \forall i, \mathbf{x}_i \in \mathcal{C}_{j,k}\}_{i=1}^{N}$. Therefore, $\mathbf{e}_i$ has a similar distribution assumption with $\mathbf{y}_i - \mathcal{F}(\mathbf{x}_i)$, i.e., $\mathbf{e}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}), i \in \mathbb{N}_{\text{inlier}}$; $\mathbf{e}_i \sim \mathcal{U}(-b\mathbf{I}, b\mathbf{I}), i \in \mathbb{N}_{\text{outlier}}$ where the random uniform distribution $\mathcal{U}$ with probability density $1/a$ [6], [58], with $a = 2b \times 2b$; $\mathbf{I}$ is a unit matrix, and $\mathbb{N}_{\text{inlier}}$ and $\mathbb{N}_{\text{outlier}}$ denote inlier set and outlier set in $\mathcal{S}$, respectively. Based on the distribution difference between inliers and outliers, we can remove the false matches with a specified filter.

Nevertheless, there are still two limitations in the aforementioned strategy. On the one hand, for an isolated sample, i.e., $\mathbf{x}_i \in \mathcal{C}_{j,k}$ and $|\mathcal{C}_{j,k}| = 1$, then $\overline{\mathbf{M}}_{j,k} = \mathbf{m}_i$, and the deviations may keep being zero for both true and false matches, thus easily leading to some misjudges. On the other hand, the connections among the neighboring cells may be neglected if only using the average operation in a single one, which may badly degrade the consistency of potential true matches particularly when there exist numerous outliers (which often occurs in the feature matching problem). Therefore, in the following, we propose an efficient filter strategy with a Gaussian convolutional kernel to recover the true motion field and remove the outliers simultaneously.

2) Convolution Operation: In order to utilize the interaction among neighboring cells, we consider the local $n_k \times n_k$ cells comprehensively according to the convolution theory. The convolution $f(\otimes)$ of putative motion vectors is defined as

$$f(\otimes) : \widetilde{\mathbf{M}} = \frac{(\mathbf{W} \cdot \overline{\mathbf{M}}) \otimes \mathbf{K}}{\mathbf{W} \otimes \mathbf{K} + \varepsilon} \tag{7}$$

where $\widetilde{\mathbf{M}}$ is the generated $n_c \times n_c \times 2$ matrix after Gaussian kernel convolution, with $\widetilde{\mathbf{M}}_{j,k}$ the typical motion vector of cell (j, k) and $\mathbf{W}$ a count matrix with $\mathbf{W}_{j,k} = |\mathcal{C}_{j,k}|$. The denominator in (7) is used for weight compensating to preserve the scale of convolution results, and ε is an infinitesimal positive number in case that there exists 0 in $\mathbf{W} \otimes \mathbf{K}$. $\mathbf{K}$ is a Gaussian kernel distance matrix of size $n_k \times n_k$, which is defined as

$$\mathbf{K}_{i,j} = \frac{\exp\{-\mathbf{D}_{i,j}\}}{\sum_{i=1}^{n_k}\sum_{j=1}^{n_k}\exp\{-\mathbf{D}_{i,j}\}}, \ \mathbf{D}_{i,j} = \|\mathbf{s}_{i,j} - \mathbf{s}^*\|_2 \tag{8}$$

where $\mathbf{s}_{i,j} = (i, j)^{\mathrm{T}}$ and $\mathbf{s}^* = (\lceil n_k/2 \rceil, \lceil n_k/2 \rceil)^{\mathrm{T}}$ are the corresponding position and the central position in the convolutional kernel $\mathbf{K}$, respectively, and $\lceil \cdot \rceil$ rounds the element to the nearest integer not less than itself. Therefore, n_k must be a positive odd number.

In addition, to avoid the influence of isolated samples, we do not take the isolated one into account during convolutional procedure by subtracting the corresponding average vectors and adjusting the weight in each cell and hence (7) can be rewritten as

$$f(\otimes) : \widetilde{\mathbf{M}} = \frac{(\mathbf{W} \cdot \overline{\mathbf{M}}) \otimes \mathbf{K} - \overline{\mathbf{M}} \cdot \mathbf{K}^*}{\mathbf{W} \otimes \mathbf{K} - B(\mathbf{W}) \cdot \mathbf{K}^* + \varepsilon} \tag{9}$$

where $B(\mathbf{W})$ indicates the binary form of $\mathbf{W}$ with the values being 0 or 1, $B(\mathbf{W}_{i,j}) = 1$ only when $\mathbf{W}_{i,j} > 0$, and $\mathbf{K}^*$ means the center element, i.e., $\mathbf{K}^* = \mathbf{K}_{\lceil n_k/2 \rceil, \lceil n_k/2 \rceil}$.

After the convolution, we obtain the typical motion vector of each cell, as shown by the cyan color vector located in the center of each cell in Fig. 1. Then, we define the deviation between $\mathbf{m}_i$ and the corresponding $\widetilde{\mathbf{M}}_{j,k}$ and constraint them between 0 and 1 with

$$d_i = 1 - \exp\Big\{-\frac{\|\mathbf{m}_i - \widetilde{\mathbf{M}}_{j,k}\|^2}{\beta^2}\Big\} \quad \forall i, \ \mathbf{x}_i \in \mathcal{C}_{j,k} \tag{10}$$

where β is used for determining the width of the range of interaction between two motion vectors, and we empirically set $\beta^2 = 0.08$. Thus, the inlier set $\mathcal{I}^*$ can be approximately detected by comparing the deviation and a given threshold λ

$$\mathcal{I}^* = \{(\mathbf{x}_i, \mathbf{y}_i) : d_i \le \lambda, i \in \mathbb{N}_N\}. \tag{11}$$

3) Progressive and Adaptive Filtering: From Fig. 1, we can find that the margin between inliers and outliers is not so separable (as shown in Iteration 1), and only a part of false matches can be filtered out by threshold λ_1. Ideally, if the typical motion vectors are constructed only by inliers, then the deviation of inliers and outliers will almost tend to 0 and 1, and therefore, we can separate inliers and outliers more easily and accurately. Nevertheless, the ground truth inliers are not available in advance. To solve this dilemma, we propose an iteration strategy to remove outliers progressively. It iteratively refines the typical motion vector and anneals the threshold λ based on the coarse-to-fine theory. The inlier set is approximated with the results of each iteration until convergence. As shown in the first line of Fig. 1, the deviation margin between inliers and outliers has been distinctly enlarged as the iteration proceeds. By the way, it is enough to obtain reliable matching performance within five iterations with λ being 0.8, 0.2, 0.1, 0.05, and 0.035 in each iteration, respectively. Even so, there are still some outliers that their motion deviations d are close even mixed to inliers, leading to the sensitivity of parameter λ. This is because a small number of outliers are slightly deviated from the correct position and may keep weak consistent with the typical motion vectors. Moreover, some false matches, with small lengths of their motion vectors, may cause small deviation when the according cells do not contain any true matches, i.e., $\widetilde{\mathbf{M}}_{j,k} = 0$.

By now, we have formulated the feature matching problem into a convolution filtering task. However, there are several hyperparameters, i.e., $\{n_c, n_k, \lambda\}$, seriously affecting the filtering results, due to that the optima parameter λ may change largely with respect to the values of n_c and n_k, especially the rotation and scaling as well as some complex nonrigid deformations between two images. Therefore, we test it using randomly selected 50 remote sensing image pairs with different types of transformations involving rigid, rotation, scale changing, and nonrigid deformations, and so on.

First, we give the motion deviation [calculated by (10)] statistics of inliers and outliers with respect to 12 combinations of (n_c, n_k) in the left plot of Fig. 2, where n_c ranges from 10 to 40 with interval 10 (denoted by four colors) and n_k ranges from 5 to 9 with the interval of 2 (denoted by three

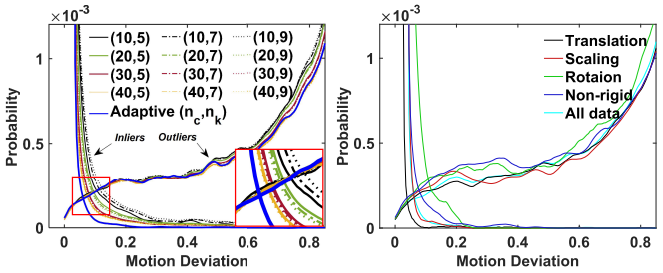

Fig. 2. Probability distributions of motion deviations [calculated by (10)] for inliers and outliers with respect to different combinations of n_c and n_k, as well as different types of image deformations. The results are obtained on 50 selected representative remote sensing image pairs. (Left) Using four colors to indicate different n_c, and using full line, chain line, and dotted line to indicate $n_k = 5$, 7, and 9, respectively, whereas the blue line (the last one in the legend) indicates the adaptive setting of (n_c, n_k). (Right) Distributions of putative matches on different types of image deformations that are divided from the selected test data. For each figure, the inliers are indicated with the downward-trend lines and outliers with upward-trend lines. The cross point of inlier line and outlier line, with the same color and type, denotes the best choice of parameter λ, and the cross point located closer to the lower left corner means better separability between inliers and outliers.

Algorithm 1: LAF Algorithm

Input: Putative set $\mathcal{S} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N}$, parameters λ, τ
Output: Inlier set $\mathcal{I}$

Initialize the inlier set as the whole set $\mathcal{S}$;
Set the parameters n_c, n_k and $\mathbf{K}$ using Eqs. (12) and (8);
Convert $\mathcal{S}$ into $\mathcal{S}'$ and gridding;
Iteration:
  Construct matrices $\overline{\mathbf{M}}$ and $\mathbf{W}$ using Eq. (4);
  Calculate matrix $\widetilde{\mathbf{M}}$ using Eq. (9);
  Calculate the deviations using Eq. (10);
  Initialize the posterior probability using Eq. (13);
  Estimate σ^2 and γ using Eqs. (14) and (15);
  Calculate probability p_i using Eq. (16);
  Determine $\mathcal{I}$ using Eq. (17);
Until convergence;
Return $\mathcal{I}$.

line types, i.e., full line, chain line, and dotted line). Note that line with the downward trend means inlier statistic, whereas the upward trend means outlier statistic. The optimal parameter λ of each group of (n_c, n_k) is the cross point of probability curves of inlier and the according outlier, i.e., two lines with the same color and type. The cross point closer to the lower left corner indicates the better separability between inliers and outliers. From the left plot of Fig. 2, we can find that the motion deviation statistic of outliers is almost unchanged with respect to different values of gridding size n_c and kernel size n_k, just as the upward-trend curves shown. While the curves of inliers (downward-trend) become lower left when increasing n_c and n_k. However, the superiority may become smaller when n_c increases to 30; instead, larger n_c will increase the computation burden. Therefore, according to the relationship between n_c and scale of the putative match set, i.e., N, we assume that the average number of putative match located in one cell must not be less than 1, but constrains it between 15 and 30 due to the robustness and efficiency

$$\begin{cases} n_c = \min\{\max\{[\sqrt{N}], 15\}, 30\} \\ n_k = \text{odd}(n_c/3) \end{cases} \tag{12}$$

where N is the number of putative set, $[\cdot]$ rounds the element, and $\text{odd}(n_c/3)$ means the nearest odd number not greater than $n_c/3$. By using the adaptive setting of n_c and n_k, the margin between inliers and outliers may be largely enlarged, as the blue curve shown in the left plot of Fig. 2.

In addition, the best choice of λ may change a lot with respect to different types of image deformations, as shown in the right plot of Fig. 2. The experiment is conducted by dividing the 50 test image pairs into five groups, saying translation, scaling, rotation, nonrigid, and all included, from which we can see that different types of image deformation may result in great differences of these cross points, meaning that it requires deferent optimal λ, especially in the situation of the rotation and nonrigid. This is because the distribution parameters, i.e., the covariance of inliers, may vary with different image data, and hence, it requires us to identify the inliers accurately based on their covariance values, instead of using the fixed λ only. Fortunately, the EM algorithm [61], [90] is often used to address such a latent variable estimation problem in (3). It can estimate the necessary parameters and is more robust to different data. Based on the estimation of posterior probability, i.e., $p_i = P(i \in \mathbb{N}_{\text{inlier}} | \mathbf{x}_i, \mathbf{y}_i, \theta^{\text{old}})$, which indicates to what degree $(\mathbf{x}_i, \mathbf{y}_i)$ being an inlier, we first initialize it using (11) and obtain

$$p_i = \begin{cases} 0, & d_i > \lambda \\ 1, & d_i \le \lambda. \end{cases} \tag{13}$$

Subsequently, let $\widetilde{\mathbf{e}}_i = \mathbf{m}_i - \widetilde{\mathbf{M}}_{j,k}, \forall i, \mathbf{x}_i \in \mathcal{C}_{j,k}$ and $\mathbf{P} = \text{diag}(p_1, \ldots, p_N)$ be a diagonal matrix. We can obtain

$$\sigma^2 = \frac{\text{tr}(\mathbf{E}^{\text{T}}\mathbf{P}\mathbf{E})}{2 \cdot tr(\mathbf{P})} \tag{14}$$

$$\gamma = \frac{\text{tr}(\mathbf{P})}{N} \tag{15}$$

where $\mathbf{E} = (\widetilde{\mathbf{e}}_1, \ldots, \widetilde{\mathbf{e}}_N)^{\text{T}}$, $\text{tr}(\cdot)$ denotes the trace of a matrix. Next, based on the Bayes rule, the probability p_i can be accurately estimated with

$$p_i = \frac{\gamma\, e^{-\frac{\|\mathbf{m}_i - \widetilde{\mathbf{M}}_{j,k}\|^2}{2\sigma^2}}}{\gamma\, e^{-\frac{\|\mathbf{m}_i - \widetilde{\mathbf{M}}_{j,k}\|^2}{2\sigma^2}} + \frac{2\pi\sigma^2(1-\gamma)}{a}} \quad \forall i,\ \mathbf{x}_i \in \mathcal{C}_{j,k}. \tag{16}$$

Finally, with a predefined threshold τ, the inlier set $\mathcal{I}$ could be obtained by the following criterion:

$$\mathcal{I} = \{(\mathbf{x}_i, \mathbf{y}_i) : p_i > \tau, i \in \mathbb{N}_N\} \tag{17}$$

which is more robust to threshold τ and can achieve better performance than the criterion in (11).

Since the proposed robust feature matching method is based on a convolution filtering strategy, which can accurately recover the motion field and remove false matches with adaptive hyperparameters setting in a linear time complexity, we name it as LAF and summarize the whole procedure in Algorithm 1.

4) Computational Complexity: To convert the putative set into nonoverlapping cells, the quotients of $\mathcal{X}$ are required to calculate over the divided interval. Therefore, the time cost of initialization, parameters setting, putative set converting, and gridding from lines 1 to 3 of Algorithm 1 is around $O(N)$. The average motion vectors and count matrix are calculated in each cell and each match is only used once, which costs $O(N)$ time complexity as well. As for the convolution operation, it depends on the cell number and the kernel size, which has time complexity close to $O(n_k^2 \times n_c^2)$. In addition, calculating the deviations and initializing the posterior probability using (10) and (13) in lines 7 and 8 cost $O(N)$ complexity. Furthermore, the estimation of σ^2 and γ and the calculation of p_i as well as to obtain inlier set in lines 9–11 cost $O(N)$ time complexity too. Since our LAF algorithm can converge in very few iterations (typically five times), the total time complexity of our LAF is about $O(n_k^2 \times n_c^2 + N)$. The space complexity of our algorithm is $O(N)$ due to the memory requirement for only storing the putative set and the deviation. Generally, n_k and n_c are constants and both much smaller than N. If the value N is large enough, both the time and space complexities of our method can be simply written as $O(N)$, that is to say, the time and space consuming of our method is linear with respect to the sample number N, which is significant for addressing large-scale or real-time remote sensing problems.

C. Transformation Estimation and Image Registration

Once we have obtained the reliable feature correspondences with the proposed LAF, then we can use it to estimate the transformation function $\mathcal{F}$ accordingly. However, since the remote sensing images typically undergo complex nonrigid transformation and local distortion or are captured from fisheye cameras, simple parameter models are no longer workable. Therefore, we choose TPS [17] for transformation parameterizing due to its generality and smooth functional mapping nature in supervised learning [91]. Thus, it can represent the nonrigid transformation in feature matching problem. In addition, TPS has no free parameters without manual tuning and also has a closed-form solution that can be decomposed into a global linear affine motion and a local nonaffine warping component. The formulation details can refer to [17].

Finally, for each pixel in the sensed image, we use the estimated transformation function $\mathcal{F}$ to calculate the corresponding coordinate in the reference image, and then use a bicubic interpolation algorithm to calculate the intensity at that coordinate in the reference image.

D. Implementation Details

There may exist multiple putative matches sharing a common feature point, i.e., $\mathbf{x}_i = \mathbf{x}_j$ or $\mathbf{y}_i = \mathbf{y}_j, i \neq j$, which would degrade the matching performance, and hence, we initialize these putative matches as outliers and recall the potential true matches from them using convolution operation. Furthermore, we normalize them ranging from 0 to 1, to eliminate the influence of the coordinate scale of feature points. In addition, we set $a = 16$, based on the area of output being $[-2, 2] \times [-2, 2]$, and empirically set $\tau = 0.8$. Note that the probabilities of inliers are close to 1 and outliers close to 0 after convergence, and hence, the choice of τ is not that sensitive. Our LAF uses an iterative strategy to filter the outliers progressively, which is similar to deterministic annealing. Thus, we set the threshold λ to a large value at the beginning and then decrease it gradually with respect to iteration. By testing on selected 50 image pairs and using the adaptive (n_c, n_k), the optimal choice of λ is related to the cross point of inlier and outlier statistic lines, as shown in Fig. 2. Therefore, guided by the cross point locations and to achieve optimal matching performance on these test data, throughout this article, we experimentally set the iteration number as 5, and $\lambda = 0.8$, 0.2, 0.1, 0.05, and 0.05 in each iteration.

IV. Experimental Results

In this section, we test the performance of our proposed LAF [2] on feature matching and image registration experiments and compare it with other representative state-of-the-art feature matching methods, such as RANSAC [57], ICF [60], GS [67], LLT [6], LMR [72], and mTopKRP [73]. The parameters are set according to the original articles and fixed throughout our experiments. For LLT, we select the adaptive model for each data set. The open-source VLFeat toolbox [92] is employed for putative math set construction with the SIFT descriptor. All the experiments are conducted on a desktop with 4.0-GHz Intel Core i7-6700K CPU, 16-GB memory, and MATLAB code.

A. Data Sets and Settings

To evaluate the performance of our method, we use seven remote sensing image data sets, in which five data sets come from [73]. These five data sets include 25 pairs of 600×337 unmanned aerial vehicle images (i.e., the UAV data set), 34 pair of 256×256 or 800×800 synthetic aperture radar images (i.e., the SAR data set), 31 pairs of 561×518 or 600×700 panchromatic aerial photograph images (i.e., the PAN data set), 40 pairs of 700×700 color infrared aerial photographs images (i.e., the CIAP data set), and 30 pairs of 1280×1024 or 1088×1088 fisheye images (i.e., the FE data set), which, respectively, undergo projective, similarity, or rigid, affine or projective, rigid, and nonrigid transformations. In addition, we use the 720Yun data set [93][3] for nonrigid test. This Cloud data set contains 30 pairs of images involving terrain, roads, buildings, terraces, and so on, with the resolution being from 496×489 to 800×800. Since the raw images on the 720 Yun platform are video panoramic images containing ground surface fluctuation and imaging viewpoint variations, each pair of images will undergo nonrigid transformation in the process of data acquisition. To evaluate the performance on images of large resolution, we collect 15 pairs of images with rigid transformations cropped from the GF-II image. We call this data set GF-II, and the image size is fixed at 2048×2048. The initial match number in each image pair ranges from 3896 to 4827, and the inlier rate ranges from 0.1251 to 0.6211. This may result in huge computational burden for many matching methods.

[2]MATLAB Code for LAF: https://github.com/StaRainJ/LAF

[3]720Yun: https://720yun.com/

To ensure objectivity, we manually check each putative match to be true or false as the ground truth before conducting any experiments. In the experimental procedure, the F-score is used for evaluating the matching performance, which is defined as $F-\text{score} = 2 * \text{Precision} * \text{Recall}/(\text{Precision} + \text{Recall})$, where the Precision (P) is defined as the ratio between the identified correct match number and the preserved match number and the Recall (R) is defined as the ratio between identified correct match number and the correct match number contained in the putative set. In addition, the root mean square error (RMSE), maximum error (MAE), and median error (MEE) are used for measuring the accuracy of image registration with the following definitions:

$$\text{RMSE} = \sqrt{1/L \sum_{i=1}^{L} (r_i^c - \mathcal{F}(s_i^c))^2} \tag{18}$$

$$\text{MAE} = \max\left\{\sqrt{(r_i^c - \mathcal{F}(s_i^c))^2}\right\}_{i=1}^{L} \tag{19}$$

$$\text{MEE} = \text{median}\left\{\sqrt{(r_i^c - \mathcal{F}(s_i^c))^2}\right\}_{i=1}^{L} \tag{20}$$

where r_i^c and s_i^c are the corresponding landmarks (i.e., pixel coordinates) of reference images and the sensed images, respectively, $\mathcal{F}$ is the transformation function from sensed image to reference image, L represents the number of selected landmarks, and $\max(\cdot)$ and $\text{median}(\cdot)$ return the maximal and median value of a set, respectively.

B. Results on Feature Matching

1) Qualitative Illustration: We first give qualitative results of our proposed LAF on some typical image pairs in Fig. 3. From top to bottom, each row contains two examples and is chosen from the seven data sets, i.e., UAV, GF-II, SAR, PAN, CIAP, 720Yun, and FE. These image pairs are challenging for the mismatch removal task due to their high outlier rates, small overlapping areas, scaling, rotation, and even nonrigid deformations. With our LAF algorithm, the precision, recall, and F-score on these image pairs are (98.96%, 99.65%, 0.9931), (99.81%, 99.07%, 0.9944), (99.63%, 100.0%, 0.9982), (99.80%, 99.95%, 0.9987), (100.0%, 100.0%, 1.000), (98.84%, 100.0%, 0.9941), (99.01%, 99.40%, 0.9921), (99.84%, 100.0%, 0.9992), (100.0%, 100.0%, 1.000), (100.0%, 100.0%, 1.000), (98.03%, 99.25%, 0.9864), (97.57%, 99.72%, 0.9863), (99.80%, 99.39%, 0.9959), and (98.21%, 98.21%, 0.9821). From the results, we can easily find that most inliers can be identified by using our LAF, with only a few misjudged. This demonstrates the generality and robustness of our method to handle different types of image deformations and a large number of outliers.

2) Quantitative Comparison: Next, we will evaluate the feature matching performance of our LAF in a quantitative way. To this end, the above mentioned seven data sets are divided into three groups such as rigid data set (SAR, CIAP, and GF-II), projective data set (UAV and PAN), and nonrigid data set (720Yun and FE), and the average putative match number of these three groups is 1, 691.0, 1, 372.1, and 651.38, respectively. The cumulative distribution of initial inlier ratios on these three data sets is provided in the first row of Fig. 4, and each column presents the rigid, projection, and nonrigid data sets, respectively. We see that the inlier ratio on the rigid data set is generally high, with a few image pairs being challenging because of low inlier ratio, some of which may undergo scaling and rotation deformations. As for the projective and nonrigid data sets, they typically suffer from low inlier rate and nonrigid deformations, respectively, which are challenging for the mismatch removal task. From the second to the last rows in Fig. 4, the statistical results about precision, recall, F-score, and runtime on the three data sets are comprehensively reported and summarized. From the statistical results, we find that all methods obtain good results on the rigid data set except for ICF. The reason is that the rigid data set almost contains simple transformation image pairs and thus easy to handle, while the poor performance of ICF is mainly because of its relaxed spatial constraint and parameter sensitivity. Compared with GS and RANSAC methods, LAF would preserve more correct matches and achieve a better recall rate. Although LLT obtains comparative results, it performs not well on low inlier rate data and the transformation model needs to be set manually. The recently proposed methods LMR and mTopKRP achieve promising performance too but are slightly worse than our proposed LAF.

The comprehensive performances are clearly characterized with the F-score statistic, from which we can observe that our method is superior to other methods. For the projective data set, the comparison results are similar with the rigid data set for the same reason. Specifically, our method can keep robust to a large number of outliers and achieve the best precision and recall performance, whereas the other methods may be degraded even fail in some cases. As for the nonrigid data set, LLT, mTopKRP, and LAF have significant superior performance, whereas the performances of RANSAC and GS methods are degraded a lot for the nonrigid deformation, and simultaneously, the ICF still performs the worst for the unworkable correspondence function definition and parameter sensitivity. As for the learning-based method LMR, it has promising precision but low recall because the fisheye images are not contained in its training samples. From the comprehensive results, i.e., F-score, we can clearly observe that our proposed LAF can obtain the best performance, which demonstrates its effectiveness and robustness in addressing nonrigid data. Most importantly, our method has a relatively low complexity, i.e., linear complexity, as shown in the last row in Fig. 4, which is surprisingly fast than other methods especially when handling tens of thousands of matches.

C. Results on Image Registration

1) Qualitative Illustration: The registration experiment focuses on whether the transformed image can maximize the alignment of the overlapping area between the reference and sensed images. To this end, we first give visual registration results on typical image pairs in Fig. 5. From top to bottom, the first row presents the original images, where the left and right in each group are reference and sensed images, respectively. The second to the last rows present the registration

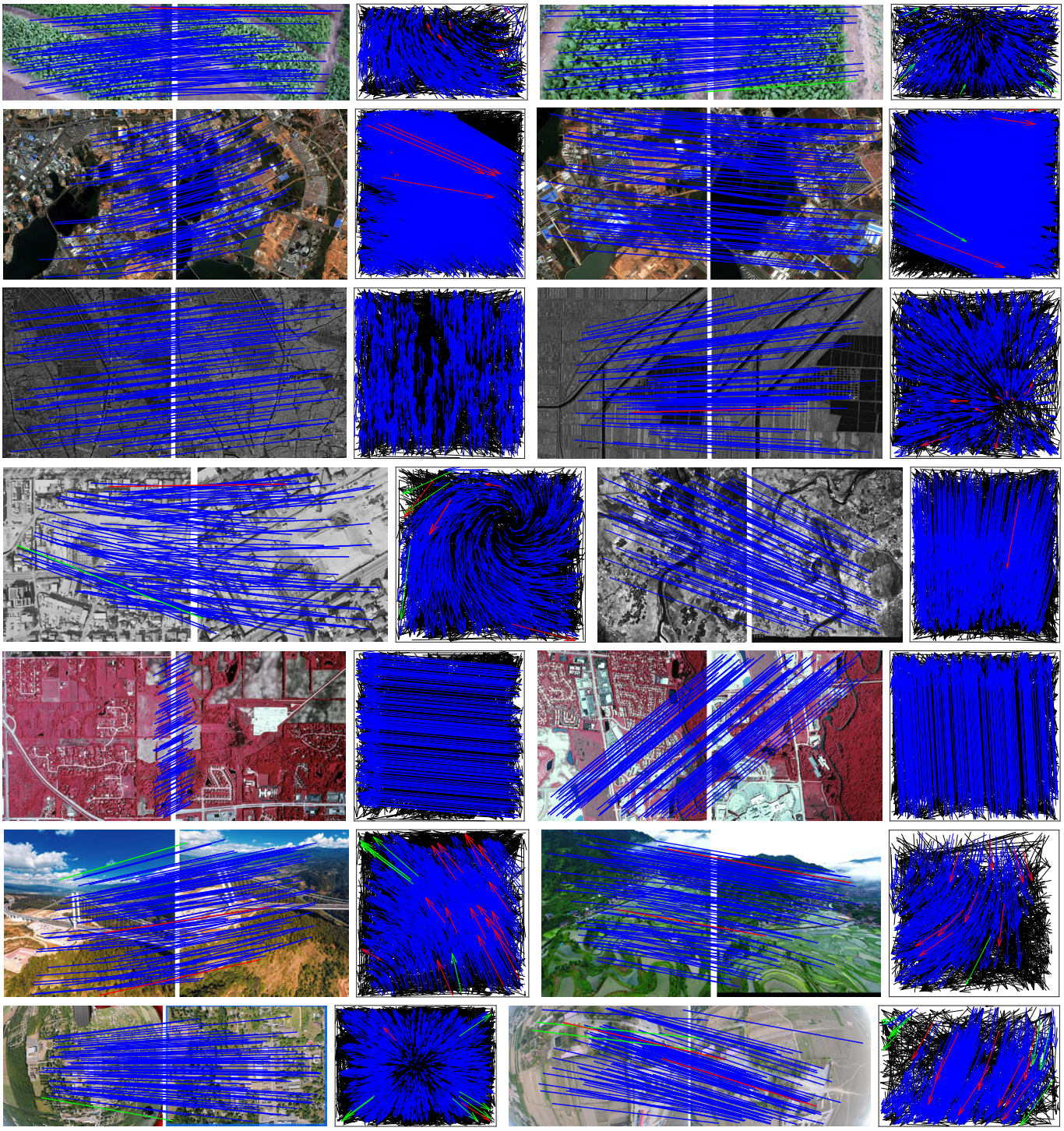

Fig. 3. Feature matching results of our LAF on 14 representative remote sensing image pairs. (From top to bottom and left to right) UAV1, UAV2, GF-II1, GF-II2, SAR1, SAR2, PAN1, PAN2, CIAP1, CIAP2, Yun1, Yun2, FE1, and FE2. The initial correspondence numbers of these 14 image pairs are 1351, 1302, 4232, 4582, 2243, 1982, 1875, 2304, 2152, 2648, 2198, 1154, 3081, and 1000, with the inlier rates being 42.49%, 41.40%, 44.58%, 42.99%, 39.77%, 42.84%, 26.83%, 26.39%, 10.59%, 11.82%, 36.58%, 31.37%, 31.97%, and 50.30%, respectively. The head and tail of each arrow in the motion field correspond to the positions of feature points in the two images (blue = true positive, black = true negative, green = false negative, and 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.

results of all comparing methods, where the left and right in each group are checkboard results and the warped sensed images, respectively. From left to right, the first two columns are chosen from projective and rigid data sets, respectively; the middle column represents fisheye image pair, and the last two are taken from the 720Yun data set. From the qualitative registration results, we can find that RANSAC, ICF, and GS can achieve satisfying performance due to their global constraints, which can obtain high precision in feature matching task and estimate the transformation model more accurately. However, they would suffer from the nonrigid deformation and/or heavy outliers, for instance, RANSAC merely estimates

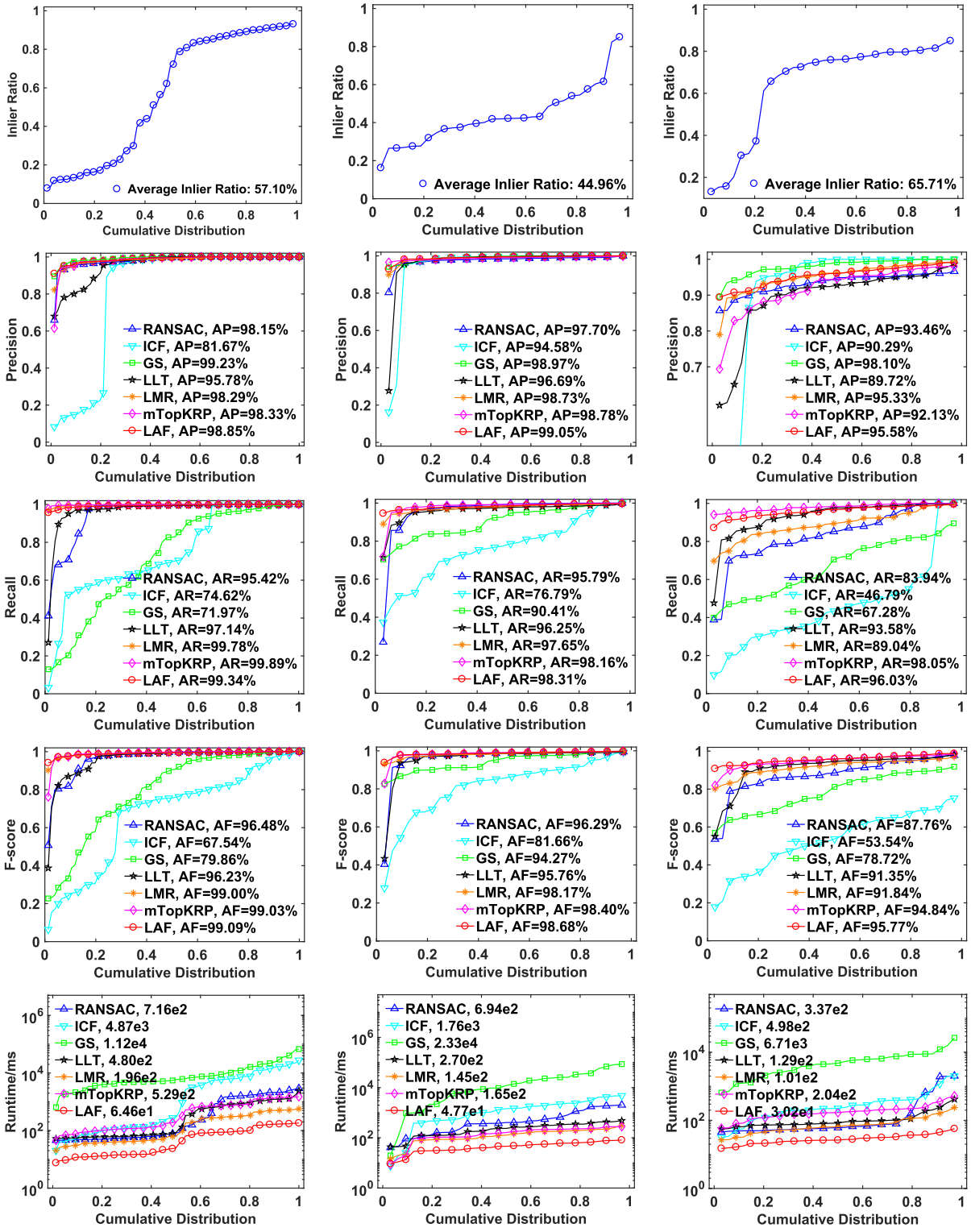

Fig. 4. Quantitative comparisons of RANSAC [57], ICF [60], GS [67], LLT [6], LMR [72], mTopKRP [73], and our LAF on seven image sets that are divided into three groups according to their transformation models. (From left to right) Rigid (SAR, CIAP, GF-II), projective (UAV, PAN), and nonrigid (720Yun, FE). (From top to bottom) Initial inlier ratio, precision, recall, F-score, and runtime 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 that have the performance values (i.e., inlier ratio, precision, recall, F-score, or runtime) no more than y, and the average performance values on the three image groups for each comparing method are shown in the legend accordingly.

a rigid model for fisheye and 720Yun image pairs, ICF fails for low inlier rate in the second image pair, and GS demands a huge computation complexity. As for LLT and mTopKRP, the main area is aligned well but not the marginal one. LMR has built promising F-scores in the mismatching removal task, but the registration results are poor with a strange visual effect. This is because LMR may preserve some bizarre and obvious false matches due to its limited match representations.

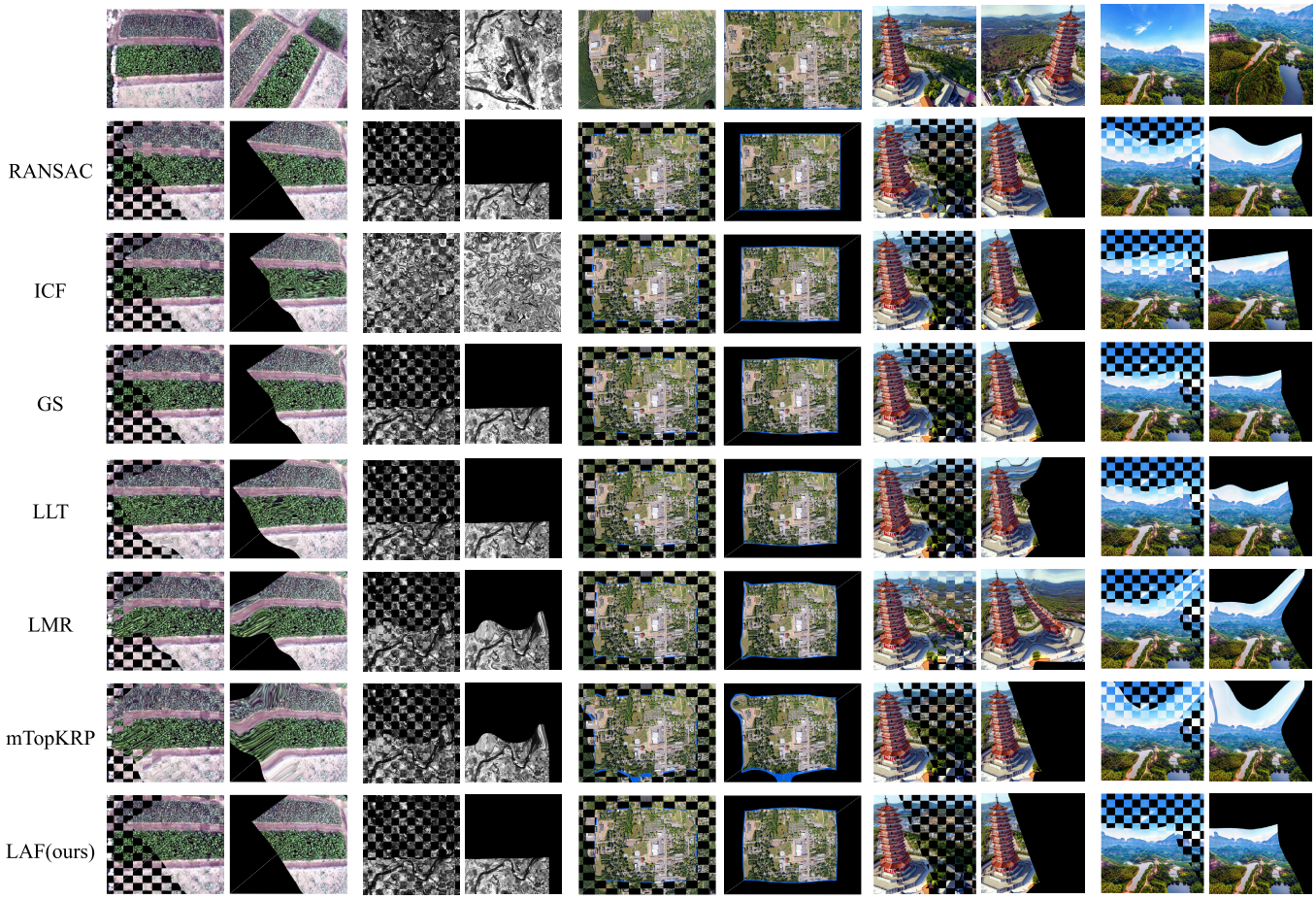

Fig. 5. Qualitative illustration of overall image registration of our LAF and other comparing methods on five representative remote sensing image pairs. First row: original input images, where the left and right in each group are reference and sensed images. Second to the last rows: registration results of all comparing methods, where the left and right in each group are checkboard results and the warped sensed images, respectively.

TABLE I

REGISTRATION RESULTS OF SEVEN COMPARING METHODS ON REMOTE SENSING DATA SETS. THE AVERAGE AND STANDARD DEVIATION OF RMSE, MAE, AND MEE ARE USED FOR EVALUATION, AND THE BEST RESULTS ARE IDENTIFIED WITH BOLD

Method	RMSE	MAE	MEE
RANSAC [57]	8.298 (±31.65)	36.59 (±92.50)	6.846 (±38.37)
ICF [60]	6.464 (±41.07)	28.23 (±101.8)	6.744 (±57.67)
GS [67]	10.72 (±23.61)	70.39 (±121.7)	**2.021 (±11.63)**
LLT [6]	23.53 (±87.35)	66.44 (±182.0)	30.42 (±121.7)
LMR [72]	15.12 (±69.33)	68.51 (±161.8)	16.61 (±96.41)
mTopKRP [73]	6.161 (±37.74)	27.59 (±115.2)	6.715 (±47.87)
LAF (ours)	**4.406 (±23.22)**	**26.09 (±81.29)**	3.339 (±26.95)

In contrast, our proposed LAF can align the overlapping area more accurately, especially the marginal areas.

2) Quantitative Comparison: To evaluate the registration performance in a quantitative way, 78 image pairs, including different types of deformation, are selected from the abovementioned data sets and used to adequately compare the registration performance of these comparing methods. Specifically, the average initial correspondence number of the registration data is 1099.73, with an average inlier rate being 27.23%. In addition, the quantitative experiment is conducted on the selected landmarks $\{r_i^c, s_i^c\}_{i=1}^L$ manually, and performance evaluation is measured by calculating the RMSE, MAE, and MEE of 20 pairs of landmarks that are evenly distributed in easily identifiable locations around the region of interest. The average and standard deviation of RMSE, MAE, and MEE on the 78 selected image pairs are reported in Table I, in which we can find that our LAF achieves the best performance of RMSE and MAE. GS obtains the best MEE performance followed by our LAF, but GS is not robust to address the general registration task due to the worst MAE metric. RANSAC achieves a stable metric measurement because of its global geometrical constraint, but it suffers a lot from the nonrigid deformations. ICF and mTopKRP can obtain competitive performance for the reason that they can preserve reliable feature correspondences that are sufficient to estimate the transformation correctly. As for LLT, the relatively poor registration performance is mainly attributed to the low putative inlier rate. Furthermore, LMR performed not well for the same reason, i.e., some bizarre and obvious false matches may be preserved due to its limited match representation.

V. CONCLUSION

In this study, we propose a new feature matching method based on filtering and denoising theory. In particular, we first divide the putative set into nonoverlapping cells and then calculate the typical motion vector of each cell using the Gaussian kernel convolution operation. Finally, the outliers

are detected by checking the deviations between the putative motion vector and its corresponding typical motion vector. Also, an iterative strategy is proposed to filter out the outliers progressively. In addition, an adaptive parameter setting strategy and posterior probability estimation enable our method to be robust to different data. Our method can converge in a few iterations, and the gridding strategy enables it to achieve linear time complexity. Most importantly, some sparse point-based tasks may inspire from our method when they are achieved by deep learning techniques.

However, the proposed method may largely rely on the local coherency among potential true inliers. If there are only a few true matches but a large number of false matches or the inliers are located extremely separately, the assumption on local consistency would not satisfy, and hence, our LAF may be unworkable. Therefore, in the future research and for different scenarios in remote sensing, we plan to address the problem of feature point detection and description to create more valid feature matches. In addition, a deep convolutional pipeline based on our gridding convolutional strategy would also be studied for better matching and registration performance.

REFERENCES

[1] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, Oct. 2003.

[2] C. Pohl and J. L. Van Genderen, "Review article multisensor image fusion in remote sensing: Concepts, methods and applications," *Int. J. Remote Sens.*, vol. 19, no. 5, pp. 823–854, Jan. 1998.

[3] S. Dawn, V. Saxena, and B. Sharma, "Remote sensing image registration techniques: A survey," in *Proc. Int. Conf. Image Signal Process.*, 2010, pp. 103–112.

[4] 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.

[5] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: A systematic survey," *IEEE Trans. Image Process.*, vol. 14, no. 3, pp. 294–307, Mar. 2005.

[6] 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.

[7] Z. Shao, J. Cai, P. Fu, L. Hu, and T. Liu, "Deep learning-based fusion of Landsat-8 and Sentinel-2 images for a harmonized surface reflectance product," *Remote Sens. Environ.*, vol. 235, Dec. 2019, Art. no. 111425.

[8] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[9] G. Balakrishnan, A. Zhao, M. R. Sabuncu, A. V. Dalca, and J. Guttag, "An unsupervised learning model for deformable medical image registration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9252–9260.

[10] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vols. PAMI–8, no. 6, pp. 679–698, Nov. 1986.

[11] B. Fan, F. Wu, and Z. Hu, "Line matching leveraged by point correspondences," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Feb. 2010, pp. 390–397.

[12] K. Mikolajczyk *et al.*, "A comparison of affine region detectors," *Int. J. Comput. Vis.*, vol. 65, nos. 1–2, pp. 43–72, Nov. 2005.

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

[14] 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.

[15] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*. [Online]. Available: http://arxiv.org/abs/1609.02907

[16] X. Jiang, J. Ma, and J. Chen, "Progressive filtering for feature matching," in *Proc. ICASSP - IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 2217–2221.

[17] 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.

[18] 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.

[19] J. L. Schonberger and J.-M. Frahm, "Structure-from-motion revisited," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4104–4113.

[20] J. Ma, J. Zhao, J. Tian, Z. Tu, and A. L. Yuille, "Robust estimation of nonrigid transformation for point set registration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2147–2154.

[21] J. Maier, M. Humenberger, M. Murschitz, O. Zendel, and M. Vincze, "Guided matching based on statistical optical flow for fast and robust correspondence analysis," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 101–117.

[22] Y. Wu, J. Lim, and M. H. Yang, "Object tracking benchmark," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1834–1848, Sep. 2015.

[23] L. Zheng, Y. Yang, and Q. Tian, "SIFT meets CNN: A decade survey of instance retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 5, pp. 1224–1244, May 2018.

[24] X. Jiang, J. Ma, J. Jiang, and X. Guo, "Robust feature matching using spatial clustering with heavy outliers," *IEEE Trans. Image Process.*, vol. 29, pp. 736–746, 2020.

[25] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," *IEEE Trans. Med. Imag.*, vol. 32, no. 7, pp. 1153–1190, Jul. 2013.

[26] A. P. Tewkesbury, A. J. Comber, N. J. Tate, A. Lamb, and P. F. Fisher, "A critical synthesis of remotely sensed optical image change detection techniques," *Remote Sens. Environ.*, vol. 160, pp. 1–14, Apr. 2015.

[27] J. Le Moigne, N. S. Netanyahu, and R. D. Eastman, *Image Registration for Remote Sensing*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[28] Y. Bentoutou, N. Taleb, K. Kpalma, and J. Ronsin, "An automatic image registration for applications in remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 9, pp. 2127–2137, Sep. 2005.

[29] L. G. Brown, "A survey of image registration techniques," *ACM Comput. Surv.*, vol. 24, no. 4, pp. 325–376, Dec. 1992.

[30] F. P. M. Oliveira and J. M. R. S. Tavares, "Medical image registration: A review," *Comput. Methods Biomech. Biomed. Eng.*, vol. 17, no. 2, pp. 73–93, 2014.

[31] J. Salvi, C. Matabosch, D. Fofi, and J. Forest, "A review of recent range image registration methods with accuracy evaluation," *Image Vis. Comput.*, vol. 25, no. 5, pp. 578–596, May 2007.

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

[33] J. Le Moigne, W. J. Campbell, and R. F. Cromp, "An automated parallel image registration technique based on the correlation of wavelet features," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 8, pp. 1849–1864, Aug. 2002.

[34] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Trans. Image Process.*, vol. 5, no. 8, pp. 1266–1271, Jul. 1996.

[35] H. Liu, B. Guo, and Z. Feng, "Pseudo-log-polar Fourier transform for image registration," *IEEE Signal Process. Lett.*, vol. 13, no. 1, pp. 17–20, Jan. 2006.

[36] Q.-S. Chen, M. Defrise, and F. Deconinck, "Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 12, pp. 1156–1168, Apr. 1994.

[37] K. Takita, T. Aoki, Y. Sasaki, T. Higuchi, and K. Kobayashi, "High-accuracy subpixel image registration based on phase-only correlation," *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, vol. 86, no. 8, pp. 1925–1934, 2003.

[38] H. Foroosh, J. B. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," *IEEE Trans. Image Process.*, vol. 11, no. 3, pp. 188–200, Mar. 2002.

[39] G. Lazaridis and M. Petrou, "Image registration using the Walsh transform," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2343–2357, Aug. 2006.

[40] W.-H. Pan, S.-D. Wei, and S.-H. Lai, "Efficient NCC-based image matching in Walsh-Hadamard domain," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 468–480.

[41] H. S. Stone, M. T. Orchard, E.-C. Chang, and S. A. Martucci, "A fast direct Fourier-based algorithm for subpixel registration of images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 10, pp. 2235–2243, May 2001.

[42] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 187–198, Apr. 1997.

[43] S. Klein, M. Staring, and J. P. W. Pluim, "Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2879–2890, Dec. 2007.

[44] D. Loeckx, P. Slagmolen, F. Maes, D. Vandermeulen, and P. Suetens, "Nonrigid image registration using conditional mutual information," *IEEE Trans. Med. Imag.*, vol. 29, no. 1, pp. 19–29, Jan. 2010.

[45] K. Johnson, A. Cole-Rhodes, I. Zavorin, and J. Le Moigne, "Mutual information as a similarity measure for remote sensing image registration," in *Proc. Geo-Spatial Image Data Exploitation*, 2001, pp. 51–61.

[46] H.-M. Chen, M. K. Arora, and P. K. Varshney, "Mutual information-based image registration for remote sensing data," *Int. J. Remote Sens.*, vol. 24, no. 18, pp. 3701–3706, Jan. 2003.

[47] H.-M. Chen, P. K. Varshney, and M. K. Arora, "Performance of mutual information similarity measure for registration of multitemporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 11, pp. 2445–2454, Nov. 2003.

[48] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: A survey," *Found. Trends Comput. Graph. Vis.*, vol. 3, no. 3, pp. 177–280, 2007.

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

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

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

[52] X. Dai and S. Khorram, "A feature-based image registration algorithm using improved chain-code representation combined with invariant moments," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2351–2362, Jul. 1999.

[53] J. Li, Q. Hu, and M. Ai, "RIFT: Multi-modal image matching based on radiation-invariant feature transform," 2018, *arXiv:1804.09493*. [Online]. Available: http://arxiv.org/abs/1804.09493

[54] F. Dellinger, J. Delon, Y. Gousseau, J. Michel, and F. Tupin, "SAR-SIFT: A SIFT-like algorithm for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 453–466, Jan. 2015.

[55] A. Sedaghat and H. Ebadi, "Remote sensing image matching based on adaptive binning SIFT descriptor," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5283–5293, Oct. 2015.

[56] K. Yang, A. Pan, Y. Yang, S. Zhang, S. Ong, and H. Tang, "Remote sensing image registration using multiple image features," *Remote Sens.*, vol. 9, no. 6, p. 581, Jun. 2017.

[57] M. A. Fischler and R. C. 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, Jun. 1981.

[58] 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.

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

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

[61] 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.

[62] 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.

[63] M. Leordeanu and M. Hebert, "A spectral technique for correspondence problems using pairwise constraints," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 1, Oct. 2005, pp. 1482–1489.

[64] J. Yan, J. Wang, H. Zha, X. Yang, and S. Chu, "Consistency-driven alternating optimization for multigraph matching: A unified approach," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 994–1009, Mar. 2015.

[65] J. Yan, X.-C. Yin, W. Lin, C. Deng, H. Zha, and X. Yang, "A short survey of recent advances in graph matching," in *Proc. ACM Int. Conf. Multimedia Retr. (ICMR)*, 2016, pp. 167–174.

[66] 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.

[67] 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.

[68] 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.

[69] 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. (CVPR)*, Jul. 2017, pp. 2828–2837.

[70] 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. (ICASSP)*, Apr. 2018, pp. 2316–2320.

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

[72] 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.

[73] 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.

[74] 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.

[75] 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.

[76] 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.

[77] 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.

[78] M. Simonovsky, B. Gutiérrez-Becker, D. Mateus, N. Navab, and N. Komodakis, "A deep metric for multimodal registration," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2016, pp. 10–18.

[79] S. Miao, Z. J. Wang, and R. Liao, "A CNN regression approach for real-time 2D/3D registration," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1352–1363, May 2016.

[80] X. Yang, R. Kwitt, M. Styner, and M. Niethammer, "Quicksilver: Fast predictive image registration—A deep learning approach," *NeuroImage*, vol. 158, pp. 378–396, Sep. 2017.

[81] X. Han, T. Leung, Y. Jia, R. Sukthankar, and A. C. Berg, "MatchNet: Unifying feature and metric learning for patch-based matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3279–3286.

[82] K. Lenc and A. Vedaldi, "Learning covariant feature detectors," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 100–117.

[83] X. Zhang, F. X. Yu, S. Kumar, and S.-F. Chang, "Learning spread-out local feature descriptors," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4595–4603.

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

[85] J. L. Schonberger, H. Hardmeier, T. Sattler, and M. Pollefeys, "Comparative evaluation of hand-crafted and learned local features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1482–1491.

[86] S. Wang, D. Quan, X. Liang, M. Ning, Y. Guo, and L. Jiao, "A deep learning framework for remote sensing image registration," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 148–164, Nov. 2018.

[87] Z. Yang, T. Dan, and Y. Yang, "Multi-temporal remote sensing image registration using deep convolutional features," *IEEE Access*, vol. 6, pp. 38544–38555, 2018.

[88] F. Ye, Y. Su, H. Xiao, X. Zhao, and W. Min, "Remote sensing image registration using convolutional neural network features," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 232–236, Feb. 2018.

[89] 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.
[90] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer-Verlag, 2006.
[91] G. Wahba, *Spline Models for Observational Data*, vol. 59. Philadelphia, PA, USA: SIAM, 1990.
[92] A. Vedaldi and B. Fulkerson, "Vlfeat: An open and portable library of computer vision algorithms," in *Proc. Int. Conf. Multimedia*, 2010, pp. 1469–1472.
[93] L. Liang *et al.*, "Image registration using two-layer cascade reciprocal pipeline and context-aware dissimilarity measure," *Neurocomputing*, vol. 371, pp. 1–14, Jan. 2020.

Xingyu Jiang (Member, IEEE) received the B.E. degree from the Department of Mechanical and Electronic Engineering, Huazhong Agricultural University, Wuhan, China, in 2017, and the M.S. degree from the Electronic Information School, Wuhan University, Wuhan, in 2019, where he is pursuing the Ph.D. degree with the Electronic Information School.

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

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

From 2012 to 2013, he was an Exchange Student with the Department of Statistics, University of California at Los Angeles, Los Angeles, CA, USA. He is a Professor with the Electronic Information School, Wuhan University, Wuhan. He has authored or coauthored over 130 refereed journal articles and conference papers, including the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE/TRANSACTIONS ON IMAGE PROCESSING/TRANSACTIONS ON SIGNAL PROCESSING, the *International Journal of Computer Vision*, the IEEE Conference on Computer Vision and Pattern Recognition, and the IEEE International Conference on Computer Vision. His research interests include the areas of computer vision, machine learning, and pattern recognition.

Dr. Ma has won the Natural Science Award of Hubei Province (first class) as the first author. He has received the Chinese Association for Artificial Intelligence (CAAI) Excellent Doctoral Dissertation Award (a total of eight winners in China) and the Chinese Association of Automation (CAA) Excellent Doctoral Dissertation Award (a total of ten winners in China). He has been identified in the 2019 Highly Cited Researchers List from the Web of Science Group. He is an Area Editor of *Information Fusion*, an Associate Editor of *Neurocomputing* and IEEE ACCESS, and a Guest Editor of *Remote Sensing*.

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

His research interests include computer vision and pattern recognition.

Haiping Xu received the B.S. degree in information and computing science and the Ph.D. degree in applied mathematics from Fuzhou University, Fuzhou, China, in 2011 and 2018, respectively.

She is a Lecturer with the College of Mathematics and Data Science, Minjiang University, Fuzhou. Her research interests include computer vision and machine learning.

Geng Lin received the B.S. degree in information and computing sciences, the M.S. degree in operations research and cybernetics, and the Ph.D. degree in applied mathematics from Fuzhou University, Fuzhou, China, in 2004, 2007, and 2010, respectively.

He is a Professor with the College of Mathematics and Data Science, Minjiang University, Fuzhou. He is also the author or coauthor of scientific articles in refereed journals, including the IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, the IEEE TRANSACTIONS ON COMPUTERS, the *Journal of Global Optimization*, *Computers & Operations Research*, *Information Sciences*, *Annals of Operations Research*, and *INFORMS Journal on Computing*. His research interests include combinatorial optimization and artificial intelligence.

Tao Lu (Member, IEEE) received the B.S. and M.S. degrees from the School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan, China, in 2003 and 2008, respectively, and the Ph.D. degree from the National Engineering Research Center for Multimedia Software, Wuhan University, Wuhan, in 2013.

He is an Associate Professor with the School of Computer Science and Engineering, Wuhan Institute of Technology, and a Research Member with the Hubei Provincial Key Laboratory of Intelligent Robot. He held a post-doctoral position at the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA, from 2015 to 2017. His research interests include image/video processing, computer vision, and artificial intelligence.

Xin Tian (Member, IEEE) received the B.S. degree from the Department of Electronic Science and Technology, Huazhong University of Science and Technology, Wuhan, China, in 2004, and the Ph.D. degree from the Institute for Pattern Recognition and Artificial Intelligence, Huazhong University of Science and Technology, in 2010.

From 2015 to 2016, he was a Visiting Faculty Member at the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA. He is an Associate Professor with the School of Electronic Information and Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan. He has authored or coauthored more than 60 scientific articles, and he holds more than ten Chinese patents. His research interests include dictionary learning, sparse coding, computational imaging, and image compression.

Dr. Tian's awards include the First Prize for Scientific and Technological Progress in Surveying and Mapping in China, the First Prize for Excellent Achievements of Information Technology in Electric Power Industry of China, and so on.