Robust Feature Matching for Remote Sensing Image Registration via Locally Linear Transforming

Jiayi Ma, Huabing Zhou, Ji Zhao, Yuan Gao, Junjun Jiang, and Jinwen Tian

***Abstract*—Feature matching, which refers to establishing reliable correspondence between two sets of features (particularly point features), is a critical prerequisite in feature-based registration. In this paper, we propose a flexible and general algorithm, which is called locally linear transforming (LLT), for both rigid and nonrigid feature matching of remote sensing images. We start by creating a set of putative correspondences based on the feature similarity and then focus on removing outliers from the putative set and estimating the transformation as well. We formulate this as a maximum-likelihood estimation of a Bayesian model with hidden/latent variables indicating whether matches in the putative set are outliers or inliers. To ensure the well-posedness of the problem, we develop a local geometrical constraint that can preserve local structures among neighboring feature points, and it is also robust to a large number of outliers. The problem is solved by using the expectation–maximization algorithm (EM), and the closed-form solutions of both rigid and nonrigid transformations are derived in the maximization step. In the nonrigid case, we model the transformation between images in a reproducing kernel Hilbert space (RKHS), and a sparse approximation is applied to the transformation that reduces the method computation complexity to linearithmic. Extensive experiments on real remote sensing images demonstrate accurate results of LLT, which outperforms current state-of-the-art methods, particularly in the case of severe outliers (even up to 80%).**

***Index Terms*—Feature matching, locally linear transforming (LLT), outlier, registration, remote sensing.**

I. Introduction

IMAGE registration is a fundamental and challenging problem in remote sensing, and it is a critical prerequisite in a wide range of applications including environment monitoring, change detection, image fusion, image mosaic, and map updating [1], [2]. The primary objective of image registration is to geometrically overlay two images of the same scene (i.e., the reference and sensed images) taken at different times, from different viewpoints, or by different sensors.

During the last decades, a variety of techniques have been developed for remote sensing image registration. These methods can be roughly classified into two categories: area based and feature based [1]. The former finds the matching information by using the original pixel intensities in the overlapped region of two images with a specified similarity metric, whereas the latter seeks correspondence between local features under descriptor similarity and/or spatial geometrical constraints. The area-based methods are preferable in case of few prominent details where the distinctive information is provided by pixel intensities rather than by local shapes and structures, but they suffer from heavy computational complexities, image distortions, and illumination changes. By contrast, feature-based methods are more robust, which allow registering images of completely different nature and can handle complex image distortions. The features used in these methods can be represented as compact geometrical entities at different levels, such as points, line segments, contours, and regions [3], [4]. Nevertheless, feature-based methods are typically formulated as a point matching problem as point representations are general and easy to extract. In this paper, we focus on point-based methods for registration of remote sensing images.

The key requirement of point-based method is feature matching, which refers to the process of establishing reliable correspondence between two sets of feature points (also known as key points or interest points). Although many methods have been proposed for different applications, it is still a challenging task to develop a unified framework for remote sensing image registration. First, remote sensing images often contain local distortions caused by ground relief variations and imaging viewpoint changes, which means that they are not "*exactly matchable*" via a simple parametric model (e.g., rigid or affine transformation) as used in most existing point-based methods. Therefore, high-dimensional nonrigid transformations are required to produce accurate alignments. Second, the complex nature of remote sensing images (e.g., unavoidable noise, occlusions, repeated structures, and multisensor data) often results in a high number of false matches, which have a significant impact on determining the transformational model required in aligning the images together. Therefore, a robust procedure of outlier removal is desirable. Third, for large remote sensing images, the scale of extracted feature points is usually extremely large, e.g., tens of thousands. This poses a significant burden on typical feature matching methods, particularly in the nonrigid

Manuscript received March 24, 2015; revised April 28, 2015; accepted May 23, 2015. This work was supported in part by the National Natural Science Foundation of China under Grant 61503288, in part by the China Postdoctoral Science Foundation under Grant 2015M570665, and in part by the Natural Science Fund of Hubei Province under Grant 2014CFB268 and Grant 2015CFB371. *(Corresponding author: Huabing Zhou.)*

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

H. Zhou is with the Hubei Provincial Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan 430073, China (e-mail: zhouhuabing@gmail.com).

J. Zhao is with the SAIT China Laboratory, Samsung Research Center, Beijing 100028, China (e-mail: zhaoji84@gmail.com).

Y. Gao is with the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong (e-mail: Ethan.Y.Gao@my.cityu.edu.hk).

J. Jiang is with the School of Computer Science, China University of Geosciences, Wuhan 430074, China (e-mail: junjun0595@163.com).

J. Tian is with the School of Automation, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: jwtian@hust.edu.cn).

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

Digital Object Identifier 10.1109/TGRS.2015.2441954

case. Therefore, it is of particular advantage to develop a more efficient technique.

To address these issues, we propose a novel method called *locally linear transforming* (LLT) in this paper. The method is general and efficient, which can handle both rigid and nonrigid transformations within linearithmic complexity, and it is also robust to a very large number of outliers. More precisely, we introduce a unified maximum-likelihood framework for robust estimation of transformation from a set of putative correspondences contaminated by noise and outliers. Our approach associates each correspondence with a latent variable, which indicates whether it is an inlier or not, and then alternatively recovers the underlying transformation (either rigid or nonrigid) and estimates the inlier set by using an expectation–maximization (EM) algorithm [5]. To ensure the well-posedness[1] of the optimization problem, we develop a local geometrical constraint that is similar to the locally linear embedding (LLE) [6]. The local geometrical constraint can also preserve local structures among neighboring feature points after the transformation and hence helps to recover the accurate correspondence. Moreover, in the nonrigid case, we model the transformation in a functional space, called the reproducing kernel Hilbert space (RKHS) [7], in which the function has an explicit kernel representation.

Our contribution in this paper includes the following three aspects. First, we propose a unified maximum-likelihood formulation for robust feature matching of remote sensing images. Compared with the existing methods that typically rely on a parametric transformational model, our formulation can also handle nonparametric model such as nonrigid transformation. Second, we develop a local geometrical constraint to regularize the transformation. This constraint can preserve local structures among neighboring feature points, which is an intrinsic property of remote sensing image pairs, and hence can handle a large number of outliers. Third, we introduce a fast implementation based on sparse approximation to improve the computational efficiency. This makes our method to have linearithmic time complexity and linear space complexity with respect to the scale of feature correspondence, which is significant for a large size of remote sensing images.

The remainder of this paper is organized as follows. Section II describes background material and related work. In Section III, we present our robust LLT algorithm and apply it to rigid, affine,[2] and nonrigid feature matching. Section IV illustrates the registration performance of our method on various types of remote sensing image pairs with comparisons to other approaches, followed by some concluding remarks in Section V.

II. Related Work

Image registration has been widely applied in many fields such as computer vision [8]–[11], pattern recognition [12]–[15], image analysis [16]–[20], and particularly in the field of remote sensing [1], [21]–[24]. Exhaustive reviews on image registration methods can be found in the literature [1], [3], [22], [25]. Then, we briefly review the two major types of methods, i.e., the area-based methods and feature-based methods, particularly in the context of remote sensing image registration.

A. Area-Based Methods

Area-based methods deal directly with the image intensity values without attempting to detect salient structures such as features. These methods can be broadly classified into three types: correlation-like methods, Fourier methods, and mutual information (MI) methods [1].

Correlation-like methods such as cross-correlation and its modifications are a classical representative of the area-based methods [26]. The main idea of these methods is to compute the similarities of window pairs in two images, and consider the one with the largest similarity as a correspondence. In remote sensing applications, a correlation-like method utilizing maxima of wavelet coefficients has been developed for automatic registration [27]. The correlation-like methods suffer from some drawbacks such as the flatness of the similarity measure in textureless regions and high computational complexity. However, due to their easy hardware implementation, which is beneficial for real-time applications, these methods are still often in use.

Fourier methods exploit the Fourier representation of images in the frequency domain [28]. A common technique is the phase correlation method based on the Fourier shift theorem, which was later extended to account for rotation and scaling [29]. The applications to remote sensing are described in [30]. Compared with correlation-like methods, these methods have some advantages in computational efficiency and are also robust to frequency-dependent noise. However, they have some limitations in case of image pairs with significantly different spectral contents.

Finally, area-based methods also include MI methods. The MI provides an attractive metric for maximizing the dependence between two images, and it is particularly suitable for multimodal registration in remote sensing, which often makes use of the exploitation of multisensors [31], [32]. Moreover, for medical image registration, the MI-based methods not only work directly with image intensities but also with extracted features such as points of the area borders [33]. Despite its outstanding performance, the MI-based methods do not provide a global maximum of the entire search space for the transformation and hence inevitably reduces its robustness.

B. Feature-Based Methods

The second approach for image registration is based on the extraction of salient structures, i.e., features, in the images. We assume that the extracted features are represented by spatial points, which are also called *control points* in the literature [1]. The image registration then reduces to a feature matching problem, where the goal is to determine the correct correspondence

[1]A problem is well-posed if its solution exists, is unique, and depends continuously on the data (e.g., it is stable).

[2]Note that affine is also a kind of simple nonrigid transformation that allows only for translation, rotation, anisotropic scaling, and skews. Nonrigid in this paper typically refers to a nonlinear transformation.

and/or to find the underlying spatial transformation between two sets of control points [34].

A popular strategy for solving the feature matching problem is to use a two-stage process [34]–[36]. In the first stage, a set of *putative correspondences* are computed by using a similarity constraint, which requires that points can only match points with similar local descriptors (e.g., scale-invariant Fourier transform (SIFT) [37] or shape context [38]). This putative correspondence set typically includes not only most of the true matches, i.e., the inliers, but also a large number of false matches, or outliers, due to ambiguities in the similarity constraint. The second stage is designed to remove the outliers by using a geometrical constraint, which requires that the matches satisfy an underlying geometrical requirement. The inliers and the geometric parameters of the transformation are then obtained accordingly. Examples of this strategy include the classical RANSAC algorithm [39] and analogous robust hypothesize-and-verify methods [40], [41] that typically rely on parametric models, and vector field consensus [34], [42], robust point matching via L_2E [43], graph shift (GS) [44], and identifying correspondence function (ICF) [45], which are based on nonparametric models. In remote sensing, the majority of the feature-based methods also use the two-stage process. In general, the feature descriptors used in remote sensing images are not only SIFT [37] and its variants [46]–[48] but also include SURF [49] and complex wavelet feature [50]. The geometric transformational models are not only typically parametric, such as rigid, affine [51], [52], and projective [2], but also include nonparametric models, such as a nonrigid model with thin-plate spline (TPS) [21], and a graph-based model, such as restricted spatial order constraints [53]. Although these methods are very successful in many situations, none of them provide a uniform framework to handle both the rigid and nonrigid matching problems under severe outliers.

Another strategy for feature matching is to formulate this problem in terms of a correspondence matrix between control points together with a parametric or nonparametric geometrical constraint. These methods typically involve a two-step update process, which alternates between the correspondence and the transformation estimation. The well-known heuristic iterative closest point algorithm is one such example [54], [55]. Chui and Rangarajan [56] established a general framework for estimating correspondence and transformation, where in a nonrigid case, the transformation is modeled as a TPS function. Alternatively, the coherence point drift (CPD) algorithm [57] uses Gaussian radial basis functions instead of TPS, and it was later improved by using global–local topology constraints [58], [59]. In these formulations, both the rigid and nonrigid cases can be dealt with, but these methods usually throw out the similarity information (e.g., a descriptor) and use only the spatial position of each feature point; hence, the matching performance may probably be degraded. In remote sensing, a method combining spatial and similarity information has also been introduced [51], but it can only be applied to a parametric model such as an affine model. In this paper, we introduce a uniform maximum-likelihood formulation for both rigid and nonrigid feature matching. By combining it with a local geometrical constraint, the proposed approach is also able to handle a large number of outliers.

III. Locally Linear Transforming Algorithm

This section describes the proposed feature matching algorithm for remote sensing images. We start by laying out a maximum-likelihood formulation for feature matching and then introduce a local geometrical constraint and propose our LLT algorithm in the context of remote sensing. We subsequently apply the proposed approach to rigid, affine, and nonrigid feature matching. Finally, we analyze the computational complexity and provide the implementation details of the proposed approach.

A. Problem Formulation

Suppose we are given a set of N putative feature correspondences $S = \{(\mathbf{x}_n, \mathbf{y}_n)\}_{n=1}^{N}$ extracted from a remote sensing image pair, where $\mathbf{x}_n$ and $\mathbf{y}_n$ are 2-D column vectors indicating the spatial positions of feature points in the two 2-D images (adaptation to higher dimension is straightforward). Typically, S is contaminated by some unknown noise and outliers, and the goal is to distinguish inliers from the outliers to establish reliable correspondences.

Without loss of generality, we make the assumption that the noise on inliers is isotropic Gaussian with zero mean and covariance $\sigma^2\mathbf{I}$, where $\mathbf{I}$ is an identity matrix, and the outlier distribution is uniform [40]. A more general assumption of anisotropic Gaussian noise on inliers has also been considered in [60]. We then associate the nth correspondence with a latent variable $z_n \in \{0, 1\}$, where $z_n = 1$ indicates the correspondence $(\mathbf{x}_n, \mathbf{y}_n)$ being an inlier and $z_n = 0$ points to an outlier. Let $\mathcal{T}$ be the transformation that characterizes the underlying geometrical relation between the correspondences, e.g., for an inlier correspondence $(\mathbf{x}_n, \mathbf{y}_n)$, $\mathbf{y}_n = \mathcal{T}(\mathbf{x}_n)$. Thus, the mixture model takes the following form:

$$\begin{aligned} p(\mathbf{y}_n|\mathbf{x}_n, \boldsymbol{\theta}) &= \sum_{z_n} p(\mathbf{y}_n, z_n|\mathbf{x}_n, \boldsymbol{\theta}) \\ &= p(z = 1)p(\mathbf{y}_n|\mathbf{x}_n, \boldsymbol{\theta}, z_n = 1) \\ &\quad + p(z = 0)p(\mathbf{y}_n|\mathbf{x}_n, \boldsymbol{\theta}, z_n = 0) \\ &= \frac{\gamma}{2\pi\sigma^2} e^{-\frac{\|\mathbf{y}_n - \mathcal{T}(\mathbf{x}_n)\|^2}{2\sigma^2}} + \frac{1-\gamma}{a} \end{aligned} \tag{1}$$

where $\boldsymbol{\theta} = \{\mathcal{T}, \sigma^2, \gamma\}$ includes a set of unknown parameters, γ is the mixing coefficient specifying the marginal distribution over the latent variable (i.e., $\forall\, z_n$, $p(z_n = 1) = \gamma$), and $1/a$ denotes the outlier uniform distribution with a being the area of the second image (i.e., the range of $\mathbf{y}_n$). 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 $N \times 2$ matrices indicating the two feature point sets, respectively. By using the independent and identically distributed (i.i.d.) data assumption, we have the likelihood function $p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}) = \prod_{n=1}^{N} p(\mathbf{y}_n|\mathbf{x}_n, \boldsymbol{\theta})$.

We give a maximum-likelihood estimation of the parameter set $\boldsymbol{\theta}$, i.e., $\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta})$, which is equivalent to minimizing the negative log-likelihood function:

$$E(\boldsymbol{\theta}) = -\ln p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}) = -\sum_{n=1}^{N} \ln p(\mathbf{y}_n|\mathbf{x}_n, \boldsymbol{\theta}). \tag{2}$$

To solve this problem, we consider the EM algorithm. It alternates with two steps: the E-step basically estimates the responsibility that indicates to what degree a correspondence belonging to the inlier set under the given parameter set $\boldsymbol{\theta}$, whereas the M-step updates $\boldsymbol{\theta}$ based on the current estimate of the responsibility. We follow standard notations [61] and omit some terms that are independent of $\boldsymbol{\theta}$. Considering the negative log-likelihood function in (2), the complete-data log likelihood is then given by

$$\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) = -\frac{1}{2\sigma^2}\sum_{n=1}^{N} p_n \left\|\mathbf{y}_n - \mathcal{T}(\mathbf{x}_n)\right\|^2 - \ln\sigma^2 \sum_{n=1}^{N} p_n + \ln\gamma \sum_{n=1}^{N} p_n + \ln(1-\gamma)\sum_{n=1}^{N}(1-p_n) \quad (3)$$

where $p_n = P(z_n = 1|\mathbf{x}_n, \mathbf{y}_n, \boldsymbol{\theta}^{\text{old}})$ is a posterior probability indicating to what degree $(\mathbf{x}_n, \mathbf{y}_n)$ being an inlier.

E-Step: Denote $\mathbf{P} = \text{diag}(p_1, \ldots, p_N)$ as a diagonal matrix, where p_n can be computed by using the current parameter set $\boldsymbol{\theta}^{\text{old}}$ based on Bayes rule, i.e.,

$$p_n = \frac{\gamma e^{-\frac{\|\mathbf{y}_n - \mathcal{T}(\mathbf{x}_n)\|^2}{2\sigma^2}}}{\gamma e^{-\frac{\|\mathbf{y}_n - \mathcal{T}(\mathbf{x}_n)\|^2}{2\sigma^2}} + \frac{2\pi\sigma^2(1-\gamma)}{a}}. \quad (4)$$

M-Step: Reestimate the parameter set using the current responsibilities: $\boldsymbol{\theta}^{\text{new}} = \arg\max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$. Taking derivatives of $\mathcal{Q}$ with respect to σ^2 and γ, and setting them to zero, we obtain

$$\sigma^2 = \frac{\text{tr}\left((\mathbf{Y}-\mathbf{T})^{\text{T}}\mathbf{P}(\mathbf{Y}-\mathbf{T})\right)}{2\cdot\text{tr}(\mathbf{P})} \quad (5)$$

$$\gamma = \frac{\text{tr}(\mathbf{P})}{N} \quad (6)$$

where $\mathbf{T} = (\mathcal{T}(\mathbf{x}_1), \ldots, \mathcal{T}(\mathbf{x}_N))^{\text{T}}$. In order to complete the EM algorithm, the transformation $\mathcal{T}$ should be estimated in the M-step. We will discuss it later in the succeeding sections.

After the EM iteration converges, with a predefined threshold τ, the inlier set $\mathcal{I}$ could be obtained according to the following criterion:

$$\mathcal{I} = \{(\mathbf{x}_n, \mathbf{y}_n) : p_n > \tau, n \in \mathbb{N}_N\}. \quad (7)$$

B. Local Geometrical Constraint

The transformation $\mathcal{T}$ characterizes the global geometrical relation between the image pair, which is useful in keeping the overall spatial connectivity of the point correspondences during matching. However, for remote sensing image pairs, the difference between the disparities of the point correspondences in local areas are typically quite small; hence, the local structures among neighboring feature points are also very strong and stable. This is particularly beneficial when the images involve nonrigid or discontinuous motions [62]. Therefore, to establish accurate matches, a local geometrical constraint on the point correspondences is desired.

In our problem, we hope that the local structures in the point set could be preserved after the transformation $\mathcal{T}$. To this end, we introduce an efficient scheme similar to the LLE algorithm [6], [58], which is proposed as a nonlinear dimensionality reduction method to preserve the local neighborhood structure in a low-dimensional manifold. First, search the K nearest neighbors for each point in $\mathbf{X}$. Denote by $\mathbf{W}$ an $N \times N$ weight matrix, and enforce $\mathbf{W}_{ij} = 0$ if $\mathbf{x}_j$ does not belong to the set of neighbors of $\mathbf{x}_i$. Second, minimize the reconstruction errors measured by the cost function as follows:

$$\mathcal{E}(\mathbf{W}) = \sum_{i=1}^{N}\left\|\mathbf{x}_i - \sum_{j=1}^{N}\mathbf{W}_{ij}\mathbf{x}_j\right\|^2 \quad (8)$$

under a constraint that the rows of the weight matrix sum to one: $\forall i, \sum_{j=1}^{N}\mathbf{W}_{ij} = 1$. The optimal weights $\mathbf{W}_{ij}$ can be obtained by solving a least squares problem. Third, the local geometry of each inlier point after the transformation $\mathcal{T}$ could be preserved by minimizing a *transforming cost term* $\sum_{i=1}^{N} p_i\|\mathcal{T}(\mathbf{x}_i) - \sum_{j=1}^{N}\mathbf{W}_{ij}\mathcal{T}(\mathbf{x}_j)\|^2$. By combining it with (3), the objective function in the M-step then becomes

$$\widehat{\mathcal{Q}}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) = \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) - \lambda\sum_{i=1}^{N} p_i\left\|\mathcal{T}(\mathbf{x}_i) - \sum_{j=1}^{N}\mathbf{W}_{ij}\mathcal{T}(\mathbf{x}_j)\right\|^2 \quad (9)$$

where $\lambda > 0$ controls the tradeoff between the two terms. Clearly, the estimates of σ^2 and γ in (5) and (6) will not be influenced by the additional transforming cost term. To estimate $\mathcal{T}$, we consider the related terms in (9) and obtain the following minimizing problem:

$$\Psi(\mathcal{T}) = \frac{1}{2\sigma^2}\sum_{n=1}^{N} p_n\left\|\mathbf{y}_n - \mathcal{T}(\mathbf{x}_n)\right\|^2 + \lambda\sum_{i=1}^{N} p_i\left\|\mathcal{T}(\mathbf{x}_i) - \sum_{j=1}^{N}\mathbf{W}_{ij}\mathcal{T}(\mathbf{x}_j)\right\|^2 \quad (10)$$

which is composed of an empirical error term and a transforming cost term.

Note that in the linear case such as rigid or affine transformation, by using the homogeneous coordinate notation $\bar{\mathbf{x}} = (\mathbf{x}^{\text{T}}, 1)^{\text{T}}$, the transformation $\mathcal{T}$ can be represented as a 3×3 matrix $\mathbf{H} : \mathbf{y}_n = \mathbf{H}\mathbf{x}_n$. Thus, the transforming cost term in (10) has the form $\lambda\|\mathbf{P}^{1/2}(\bar{\mathbf{X}} - \mathbf{W}\bar{\mathbf{X}})\mathbf{H}^{\text{T}}\|_F^2$ with $\|\cdot\|_F$ being the Frobenius norm. As $\mathbf{P}$ and $\mathbf{W}$ are fixed during transformation estimation, the transforming cost term then becomes $\lambda\|\mathbf{G}\mathbf{H}^{\text{T}}\|_F^2$, where $\mathbf{G} = \mathbf{P}^{1/2}(\bar{\mathbf{X}} - \mathbf{W}\bar{\mathbf{X}})$ is an $N \times 3$ constant matrix. In this context, the transforming cost term plays a role of regularization on the transformation $\mathcal{T}$, which ensures the well-posedness of the empirical risk minimization [63], and it is controlled by a locally linear constraint (i.e., $\bar{\mathbf{X}} - \mathbf{W}\bar{\mathbf{X}}$) and a match correctness constraint (i.e., $\mathbf{P}$). This idea can be generalized to the nonlinear case since we expect each feature point and its neighbors to lie on or close to a locally

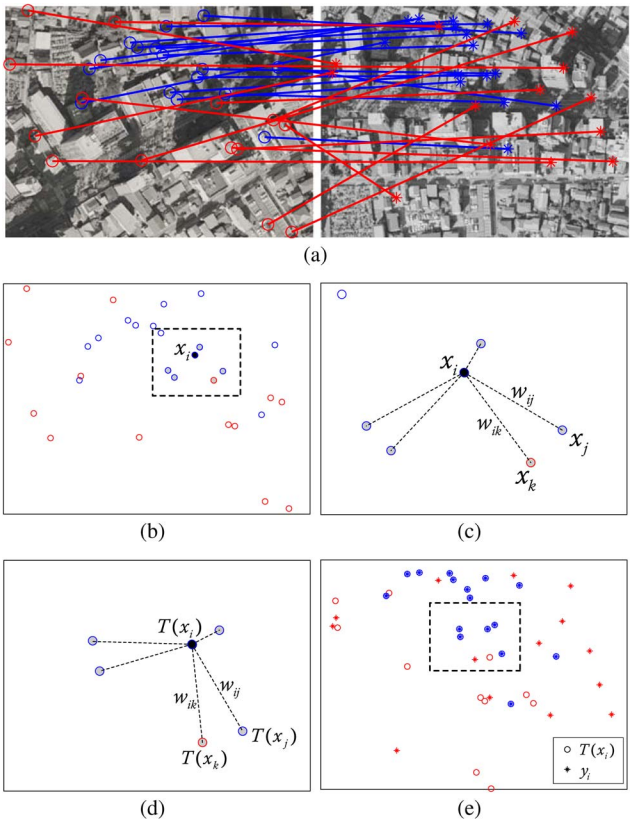


Fig. 1. Schematic of the local geometrical constraint. (a) Set of putative correspondences, where $\circ$ and $*$ are the feature points of the left and right images, and blue and red lines represent inliers and outliers, respectively. (b) Assign neighbors to each feature point $\mathbf{x}_i$, e.g., the five solid circles around $\mathbf{x}_i$. (c) Compute the weights $\mathbf{W}_{ij}$ that best linearly reconstruct $\mathbf{x}_i$ from its neighbors. (d) Optimize the transformation $\mathcal{T}$ under the constraint that each point $\mathbf{x}_i$ can be reconstructed by its neighbors with weights $\mathbf{W}_{ij}$ after the transformation. (e) Alignment of two point sets by using transformation $\mathcal{T}$.

linear patch.[3] We give a schematic illustration of the local geometrical constraint in Fig. 1. As our method recovers global transformation based on locally linear fits, we name it LLT.

We next consider the modeling of the transformation. In remote sensing tasks such as image mosaicing, environmental monitoring, and change detection, the relationships between image pairs are typically modeled by rigid or affine transformations [2], [46]. This is appropriate since remote sensing images are often taken at very long range, and then they could be approximately considered planar scenes. However, when the scene involves land drift or feature points with significant depth relief such as spatially variant terrain relief images, the planar scene assumption then cannot work well. In this case, a relatively complex nonrigid model is more preferable. Our proposed formulation in this paper is independent of the transformation model, and it is able to handle most common geometric distortions found in remotely sensed imagery. To solve the transformation $\mathcal{T}$ from (10), we specify it for rigid, affine, and nonrigid cases, separately.

[3]The locally linear assumption in the nonrigid case may fail when the neighbors cover area where the geometrical transformation nonlinearity cannot be neglected. However, the assumption works quite well as the feature points are typically in large scale and spread over the image.

C. Rigid Feature Matching

For rigid feature matching, we define the transformation as $\mathcal{T}(\mathbf{x}_n) = s\mathbf{R}\mathbf{x}_n + \mathbf{t}$, where $\mathbf{R}$ is a 2×2 rotation matrix, $\mathbf{t}$ is a 2×1 translation vector, and s is a scaling parameter. By considering that $\mathbf{R}$ is orthogonal and the constraint $\forall i$, $\sum_{j=1}^{N} \mathbf{W}_{ij} = 1$, the objective function (10) becomes

$$\Psi(\mathbf{R}, \mathbf{t}, s) = \frac{1}{2\sigma^2} \sum_{n=1}^{N} p_n \|\mathbf{y}_n - s\mathbf{R}\mathbf{x}_n - \mathbf{t}\|^2 + \lambda \sum_{i=1}^{N} p_i \left\| s\left(\mathbf{x}_i - \sum_{j=1}^{N} \mathbf{W}_{ij}\mathbf{x}_j\right)\right\|^2$$
$$\text{s.t. } \mathbf{R}^{\mathrm{T}}\mathbf{R} = \mathbf{I},\ \det(\mathbf{R}) = 1. \tag{11}$$

Note that the first term is similar to the absolute orientation problem [57], [64]. The solutions of $\mathbf{t}$ and s are straightforward, whereas the solution of $\mathbf{R}$ is complicated due to the additional constraints. To obtain the closed-form solution, we consider the following lemma [65].

Lemma 1: Let $\mathbf{R}$ be an unknown $D \times D$ rotation matrix and $\mathbf{B}$ be a known $D \times D$ real square matrix. Let $\mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}}$ be a singular value decomposition of $\mathbf{B}$, where $\mathbf{U}\mathbf{U}^{\mathrm{T}} = \mathbf{V}\mathbf{V}^{\mathrm{T}} = \mathbf{I}$ and $\mathbf{S} = \mathrm{diag}(s_i)$ with $s_1 \geq \cdots \geq s_D \geq 0$. Then, the optimal rotation matrix $\mathbf{R}$ that maximizes $\mathrm{tr}(\mathbf{B}^{\mathrm{T}}\mathbf{R})$ is $\mathbf{R} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}}$, where $\mathbf{D} = \mathrm{diag}(1, \ldots, 1, \det(\mathbf{U}\mathbf{V}^{\mathrm{T}}))$.

To solve the rotation matrix $\mathbf{R}$, we rewrite the objective function (11) so that it has the form $\mathrm{tr}(\mathbf{B}^{\mathrm{T}}\mathbf{R})$. To this end, we first eliminate the translation parameter $\mathbf{t}$. By taking the derivative of Ψ with respect to $\mathbf{t}$ and setting it to zero, we obtain

$$\mathbf{t} = \frac{1}{\mathrm{tr}(\mathbf{P})}\mathbf{Y}^{\mathrm{T}}\mathbf{P}\mathbf{1} - \frac{1}{\mathrm{tr}(\mathbf{P})}s\mathbf{R}\mathbf{X}^{\mathrm{T}}\mathbf{P}\mathbf{1} = \mu_{\mathbf{y}} - s\mathbf{R}\mu_{\mathbf{x}} \tag{12}$$

where $\mu_{\mathbf{x}}$ and $\mu_{\mathbf{y}}$ are the mean vectors defined as

$$\mu_{\mathbf{x}} = \frac{1}{\mathrm{tr}(\mathbf{P})}\mathbf{X}^{\mathrm{T}}\mathbf{P}\mathbf{1} \quad \mu_{\mathbf{y}} = \frac{1}{\mathrm{tr}(\mathbf{P})}\mathbf{Y}^{\mathrm{T}}\mathbf{P}\mathbf{1}. \tag{13}$$

By substituting $\mathbf{t}$ back into the objective function and omitting the terms that are independent of $\mathbf{R}$ and s, we obtain

$$\Psi(\mathbf{R}, s) = \frac{1}{2\sigma^2}\mathrm{tr}(s^2\hat{\mathbf{X}}^{\mathrm{T}}\mathbf{P}\hat{\mathbf{X}} - 2s\hat{\mathbf{Y}}^{\mathrm{T}}\mathbf{P}\hat{\mathbf{X}}\mathbf{R}^{\mathrm{T}}) + \lambda \cdot \mathrm{tr}(s^2\mathbf{X}^{\mathrm{T}}\mathbf{Q}\mathbf{X}) \tag{14}$$

where $\hat{\mathbf{X}} = \mathbf{X} - \mathbf{1}\mu_{\mathbf{x}}^{\mathrm{T}}$ and $\hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{1}\mu_{\mathbf{y}}^{\mathrm{T}}$ are centered point matrices, amd $\mathbf{Q} = (\mathbf{I} - \mathbf{W})^{\mathrm{T}}\mathbf{P}(\mathbf{I} - \mathbf{W})$. Specifically, we consider the term related to $\mathbf{R}$, which has the following form:

$$\Psi(\mathbf{R}) = -\frac{s}{\sigma^2}\mathrm{tr}\left((\hat{\mathbf{Y}}^{\mathrm{T}}\mathbf{P}\hat{\mathbf{X}})^{\mathrm{T}}\mathbf{R}\right). \tag{15}$$

Therefore, by applying Lemma 1, the optimal $\mathbf{R}$ of the problem in (11) is given by

$$\mathbf{R} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}} \tag{16}$$

where $\mathbf{U}$ and $\mathbf{V}$ can be obtained from $\mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}} = \mathrm{svd}(\hat{\mathbf{Y}}^{\mathrm{T}}\mathbf{P}\hat{\mathbf{X}})$, and $\mathbf{D} = \mathrm{diag}(1, \det(\mathbf{U}\mathbf{V}^{\mathrm{T}}))$.

To solve the scaling parameter s, we equate the corresponding derivative of (14) to zero and obtain

$$s = \frac{\mathrm{tr}\left((\hat{\mathbf{Y}}^{\mathrm{T}}\mathbf{P}\hat{\mathbf{X}})^{\mathrm{T}}\mathbf{R}\right)}{\mathrm{tr}(\hat{\mathbf{X}}^{\mathrm{T}}\mathbf{P}\hat{\mathbf{X}}) + 2\lambda\sigma^2\mathrm{tr}(\mathbf{X}^{\mathrm{T}}\mathbf{Q}\mathbf{X})}. \tag{17}$$

So far, all the parameters in the M-step have been solved. We summarize our LLT algorithm for rigid matching in Algorithm 1.

Algorithm 1: The LLT algorithm for rigid matching

Input: Correspondence set $S = \{(\mathbf{x}_n, \mathbf{y}_n) : n \in \mathbb{N}_N\}$, parameters K, λ, τ
Output: Inlier set $\mathcal{I}$

Initialize γ, $\mathbf{R} = \mathbf{I}$, $\mathbf{t} = \mathbf{0}$, $\mathbf{P} = \mathbf{I}_{N\times N}$;
Set a to the volume of the output space;
Initialize σ^2 by Eq. (5);
Search the K nearest neighbors for each point in $\mathbf{X}$;
Compute $\mathbf{W}$ by minimizing the cost function (8);
repeat
  E-step:
    Update $\mathbf{P}$ by Eq. (4);
  M-step:
    Compute $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$ according to Eq. (13);
    Compute $\mathbf{USV}^{\mathrm{T}} = \mathrm{svd}(\hat{\mathbf{Y}}^{\mathrm{T}}\mathbf{P}\hat{\mathbf{X}})$;
    Update $\mathbf{R}$, s, $\mathbf{t}$ by Eqs. (16), (17) and (12);
    Update σ^2 and γ by Eqs. (5) and (6);
until $\mathcal{Q}$ *converges*;
The consensus set $\mathcal{I}$ is determined by Eq. (7).

D. Affine Feature Matching

Compared with the rigid case, affine feature matching is simpler since the optimization is unconstrained. We define the transformation as $\mathcal{T}(\mathbf{x}_n) = \mathbf{A}\mathbf{x}_n + \mathbf{t}$, where $\mathbf{A}$ is a 2×2 affine matrix, and $\mathbf{t}$ is a 2×1 translation vector. The objective function (10) then becomes

$$\Psi(\mathbf{A}, \mathbf{t}) = \frac{1}{2\sigma^2}\sum_{n=1}^{N} p_n\|\mathbf{y}_n - \mathbf{A}\mathbf{x}_n - \mathbf{t}\|^2 + \lambda\sum_{i=1}^{N} p_i\left\|\mathbf{A}\left(\mathbf{x}_i - \sum_{j=1}^{N}\mathbf{W}_{ij}\mathbf{x}_j\right)\right\|^2. \tag{18}$$

The solution of $\mathbf{t}$ is similar to the rigid case. The solution of $\mathbf{A}$ can be obtained by directly taking the partial derivative of Ψ, setting it to zero, and solving the resulting linear system of equations. The optimal $\mathbf{t}$ and $\mathbf{A}$ are given by

$$\mathbf{t} = \mu_{\mathbf{y}} - \mathbf{A}\mu_{\mathbf{x}} \tag{19}$$

$$\mathbf{A} = (\hat{\mathbf{Y}}^{\mathrm{T}}\mathbf{P}\hat{\mathbf{X}})(\hat{\mathbf{X}}^{\mathrm{T}}\mathbf{P}\hat{\mathbf{X}} + 2\lambda\sigma^2\mathbf{X}^{\mathrm{T}}\mathbf{Q}\mathbf{X})^{-1}. \tag{20}$$

So far, the affine transformation in the M-step has been solved. We summarize our LLT for affine matching in Algorithm 2.

E. Nonrigid Feature Matching

We now consider the case of nonrigid transformation. To this end, we define the transformation $\mathcal{T}$ as the initial position plus a displacement function $\mathbf{f} : \mathcal{T}(\mathbf{x}) = \mathbf{x} + \mathbf{f}(\mathbf{x})$, where $\mathbf{f}$ is modeled by requiring it to lie within a specific functional space $\mathcal{H}$, namely, a vector-valued RKHS [66] (associated with a particular kernel), as described in detail in Appendix A. We define $\mathcal{H}$ by a matrix-valued kernel $\Gamma : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^{2\times 2}$, and a diagonal Gaussian kernel $\Gamma(\mathbf{x}_i, \mathbf{x}_j) = \kappa(\mathbf{x}_i, \mathbf{x}_j)\cdot\mathbf{I} = e^{-\beta\|\mathbf{x}_i - \mathbf{x}_j\|^2}\cdot\mathbf{I}$ is chosen in this paper. Thus, we have the following theorem, and the proof is given in Appendix B.

Algorithm 2: The LLT algorithm for affine matching

Input: Correspondence set $S = \{(\mathbf{x}_n, \mathbf{y}_n) : n \in \mathbb{N}_N\}$, parameters K, λ, τ
Output: Inlier set $\mathcal{I}$

Initialize γ, $\mathbf{A} = \mathbf{I}$, $\mathbf{t} = \mathbf{0}$, $\mathbf{P} = \mathbf{I}_{N\times N}$;
Set a to the volume of the output space;
Initialize σ^2 by Eq. (5);
Search the K nearest neighbors for each point in $\mathbf{X}$;
Compute $\mathbf{W}$ by minimizing the cost function (8);
repeat
  E-step:
    Update $\mathbf{P}$ by Eq. (4);
  M-step:
    Compute $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$ according to Eq. (13);
    Update $\mathbf{A}$ and $\mathbf{t}$ by Eqs. (20) and (19);
    Update σ^2 and γ by Eqs. (5) and (6);
until $\mathcal{Q}$ *converges*;
The consensus set $\mathcal{I}$ is determined by Eq. (7).

Theorem 1: The optimal solution of the objective function (10) in the nonrigid case is given by

$$\mathcal{T}(\mathbf{x}) = \mathbf{x} + \mathbf{f}(\mathbf{x}) = \mathbf{x} + \sum_{n=1}^{N}\Gamma(\mathbf{x}, \mathbf{x}_n)\mathbf{c}_n \tag{21}$$

with the coefficient set $\{\mathbf{c}_n : n \in \mathbb{N}_N\}$ determined by a linear system

$$(\mathbf{P} + 2\lambda\sigma^2\mathbf{Q})\mathbf{\Gamma}\mathbf{C} = \mathbf{P}\mathbf{Y} - (\mathbf{P} + 2\lambda\sigma^2\mathbf{Q})\mathbf{X} \tag{22}$$

where $\mathbf{C} = (\mathbf{c}_1, \ldots, \mathbf{c}_N)^{\mathrm{T}}$, $\mathbf{\Gamma} \in \mathbb{R}^{N\times N}$ is the so-called Gram matrix with $\mathbf{\Gamma}_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j) = e^{-\beta\|\mathbf{x}_i - \mathbf{x}_j\|^2}$.

Fast Implementation: In a feature matching problem, the point set typically contains hundreds or thousands of points, which causes significant complexity problems (both in time and space). Consequently, we adopt a sparse approximation and randomly pick only a subset of size M input points $\{\tilde{\mathbf{x}}_m\}_{m=1}^{M}$ to have nonzero coefficients in the expansion of the solution [i.e., (21)]. This follows the work in [67] and [68], where it was found that this approximation works well and that simply selecting a random subset of the input points in this manner performs no worse than more sophisticated and time-consuming methods. Thus, we seek a solution of the following form:

$$\mathbf{f}(\mathbf{x}) = \sum_{m=1}^{M}\Gamma(\mathbf{x}, \tilde{\mathbf{x}}_m)\mathbf{c}_m. \tag{23}$$

The chosen point set $\{\tilde{\mathbf{x}}_m\}_{m=1}^{M}$ is somewhat analogous to control points. By using the sparse approximation, the linear system (22) becomes

$$\mathbf{E}^{\mathrm{T}}(\mathbf{P} + 2\lambda\sigma^2\mathbf{Q})\mathbf{E}\mathbf{C}^s = \mathbf{E}^{\mathrm{T}}\mathbf{P}\mathbf{Y} - \mathbf{E}^{\mathrm{T}}(\mathbf{P} + 2\lambda\sigma^2\mathbf{Q})\mathbf{X} \tag{24}$$

where the coefficient matrix $\mathbf{C}^s = (\mathbf{c}_1, \ldots, \mathbf{c}_M)^{\mathrm{T}} \in \mathbb{R}^{M\times 2}$, and $\mathbf{E} \in \mathbb{R}^{N\times M}$ with $\mathbf{E}_{ij} = \kappa(\mathbf{x}_i, \tilde{\mathbf{x}}_j) = e^{-\beta\|\mathbf{x}_i - \tilde{\mathbf{x}}_j\|^2}$.

We summarize our LLT for nonrigid matching in Algorithm 3.

Algorithm 3: The LLT algorithm for non-rigid matching

Input: Correspondence set $S = \{(\mathbf{x}_n, \mathbf{y}_n) : n \in \mathbb{N}_N\}$, kernel Γ, parameters K, λ, β, M, τ

Output: Inlier set $\mathcal{I}$

```
Initialize γ, C = 0, P = I_{N×N};
Set a to the area of the image;
Initialize σ² by Eq. (5);
Construct the Gram matrix Γ using the definition of Γ;
Search the K nearest neighbors for each point in X;
Compute W by minimizing the cost function (8);
repeat
    E-step:
      Update P by Eq. (4);
    M-step:
      Update C by solving linear system (22) or (24);
      Update σ² and γ by Eqs. (5) and (6);
until Q converges;
The consensus set I is determined by Eq. (7).
```

F. Convergence Analysis

Note that the negative log-likelihood function (2) is not convex; therefore, it is unlikely that any algorithm can find its global minimum. Our strategy is to initialize the variance σ^2 by a large initial value and then use the EM algorithm. At large σ^2, the objective function will be convex in a large region surrounding the global minimum. Hence, we are likely to find the global minimum for large variance. As σ^2 decreases, the position of the global minimum will tend to change smoothly. The objective function will be convex in a small region around its minimum, which makes it likely that using the old global minimum as an initial value could converge to the new global minimum. Therefore, as the iteration proceeds, we have a good chance of reaching the global minimum. This is conceptually similar to deterministic annealing [69], which uses the solution of an easy (e.g., smoothed) problem to recursively give initial conditions to increasingly harder problems, but it differs in several respects (e.g., by not requiring any annealing schedule). Moreover, the adaptive estimation of σ^2 also allows to reduce the number of free parameters and, more importantly, to obtain good estimates very quickly (e.g., avoiding many of the local minima inherent in the formulation).

G. Computational Complexity

To search the K nearest neighbors for each point in $\mathbf{X}$, the time complexity is close to $O((K+N)\log N)$ by using the $k-d$ tree [70]. According to (8), the time complexity of obtaining the weight matrix $\mathbf{W}$ is $O(K^3N)$ because each row of $\mathbf{W}$ can be solved separately with $O(K^3)$ time complexity. For the rigid and affine cases, the time complexities of solving the transformations are both $O(KN)$; hence, the total time complexities for rigid and affine matching are both $O(K^3N + N\log N)$. The space complexities for rigid and affine matching are both $O(KN)$ due to the memory requirements for storing the weight matrix $\mathbf{W}$.

TABLE I
COMPUTATIONAL COMPLEXITIES OF OUR LLT ALGORITHM

	Rigid	Affine	Non-rigid
Time	$K^3N + N\log N$	$K^3N + N\log N$	$K^3N + M^2N + N\log N$
Space	KN	KN	$MN + KN$

For the nonrigid case, the time complexity of solving the linear system (22) is $O(N^3)$; hence, the total complexity can be written as $O(N^3)$. The space complexity scales like $O(N^2)$ due to the memory requirements for storing the Gram matrix $\mathbf{\Gamma}$. By using the sparse approximation, the time complexity of solving the linear system (24) reduces to $O(M^2N + KMN + K^2N)$. Therefore, the total time complexity is $O(K^3N + M^2N + N\log N)$, which is about linearithmic with respect to the scale of the given correspondence set. The space complexity reduces to $O(MN + KN)$ due to the memory requirements for storing $\mathbf{E}$ and $\mathbf{W}$.

We summarize the time and space complexities of our LLT algorithm in Table I. We see that the time complexities are all linearithmic, and the space complexities are all linear, with respect to the scale of the given correspondence set. This is significant for large-scale problems.

H. Implementation Details

The performance of feature matching algorithms depends, typically, on the coordinate system in which feature points are expressed. We use data normalization to control for this. More specifically, we perform a linear rescaling so that the spatial positions of the two feature point sets both have zero mean and unit variance. Note that the constant a of the uniform distribution in (1) should be set according to the data normalization.

Parameter Setting: There are mainly six parameters in our method: K, λ, τ, γ, β, and M. Parameter K controls the number of nearest neighbors for linear reconstruction. Parameter λ controls the influence of the local geometrical constraint on the transformation $\mathcal{T}$. Parameter τ is a threshold, which is used for deciding the correctness of a correspondence. Parameter γ reflects our initial assumption on the amount of inliers in the correspondence sets. Parameters β and M are used in our nonrigid matching algorithm, where the former determines how wide the range of interaction between feature points, and the latter is the required number of control points for sparse approximation. We set $K = 15$, $\lambda = 1000$, $\tau = 0.5$, $\gamma = 0.9$, $\beta = 0.1$ and $M = 15$, throughout this paper.

IV. EXPERIMENTAL RESULTS

Here, we test the performance of our proposed LLT and compare it with other three state-of-the-art feature matching methods such as RANSAC [39], ICF [45], and GS [44]. Throughout the experiments, four algorithms' parameters are all fixed. The experiments are performed on a laptop with 2.5-GHz Intel Core CPU, 8-GB memory, and MATLAB code.

A. Data Sets and Settings

To evaluate our LLT algorithm, we design a set of experiments on feature matching of remote sensing images involving

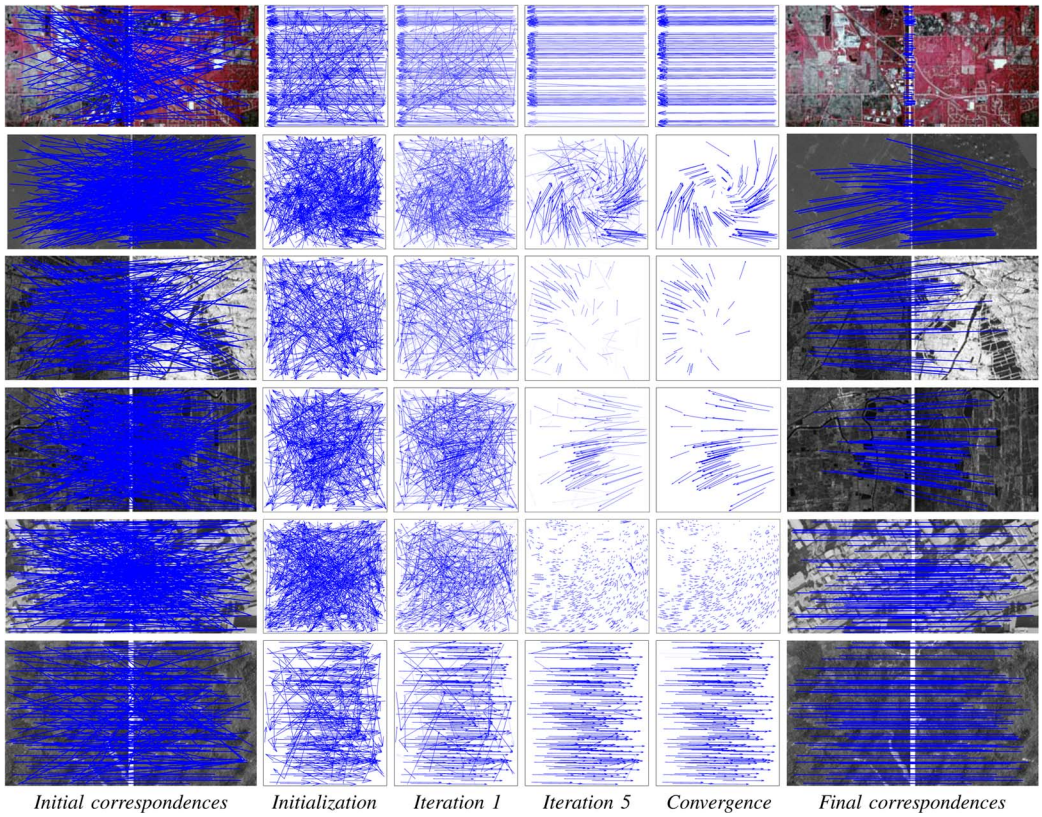


Fig. 2. Qualitative matching results on several typical remote sensing image pairs. The first two rows satisfy the rigid model, the middle two rows satisfy the affine model, and the last two rows satisfy the nonrigid model. The columns show the matching process during the EM iteration, and the level of the blue color indicates to what degree a correspondence belongs to the inlier. For visibility, at most 300 randomly selected matches are presented in the first column.

rigid, affine, and nonrigid transformations. Next, we discuss about the data sets used in this paper.

- *Rigid data set*. The test data set consists of two types of image pairs, which were captured over eastern Illinois, USA (from Erdas example data[4]) and Shanghai, China. The first contains 20 color-infrared aerial photograph image pairs with small overlap areas of sizes from 1391×1374 to 1450×1380, which typically arises in the image mosaic problem. The second contains eight SPOT image pairs of all size 3086×2865 with each pair representing the same area taken at different times, which typically arises in the change detection problem. The latter type of image pairs have already been aligned; hence, we manually add random rigid transformations to make the matching problem more challenging.
- *Affine data set*. The test data set contains 16 image pairs with sizes from 400×400 to 1072×1072 , where the reference images are the satellite-based SAR data of RADARSAT II, and the sensed images are the airborne SAR data of an unmanned aerial vehicle. All the image pairs involved affine distortion and were captured over Nantong, Jiangsu Province, China, during October and November in 2013.
- *Nonrigid data set*. The test data set contains 16 pairs of panchromatic aerial photographs with sizes from 638×750 to 1432×1632, which were captured by a frame camera over Tokyo, Japan and Wuhan, China. The image pairs involve ground relief variations and imaging viewpoint changes and hence are not exactly matchable via a parametric model such as rigid or affine transformation.

In all the experimental data sets, different image pairs are captured from different locations. We adopt the SIFT algorithm to determine the putative feature correspondences, which is implemented by the open-source VLFEAT toolbox [71]. The experimental results are evaluated by precision and recall, where the precision is defined as the ratio of the preserved inlier number and the preserved correspondence number, and the recall is defined as the ratio of the preserved inlier number and the inlier number contained in the putative correspondence set. To establish the ground truth, i.e., determining the matching correctness of each correspondence, we first use LLT to establish rough correspondences and then confirm the results artificially, including both the preserved and removed correspondences produced by LLT.

B. Qualitative Experiments

Our first experiment involves feature matching on several typical remote sensing image pairs satisfying rigid, affine, or nonrigid model, as shown in the first column of Fig. 2, where each case contains two image pairs. We see that the matching problem on these image pairs is quite a challenge due to the small overlap areas in the first pair, the severe noise in the

[4]The data set is available at: http://download.intergraph.com/downloads/erdas-imagine-2013-2014-example-data.

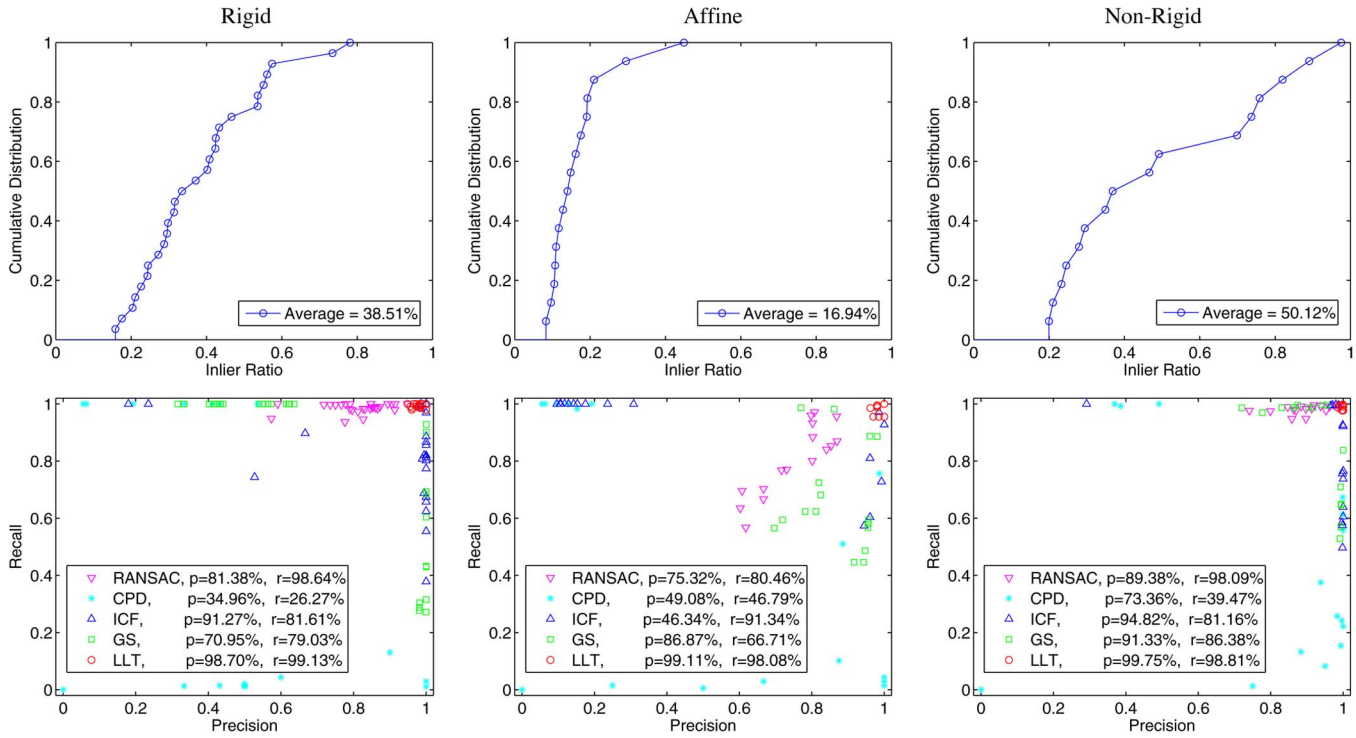


Fig. 3. Quantitative comparisons of LLT with RANSAC [39], GS [44], and ICF [45] on the rigid (left column), affine (middle column) and nonrigid (right column) data sets. For each column, the top figure is the cumulative distribution function of the initial inlier ratio in the data set, and the bottom figure is the precision–recall statistics of the four methods. Our LLT (red circles, upper right corner) has the best precision and recall overall.

middle two pairs, the ground relief variations in the last two pairs, etc. We use our LLT algorithm with corresponding model to establish accurate feature correspondences.

The whole matching process on these image pairs is illustrated schematically in Fig. 2. The columns show the iterative process, the level of the blue color indicates to what degree a correspondence belongs to inlier, and it is also the posterior probability p_n in (4). In the beginning, we establish the putative correspondence sets by using the SIFT algorithm, and all the SIFT matches are assumed to be inlier, as shown in the first column. To visualize the matching process, we represent the correspondences as motion fields, as shown in the second column, where the head and tail of each arrow correspond to the positions of a SIFT match in two images. As the EM iterative process continues, progressively more refined matches are shown in the third, fourth, and fifth columns. The fifth column shows that LLT almost converges to a nearly binary decision on the match correctness. The SIFT matches finally preserved by our LLT are presented in the last column. From the results, we see that our LLT is able to distinguish inliers form the outliers on all the six typical pairs.

C. Quantitative Evaluation

We next give quantitative comparisons of our LLT with RANSAC [39], CPD [57], ICF [45], and GS [44] on the rigid, affine, and nonrigid data sets. The cumulative distribution functions of the initial inlier ratio are shown in the top row of Fig. 3, where the average inlier percentages are about 38.51%, 16.94%, and 52.10%, and the average numbers of correct correspondences in the putative sets are about 117, 70, and 621, respectively. We see that the data sets are quite challenging, particularly the affine data set, where most image pairs have inlier percentages below 20% due to the severe noise in the SAR data.

The results of five methods on the three data sets are summarized in the bottom row of Fig. 3, in which each scattered dot represents a precision–recall pair on an image pair. For RANSAC, the rigid transformation, affine transformation, and fundamental matrix are chosen as the parametric models on the rigid, affine, and nonrigid data sets, respectively. From the results, we see that RANSAC can produce satisfying results when the initial putative set does not contain many outliers, e.g., in the nonrigid data set. However, its performance degenerates rapidly as the percentage of outliers increases, such as in the affine data set. The performance of CPD is not that satisfied, and it completely fails on many image pairs (indeed the image pairs with more outlier correspondences). This is not surprising since correspondence matrix-based methods do not use local descriptors to establish initial correspondences; hence, they cannot handle a large number of outliers. ICF usually has high precision or recall, but not simultaneously. It lacks robustness when the outlier percentage is high. GS has a better precision–recall tradeoff compared with ICF, but the matching results are still not that satisfactory. This is probably because that GS cannot estimate the factor for affinity matrix automatically, and it is not affine invariant. In contrast, our proposed method LLT has the best matching performance, where the precisions and recalls are all close to 1. Note that all methods perform best on the nonrigid data set. This is because that the nonrigid data set contains high-quality images; hence, both the initial inlier percentages

TABLE II
COMPARISON OF AVERAGE RUN TIMES (SEC.) ON THE THREE DATA SETS

	Rigid	Affine	Non-Rigid
RANSAC [39]	16.32 ± 10.08	15.36 ± 3.76	12.23 ± 19.17
CPD [57]	1.33 ± 1.05	0.82 ± 0.30	6.24 ± 7.77
ICF [45]	2.13 ± 2.35	1.66 ± 1.21	2.36 ± 2.48
GS [44]	0.98 ± 0.74	0.64 ± 0.35	15.28 ± 18.95
LLT	$\mathbf{0.34 \pm 0.16}$	$\mathbf{0.34 \pm 0.23}$	$\mathbf{1.85 \pm 2.81}$

TABLE III
AVERAGE PRECISION–RECALL PAIRS OF OUR LLT ON THE RIGID, AFFINE, AND NONRIGID DATA SETS. EACH ROW IS THE RESULTS WITH A CERTAIN MODEL, SUCH AS RIGID (R), AFFINE (A), AND NONRIGID (NR)

	Rigid	Affine	Non-Rigid
LLT (R)	**(98.70%, 99.13%)**	(91.61%, 75.96%)	(98.32%, 95.41%)
LLT (A)	(98.68%, 98.89%)	**(99.11%, 98.08%)**	(98.84%, 95.98%)
LLT (NR)	(98.70%, 98.91%)	(98.51%, 97.86%)	**(99.75%, 98.81%)**

(i.e., 50.12%) and inlier number (621) are quite large (they are the two critical factors influencing the matching performance which we will discuss in the next section).

The average run times of the four methods on the three data sets are provided in Table II. We see that RANSAC is not very efficient due to the low initial inlier percentages in the data sets. CPD and GS are quite efficient when the scale of the data set is not large but degrades rapidly as the scale of the data set grows. ICF has better efficiency compared with the former two algorithms. However, our LLT clearly is even faster on all the three data sets. The large variances of the run times is because that the initial inlier percentages have large variances.

We also have tested LLT with different models on the three data sets, and the average precision–recall pairs are summarized in Table III. As can be seen from the diagonal of the table, for each data set, the "correct model" produces the best performance. On the rigid data set, the affine and nonrigid models can also work well. This is not surprising since the rigid model is a special case of the affine or nonrigid model. On the affine data set, the affine distortions are quite large (as can be seen from the middle two rows in Fig. 2); hence, the rigid model fails. On the nonrigid data set, we found that the main reason leading to the inferior performance (e.g., low recall) of the rigid and affine model is that they falsely remove the correspondences with large disparities caused by the nonrigid distortions (e.g., the spatially variant terrain relief). Nevertheless, they still generate good results. This can be attributed to that the nonrigid distortions are relatively slight, and a rigid or an affine model can approximate the image relationship sufficiently well.

The influence of the transforming cost term on the performance of LLT has also been investigated on the three data sets. To this end, we test the average precision–recall pairs of the proposed maximum-likelihood framework without the transforming cost term, i.e., setting $\lambda = 0$ in (9). The statistic results are given in Table IV. Clearly, LLT with the transforming cost term can achieve much better performance, particularly in the affine and nonrigid cases. This is appropriate since the severe outliers in the affine data set and the nonrigid deformations in the nonrigid data set will probably make the matching problem not well posed, which means that the transforming cost term indeed plays an important role for solving the problem.

TABLE IV
AVERAGE PRECISION–RECALL PAIRS OF OUR LLT AND LLT WITHOUT THE TRANSFORMING COST TERM (I.E., $\lambda = 0$) ON THE RIGID, AFFINE, AND NONRIGID DATA SETS

	Rigid	Affine	Non-Rigid
LLT ($\lambda = 0$)	(97.98%, 98.15%)	(84.57%, 90.05%)	(74.20%, 75.78%)
LLT	**(98.70%, 99.13%)**	**(99.11%, 98.08%)**	**(99.75%, 98.81%)**

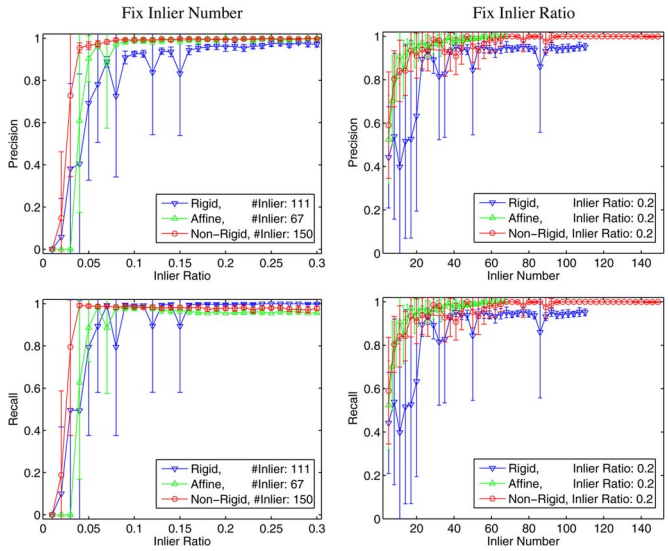


Fig. 4. Robustness tests of our LLT algorithm. Each column is a group of results, where the top figure is the precision curves and the bottom figure is the recall curves. The curves marked by "$\triangledown$" are the results of LLT with a rigid model on a rigid image pair (i.e., the first row in Fig. 2); the curves marked by "$\triangle$" are the results of LLT with affine model on an affine image pair (i.e., the third row in Fig. 2); the curves marked by "$\circ$" are the results of LLT with a nonrigid model on a nonrigid image pair (i.e., the fifth row in Fig. 2). In the left column, we fix the inlier number and vary the inlier ratio; in the right column, we fix the inlier ratio and vary the inlier number. The error bars indicate the precision/recall means and standard deviations over ten trials.

D. Robustness Test

Finally, we test the robustness of our LLT with different models. To this end, we use the three typical image pairs shown in the first, third, and fifth rows of Fig. 2 to test LLT with rigid, affine, and nonrigid models, respectively. In our evaluation, we consider the following two scenarios.

On one hand, the initial inlier percentage in the data set is clearly an important factor that influences the matching performance. Thus, we fix the inlier number and vary the inlier ratio, and obtain the precision and recall curves of LLT with different models on the corresponding image pairs as shown in the left column of Fig. 4. In the three image pairs, the initial inlier numbers are 111, 67, and 150, respectively. We fix them and vary the inlier ratio from 0.01 to 0.3 at an interval of 0.01. From the results, we see that LLT can work well even the inlier ratio below 0.1, and the matching performance becomes better and stable as the inlier ratio increases above 0.1.

On the other hand, the absolute number of correct correspondences in the data set can also influence the matching performance. To validate this idea, we fix the inlier ratio to 0.2 and then vary the inlier number. More specifically, we vary the inlier number from 5 to 110 at an interval of 3 in the rigid image pair, from 5 to 67 at an interval of 2 in the affine image pair,

and from 5 to 149 at an interval of 3 in the nonrigid image pair, respectively. The precision and recall curves are given in the right column of Fig. 4. We see that the matching performance becomes better as the inlier number grows, and the results are in general satisfactory when the initial inlier numbers are large than about 30. That is, at a fixed inlier ratio, the putative sets with fewer inliers are more sensitive to the noise and outliers.

V. Discussion and Conclusion

In this paper, we proposed a new feature matching algorithm called LLT for remote sensing image registration. It simultaneously estimates the transformation and generates accurate correspondences by an iterative EM algorithm under a maximum-likelihood framework with a local geometrical constraint. The method is robust and general, which is able to handle both rigid and nonrigid transformations in case of severe outliers and hence can be applied to various registration tasks in remote sensing. The qualitative and quantitative comparisons on several types of remote sensing image pairs demonstrate that our method significantly outperforms other state-of-the-art methods for feature matching.

The local geometrical constraint plays a role of regularization on the transformation $\mathcal{T}$. In the nonrigid case, the regularization may be assured in a more simple way by defining a smoothness functional $\phi(\mathcal{T})$, where a lower value of the functional corresponds to smoother function, e.g., $\phi(\mathcal{T}) = \|\mathcal{T}\|_\Gamma^2$ with $\|\cdot\|_\Gamma$ being a norm in the RKHS $\mathcal{H}$ defined by the kernel Γ [63]. However, this type of regularization typically imposes global smoothness on the transformation $\mathcal{T}$ and cannot be generalized to the rigid or affine case since a linear transformation is always smooth and $\phi(\mathcal{T})$ is a constant. Our proposed LLT is general, which can impose regularization in both linear and nonlinear cases based on sustaining the local structure.

Our maximum-likelihood formulation and local geometrical constraint are applicable for both 2-D and 3-D images. Using the SIFT algorithm for putative correspondence construction, we only demonstrate it on 2-D images. We will test the effectiveness of our LLT for 3-D feature matching by extending the feature extraction method to 3-D images.

Appendix A. Vector-Valued Reproducing Kernel Hilbert Space

We review the basic theory of vector-valued RKHS, and for further details and references, we refer to [66] and [72].

Let $\mathcal{X}$ be a set, e.g., $\mathcal{X} \subseteq \mathbb{R}^P$, $\mathcal{Y}$ a real Hilbert space with inner product (norm) $\langle\cdot,\cdot\rangle$, $(\|\cdot\|)$, e.g., $\mathcal{Y} \subseteq \mathbb{R}^D$; and $\mathcal{H}$ a Hilbert space with inner product (norm) $\langle\cdot,\cdot\rangle_\mathcal{H}$, $(\|\cdot\|_\mathcal{H})$, where $P = D = 2$ or 3 for point matching problem. Note that a norm can be induced by an inner product, e.g., $\forall\, \mathbf{f} \in \mathcal{H}$, $\|\mathbf{f}\|_\mathcal{H} = \sqrt{\langle \mathbf{f}, \mathbf{f}\rangle_\mathcal{H}}$. A Hilbert space is a real or complex inner product space that is also a complete metric space with respect to the distance function induced by the inner product. Thus, a vector-valued RKHS can be defined as follows.

Definition 1: A Hilbert space $\mathcal{H}$ is an RKHS if the evaluation maps $ev_\mathbf{x} : \mathcal{H} \to \mathcal{Y}$ (i.e., $ev_\mathbf{x}(\mathbf{f}) = \mathbf{f}(\mathbf{x})$) are bounded, i.e., if $\forall\, \mathbf{x} \in \mathcal{X}$, there exists a positive constant $C_\mathbf{x}$ such that

$$\|ev_\mathbf{x}(\mathbf{f})\| = \|\mathbf{f}(\mathbf{x})\| \leq C_\mathbf{x}\|\mathbf{f}\|_\mathcal{H}, \quad \forall\, \mathbf{f} \in \mathcal{H}. \tag{25}$$

A reproducing kernel $\Gamma : \mathcal{X} \times \mathcal{X} \to \mathcal{B}(\mathcal{Y})$ is then defined as $\Gamma(\mathbf{x}, \mathbf{x}') := ev_\mathbf{x} ev_{\mathbf{x}'}^*$, where $\mathcal{B}(\mathcal{Y})$ is the Banach space of bounded linear operators (i.e., $\Gamma(\mathbf{x}, \mathbf{x}'), \forall\, \mathbf{x}, \mathbf{x}' \in \mathcal{X}$) on $\mathcal{Y}$, e.g., $\mathcal{B}(\mathcal{Y}) \subseteq \mathbb{R}^{D\times D}$, and $ev_\mathbf{x}^*$ is the adjoint of $ev_\mathbf{x}$. We have the following two properties about the RKHS and kernel.

Remark 1: The kernel Γ reproduces the value of a function $\mathbf{f} \in \mathcal{H}$ at a point $\mathbf{x} \in \mathcal{X}$. Indeed, $\forall\, \mathbf{x} \in \mathcal{X}$ and $\mathbf{y} \in \mathcal{Y}$, we have $ev_\mathbf{x}^*\mathbf{y} = \Gamma(\cdot, \mathbf{x})\mathbf{y}$, so that $\langle \mathbf{f}(\mathbf{x}), \mathbf{y}\rangle = \langle \mathbf{f}, \Gamma(\cdot, \mathbf{x})\mathbf{y}\rangle_\mathcal{H}$.

Remark 2: An RKHS defines a corresponding reproducing kernel. Conversely, a reproducing kernel defines a unique RKHS.

More specifically, for any $N \in \mathbb{N}$, $\{\mathbf{x}_n\}_{n=1}^N \subseteq \mathcal{X}$, and a reproducing kernel Γ, a unique RKHS can be defined by considering the completion of the space, i.e.,

$$\mathcal{H}_N = \left\{ \sum_{n=1}^N \Gamma(\cdot, \mathbf{x}_n)\mathbf{c}_n : \mathbf{c}_n \in \mathcal{Y} \right\} \tag{26}$$

with respect to the norm induced by the inner product

$$\langle \mathbf{f}, \mathbf{g}\rangle_\mathcal{H} = \sum_{i,j=1}^N \langle \Gamma(\mathbf{x}_j, \mathbf{x}_i)\mathbf{c}_i, \mathbf{d}_j\rangle \quad \forall\, \mathbf{f}, \mathbf{g} \in \mathcal{H}_N \tag{27}$$

where $\mathbf{f} = \sum_{i=1}^N \Gamma(\cdot, \mathbf{x}_i)\mathbf{c}_i$ and $\mathbf{g} = \sum_{j=1}^N \Gamma(\cdot, \mathbf{x}_j)\mathbf{d}_j$.

Appendix B. Proof of Theorem 1

For any given reproducing kernel Γ, we can define a unique RKHS $\mathcal{H}_N$ as in (26) in Appendix A. Let $\mathcal{H}_N^\perp$ be a subspace of $\mathcal{H}$

$$\mathcal{H}_N^\perp = \{\mathbf{f} \in \mathcal{H} : \mathbf{f}(\mathbf{x}_n) = 0, n \in \mathbb{N}\}. \tag{28}$$

From the reproducing property, i.e., Remark 1, $\forall\, \mathbf{f} \in \mathcal{H}_N^\perp$

$$\left\langle \mathbf{f}, \sum_{n=1}^N \mathbf{\Gamma}(\cdot, \mathbf{x}_n)\mathbf{c}_n \right\rangle_\mathcal{H} = \sum_{n=1}^N \langle \mathbf{f}(\mathbf{x}_i), \mathbf{c}_n\rangle = 0. \tag{29}$$

Thus, $\mathcal{H}_N^\perp$ is the orthogonal complement of $\mathcal{H}_N$; then, every $\mathbf{f} \in \mathcal{H}$ can be uniquely decomposed in components along and perpendicular to $\mathcal{H}_N$: $\mathbf{f} = \mathbf{f}_N + \mathbf{f}_N^\perp$, where $\mathbf{f}_N \in \mathcal{H}_N$ and $\mathbf{f}_N^\perp \in \mathcal{H}_N^\perp$. That is, $\forall\, \mathbf{f} \in \mathcal{H}$, we have $\mathbf{f}(\mathbf{x}_n) = \mathbf{f}_N(\mathbf{x}_n)$. Therefore, the optimal displacement function $\mathbf{f}$ comes from the space $\mathcal{H}_N$; hence, the optimal solution of the objective function (10) has the form (21).

To solve the coefficient set $\mathbf{C}$, we consider the terms of Ψ that are related to $\mathbf{C}$ and rewrite them in matrix form as follows:

$$\begin{aligned}\Psi(\mathbf{C}) = \frac{1}{2\sigma^2}\mathrm{tr}\left(\mathbf{C}^\mathrm{T}\mathbf{\Gamma}\mathbf{P}\mathbf{\Gamma}\mathbf{C} - 2\mathbf{C}^\mathrm{T}\mathbf{\Gamma}\mathbf{P}(\mathbf{Y} - \mathbf{X})\right) \\ + \lambda\mathrm{tr}(\mathbf{C}^\mathrm{T}\mathbf{\Gamma}\mathbf{Q}\mathbf{\Gamma}\mathbf{C} + 2\mathbf{C}^\mathrm{T}\mathbf{\Gamma}\mathbf{Q}\mathbf{X}).\end{aligned} \tag{30}$$

Taking derivative of (30) with respect to $\mathbf{C}$ and setting it to zero, we obtain the linear system in (22). Thus, the coefficient set $\{\mathbf{c}_n : n \in \mathbb{N}_N\}$ of the optimal solution is determined by the linear system (22).

References

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

[2] A. Wong and D. A. Clausi, "ARRSI: Automatic registration of remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1483–1493, May 2007.

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

[4] J. Ma, J. Zhao, Y. Ma, and J. Tian, "Non-rigid visible and infrared face registration via regularized gaussian fields criterion," *Pattern Recognit.*, vol. 48, no. 3, pp. 772–784, Mar. 2015.

[5] A. Dempster, N. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *J. R. Stat. Soc. Ser. B*, vol. 39, no. 1, pp. 1–38, 1977.

[6] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[7] N. Aronszajn, "Theory of reproducing kernels," *Trans. Amer. Math. Soc.*, vol. 68, no. 3, pp. 337–404, May 1950.

[8] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. Int. Joint Conf. Artif. Intell.*, 1981, vol. 2, pp. 674–679.

[9] Y. Liu, "Automatic range image registration in the markov chain," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 12–29, Jan. 2010.

[10] X. Wang, B. Feng, X. Bai, W. Liu, and L. J. Latecki, "Bag of contour fragments for robust shape classification," *Pattern Recognit.*, vol. 47, no. 6, pp. 2116–2125, Jun. 2014.

[11] C. Yao, X. Bai, and W. Liu, "A unified framework for multi-oriented text detection and recognition," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4737–4749, Nov. 2014.

[12] X. Bai and L. J. Latecki, "Path similarity skeleton graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 7, pp. 1282–1292, Jul. 2008.

[13] Y. Zhou, X. Bai, W. Liu, and L. J. Latecki, "Fusion with diffusion for robust visual tracking," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2012, vol. 4, pp. 2978–2986.

[14] X. Guo and X. Cao, "Good match exploration using triangle constraint," *Pattern Recognit. Lett.*, vol. 33, no. 7, pp. 872–881, May 2012.

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

[16] D. L. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, "Medical image registration," *Phys. Med. Biol.*, vol. 46, no. 3, R1–45, Mar. 2001.

[17] Y. Li, C. Chen, F. Yang, and J. Huang, "Deep sparse representation for robust image registration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2015, pp. 4894–4901.

[18] J. Jiang, R. Hu, Z. Wang, and Z. Han, "Face super-resolution via multilayer locality-constrained iterative neighbor embedding and intermediate dictionary learning," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4220–4231, Oct. 2014.

[19] J. Jiang, R. Hu, Z. Wang, and Z. Han, "Noise robust face hallucination via locality-constrained representation," *IEEE Trans. Multimedia*, vol. 16, no. 5, pp. 1268–1281, Aug. 2014.

[20] X. Bai, X. Yang, L. J. Latecki, W. Liu, and Z. Tu, "Learning context-sensitive shape similarity by graph transduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 861–874, May 2010.

[21] 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.

[22] S. Dawn, V. Saxena, and B. Sharma, "Remote sensing image registration techniques: A survey," in *Image and Signal Processing*, vol. 6134, Berlin, Germany: Springer-Verlag, 2010, pp. 103–112.

[23] C. Chen, Y. Li, W. Liu, and J. Huang, "Image fusion with local spectral consistency and dynamic gradient sparsity," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2014, pp. 2760–2765.

[24] L. Ma, M. M. Crawford, X. Yang, and Y. Guo, "Local-manifold-learning-based graph construction for semisupervised hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2832–2844, May 2015.

[25] J. Maintz and M. A. Viergever, "A survey of medical image registration," *Med. Image Anal.*, vol. 2, no. 1, pp. 1–36, Mar. 1998.

[26] R. C. Gonzalez and P. Wintz, *Digital Image Processing*. New York, NY, USA: Addison-Wesley, 1987.

[27] J. Le Moigne, W. J. Campbell, and R. 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.

[28] R. N. Bracewell, *The Fourier Transform and its Applications*. New York, NY, USA: McGraw-Hill, 1986.

[29] 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, Aug. 1996.

[30] 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, Dec. 1994.

[31] J. Inglada, V. Muron, D. Pichard, and T. Feuvrier, "Analysis of artifacts in subpixel remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 1, pp. 254–264, Jan. 2007.

[32] J. Liang *et al.*, "Automatic registration of multisensor images using an integrated spatial and mutual information (SMI) metric," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 603–615, Jan. 2014.

[33] A. Rangarajan, H. Chui, and J. S. Duncan, "Rigid point feature registration using mutual information," *Med. Image Anal.*, vol. 3, no. 4, pp. 425–440, Dec. 1999.

[34] 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.

[35] S. Pang, J. Xue, Q. Tian, and N. Zheng, "Exploiting local linear geometric structure for identifying correct matches," *Comput. Vis. Image Understand.*, vol. 128, pp. 51–64, Nov. 2014.

[36] J. Ma, Y. Ma, J. Zhao, and J. Tian, "Image feature matching via progressive vector field consensus," *IEEE Signal Process. Lett.*, vol. 22, no. 6, pp. 767–771, Jun. 2015.

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

[38] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, Apr. 2002.

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

[40] 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.

[41] O. Chum and J. Matas, "Matching with PROSAC—Progressive sample consensus," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2005, vol. 1, pp. 220–226.

[42] J. Zhao, J. Ma, J. Tian, J. Ma, and D. Zhang, "A robust method for vector field learning with application to mismatch removing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2011, pp. 2977–2984.

[43] J. Ma *et al.*, "Robust L_2E estimation of transformation for non-rigid registration," *IEEE Trans. Signal Process.*, vol. 63, no. 5, pp. 1115–1129, Mar. 2015.

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

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

[46] A. Sedaghat, M. Mokhtarzade, and H. Ebadi, "Uniform robust scale-invariant feature matching for optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4516–4527, Nov. 2011.

[47] G.-R. Cai *et al.*, "Perspective-sift: An efficient tool for low-altitude remote sensing image registration," *Signal Process.*, vol. 93, no. 11, pp. 3088–3110, Nov. 2013.

[48] 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.

[49] R. Bouchiha and K. Besbes, "Automatic remote-sensing image registration using surf," *Int. J. Comput. Theory Eng.*, vol. 5, no. 1, pp. 88–92, Feb. 2013.

[50] A. Wong and D. A. Clausi, "AISIR: Automated inter-sensor/inter-band satellite image registration using robust complex wavelet feature representations," *Pattern Recognit. Lett.*, vol. 31, no. 10, pp. 1160–1167, Jul. 2010.

[51] 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.

[52] Z. Song, S. Zhou, and J. Guan, "A novel image registration algorithm for remote sensing under affine transformation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4895–4912, Aug. 2014.

[53] 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.

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

[55] Y. Liu, "Improving ICP with easy implementation for free-form surface matching," *Pattern Recognit.*, vol. 37, no. 2, pp. 211–226, Feb. 2004.

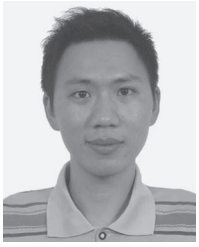
[56] H. Chui and A. Rangarajan, "A new point matching algorithm for non-rigid registration," *Comput. Vis. Image Understand.*, vol. 89, no. 2/3, pp. 114–141, Feb. 2003.

[57] 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.

[58] S. Ge, G. Fan, and M. Ding, "Non-rigid point set registration with global-local topology preservation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. Workshops*, Jun. 2014, pp. 245–251.
[59] S. Ge and G. Fan, "Non-rigid articulated point set registration for human pose estimation," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, Jan. 2015, pp. 94–101.
[60] R. Horaud, F. Forbes, M. Yguel, G. Dewaele, and J. Zhang, "Rigid and articulated point registration with expectation conditional maximization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 587–602, Mar. 2011.
[61] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer-Verlag, 2006.
[62] J. Ma, J. Chen, D. Ming, and J. Tian, "A mixture model for robust point matching under multi-layer motion," *PloS One*, vol. 9, no. 3, Mar. 2014, Art. ID. e92282.
[63] T. Evgeniou, M. Pontil, and T. Poggio, "Regularization networks and support vector machines," *Comput. Math.*, vol. 13, no. 1, pp. 1–50, Apr. 2000.
[64] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 4, pp. 376–380, Apr. 1991.
[65] A. Myronenko and X. Song, "On the closed-form solution of the rotation matrix arising in computer vision problems," Dept. Sci. Eng., Oregon Health Sci. Univ., Portland, OR, USA. unpublished paper. [Online]. Available: http://arxiv.org/abs/0904.1613
[66] C. A. Micchelli and M. Pontil, "On learning vector-valued functions," *Neural Comput.*, vol. 17, no. 1, pp. 177–204, Jan. 2005.
[67] R. Rifkin, G. Yeo, and T. Poggio, "Regularized least-squares classification," in *Advances in Learning Theory: Methods, Models and Applications*, Amsterdam, The Netherlands: IOS Press, 2003.
[68] J. Ma, J. Zhao, J. Tian, X. Bai, and Z. Tu, "Regularized vector field learning with sparse approximation for mismatch removal," *Pattern Recognit.*, vol. 46, no. 12, pp. 3519–3532, Dec. 2013.
[69] A. L. Yuille, "Generalized deformable models, statistical physics, and matching problems," *Neural Comput.*, vol. 2, no. 1, pp. 1–24, Mar. 1990.
[70] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Commun. ACM*, vol. 18, no. 9, pp. 509–517, Sep. 1975.
[71] A. Vedaldi and B. Fulkerson, "VLFeat—An open and portable library of computer vision algorithms," in *Proc. Int. Conf. Multimedia*, 2010, pp. 1469–1472.
[72] C. Carmeli, E. De Vito, and A. Toigo, "Vector valued reproducing kernel Hilbert spaces of integrable functions and mercer theorem," *Anal. Appl.*, vol. 4, no. 10, pp. 377–408, Oct. 2006.

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 with the Department of Statistics, University of California at Los Angeles, Los Angeles, CA, USA. He is currently a Postdoctoral Researcher with the Electronic Information School, Wuhan University, Wuhan. His current research interests include computer vision, machine learning, and pattern recognition.



Huabing Zhou received the Ph.D. degree in control science and engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2012.

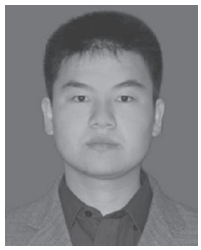
He is currently an Assistant Professor with the School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan. His research interests include remote sensing image analysis, computer vision, and intelligent robotics.

Ji Zhao received the B.S. degree in automation from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2005 and the Ph.D. degree in control science and engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2012.

From 2012 to 2014, he was a Postdoctoral Research Associate with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA. He is currently a Researcher with Samsung Research Center, Beijing, China. His research interests include computer vision and machine learning.

Yuan Gao received the B.S. degree in biomedical engineering and the M.S. degree in pattern recognition and intelligent systems from the Huazhong University of Science and Technology, Wuhan, China, in 2009 and 2012, respectively. He is currently working toward the Ph.D. degree with the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong.

His research interests include machine learning, pattern recognition, and applications.



Junjun Jiang received the B.S. degree in information and computing science from Huaqiao University, Quanzhou, China, in 2009 and the Ph.D. degree in communication and information system from Wuhan University, Wuhan, China, in 2014.

He is currently an Associate Professor with the School of Computer Science, China University of Geosciences, Wuhan. He is the author or coauthor of more than 40 scientific articles and is the holder of four Chinese patents. His research interests include applications of image processing and pattern recognition in video surveillance, image super-resolution, image interpolation, and face recognition.

Jinwen Tian received the Ph.D. degree in pattern recognition and intelligent systems from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 1998.

He is currently a Professor and a Ph.D. Supervisor of pattern recognition and artificial intelligence with HUST. His main research interests include remote sensing image analysis, wavelet analysis, image compression, computer vision, and fractal geometry.