Feature Matching for Remote Sensing Image Registration via Manifold Regularization

Huabing Zhou, Anna Dai, Tian Tian, Yulu Tian, Zhenghong Yu, Yuntao Wu, and Yanduo Zhang

***Abstract*—Feature matching is critical in analyzing remote sensing images, aiming to find the optimal mapping between correspondences. Regularization technology is essential to ensure the well-posedness of feature matching. However, current regularization-based methods scarcely consider the geometry structure of the image, which is beneficial for estimating the mapping, especially when the image pairs have a large view or scale change and local distortion. In this article, we introduce manifold regularization to overcome this limit and formulate feature matching as a unified semisupervised latent variable mixture model for both rigid and nonrigid transformations. Especially, we apply a Bayesian model with latent variables indicating whether matches in the putative correspondences are outliers or inliers. Moreover, we employ all the feature points, only part of which have correct matches, to express the intrinsic structure, which is preserved by manifold regularization. Finally, we combine manifold regularization with three different transformation models (e.g., rigid, affine, and thin-plate spline) to estimate the corresponding mappings. Experimental results on four remote sensing image datasets demonstrate that our method can significantly outperform the state of the art.**

***Index Terms*—Feature matching, image registration, manifold regularization.**

I. Introduction

IMAGE registration is a fundamental and challenging problem in the image process, aiming to align geometrically two or more images, which are from different views, different times, even various sensors, but the same scene [1]. It is a critical prerequisite of the tasks of analyzing two or more relevant images. After the first full digit image registration was applied in remote sensing field [2], image registration becomes popular in remote sensing, which can be applied for environmental monitoring, change detection, image mosaic, and so on [3]–[6].

Manuscript received May 13, 2020; revised July 29, 2020; accepted August 5, 2020. Date of publication August 10, 2020; date of current version August 24, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61771353 and Grant 41501505, in part by Hubei Technology Innovation Project 2019AAA045, and in part by the Guangdong Provincial Department of Education 2017 "Innovation and Strong School Project" and Scientific Research Project: Natural Science Characteristic Innovation Project 2017GKTSCX014. *(Corresponding authors: Anna Dai; Tian Tian; Yuntao Wu.)*

Huabing Zhou, Anna Dai, Yulu Tian, Yuntao Wu, and Yanduo Zhang are with the Wuhan Institute of Technology, Wuhan 430205, China (e-mail: zhouhuabing@gmail.com; daidiana88521@gmail.com; lucytianyulu@163.com; ytwu@sina.com; zhangyanduo@hotmail.com).

Tian Tian is with the School of Computer Science, China University of Geosciences, Wuhan 430074, China (e-mail: tiantian@cug.edu.cn).

Zhenghong Yu is with the College of Robotics,Guangdong Polytechnic of Science and Technology, Zhuhai 519090, China (e-mail: honger1983@gmail.com).

Digital Object Identifier 10.1109/JSTARS.2020.3015350

The registration methods can be classified as area- and feature-based methods. As the geometric structure will not change in some situations, the feature-based methods are robust to the variety of illumination, different sensors, or noise. Besides, as remote sensing images have scaling changing between the pairs, we use the feature points to model the images. The general procedure of feature-based methods is feature extraction, feature matching, transformation function design, and image resampling.

In recent years, some methods have tended to locate feature points and find correspondences between them. RANSAC [7] is a classical method that can estimate the transformation function and then apply it to verify correspondences between the remaining points in the two sets; however, efficiency will decrease when the proportion of mismatches becomes high. Thus, Li and Hu [8] proposed the Identifying point correspondences by Correspondence Function (ICF) algorithm, and it can pick out the correspondences that satisfy the consistency of transformation function to reject mismatches. Some methods focus on the dissimilarity metric of the descriptor [9], [10], which utilize eigenvector correlation based on the eigenvector properties or signal directional differences.

However, the methods mentioned above could not get satisfactory results when the images suffer from nonrigid transformation. Therefore, some works introduce additional constraints to approximate the mapping function well. For example, vector field consensus (VFC) [11] algorithm, based on the Tikhonov regularization in vector-valued reproducing kernel Hilbert space, learns a vector field fitting for the inliers to remove outliers. Locally linear transforming [12] designs a local geometrical constraint to preserve local structures among neighboring feature points; it performs well even the outlier rate is too high. Kahaki *et al.* [13] use the Frobenius norm squared of the difference between the deformation gradient tensor and the identity matrix to regularize affine and B-spline transformations, which can progressively improve the registration accuracy with the increasing complexity of the transformations. This method mainly deals with the medical images suffering from a small view change, where the transformation estimating model need not be robust to a large number of outliers.

Although many methods have been proposed to solve the problem of image registration, the complex variety happens in the remote sensing image process making it difficult to get satisfying results. On the one hand, these methods only based on the local geometric constraint that regularizes the transformation function, and they do not fit the images suffering from local

distortion. On the other hand, the constraints are obtained from the putative correspondences under the similarity constraint, which have discarded the mismatches that may contain important information.

In this article, for solving the problem mentioned above, we propose a novel method based on manifold regularization. First, under the Bayes framework, we apply a joint model with a latent parameter to formulate the problem as the probability estimation, and then, we introduce the manifold regularization technology to explore the additional information about images to achieve the goal of improving performance; finally, we use three typical transformation models and combine them with the manifold regularization to approximate the mapping function well.

Our contribution in this article can be summarized in the following three aspects.

1) We propose a new method that can be applied in remote sensing image registration; unlike the other method that merely uses local features to approximate the transformation function, we introduce the manifold regularization, which is constructed by all feature points preserving more information about images.
2) We combine the manifold regularization term with three transformation models that obtain different mapping functions models—rigid, affine, and thin-plate spline (TPS) model, which are called MRR, MRA, and MR-TPS.
3) We derive the solutions of these three models and test their performance on four datasets; these models are general, and MR-TPS can get the best performance compared with the comparable state-of-the-arts methods and the other two models.

The remainder of this article is organized as follows. Section II describes the background material and related work. In Section III, we present our manifold regularization algorithm and apply it to the rigid, affine, and TPS feature matching. Section IV describes our experiment to show the performance of our algorithm and the result of comparing it with the other algorithm. Finally, Section V concludes this article.

II. Related Work

Image registration has a number of applications from computer vision [14], [15], medical imaging [16], [17], and pattern recognition [18]–[20] to remote sensing imaging [21]–[23]. The remote sensing image registration methods can be classified as area-based methods and feature-based methods. Area-based methods deal with the image rather than detecting the distinctive object but more focusing on the image intensity values. There are three main types of area-based methods: correlation-like methods [24], Fourier methods [25], and mutual information methods [26]. However, when there are sufficient salient and easily detected features in remote sensing images, the feature-based methods will be the better choice for small complexity.

A. Correspondence Construction

Our method is based on features. For establishing reliable correspondences, we should first detect the features and then, based on these features, find out the correspondences between them. According to the type of features used in remote sensing image registration, we can classify them as region features, line features, and point features. Region features are always the projection of the closed-boundary regions with high contrast, which are usually detected using segmentation methods [27]. Line features can be obtained from general line segments, object contours, coastal lines, or roads [28], [29]. Existing edge detection methods are used for line feature detection, such as Canny detector [30] or based on the Laplacian of Gaussian [31]. Because the region feature can be represented by their center of gravity, and the line feature can be represented by the pairs of line ends or middle points, point features are applied more widely. Moravec [32] is the first one who uses a set of local interest points in image matching; then, Harris and Stephens [33] improved the performance of the Moravec's method to make it more repeatable under small image size. In 2004, Lowe [34] proposed the SIFT descriptor, which assembles a high-dimensional vector representing the gradients of the image within a local region of the image to keep the distinctiveness of the point. SIFT is invariant to image rotation and scale and robust across a substantial range of affine distortion, addition of noise, and change in illumination.

Establishing a robust correspondence is a crucial step for image registration. Here are two steps for establishing the reliable point correspondences between remote sensing images [11], [35], [36]: first, computing a set of putative correspondences and then using the geometrical constraints to remove the outliers causing a bad influence on matching. During the first step, there are three methods to construct the correspondence, which use spatial relations, invariant descriptors, and relaxation methods. Besl and McKay [37] proposed the well-known iterative closest point algorithm for registering 3-D shapes, which minimize the distance between two point sets through rotation and translation. Li *et al.* [38] proposed using the SITF to overcome the difference in the gradient intensity and orientation between remote image pairs. In this way, we need to find out the correspondences that are highly similar to all possible correspondences and combine them with parametric geometrical models providing more information about the images to remove the outliers for improving the robustness of estimation. Fischler and Bolles [7] proposed a random sample consensus (RANSAC), which aims to solve the estimation of a model for the given set. Through hypothesizing and verifying, it resamples to obtain the smallest outlier subset to estimate the parameters of the model. There are many variant algorithms based on RANSAC that have been proposed, such as MLESAC [39], LO-RANSAC [40], and PROSAC [41].

B. Manifold Regularization

Regularization was first proposed by Tikhonov in 1963 [42], which aims to solve the ill-posed inverse problems. Considering that most of the image problems are of the same type, many researchers try to apply this in solving image problems [43]–[45]. Ma *et al.* [46] proposed moving regularized least squares (MRLS) which based on the moving least squares (MLS) [47] introduced the regularization in reproducing kernel Hilbert space (RKHS) to get a closed-form solution and a detail-preserving deformation. Zhou *et al.* [48] proposed MRLS-TPS, which

models the deformation function by a nonrigid TPS function with a regularization coefficient that can satisfy global linear affine transformation and local nonrigid deformation to get realistic deformation. Manifold regularization was proposed by Belkin *et al.* [49], which is used for reducing the overfitting in machine learning and for ensuring that a problem is well-posed by penalizing complex solutions. Ma *et al.* [50] cast the point registration into a semisupervised learning problem, where a set of indicator variables is adopted to help distinguish outliers in a mixture model; they constrain the transformation with manifold regularization to play a role of prior knowledge to preserve the intrinsic geometry structure of the input data. Zhou *et al.* [51] applied manifold regularization as a prior on the feature-guided MLS transformation and get a fast deformation in the 2-D image and 3-D surface. Considering the similarity of image registration and image deformation, we apply the manifold regularization to get better performance.

III. Method

In our method, we use the feature detection algorithm to locate the feature points $X_M = \{x_i\}_{i=1}^m$ and $Y_N = \{y_j\}_{j=1}^n$, which come from the reference and sensed image, respectively. Both of them are 2-D column vectors and denote the positions of feature points, where $m > n$. After comparing the similarity of feature points, we can obtain a set of putative correspondences containing the outliers. In theory, the set of inliers could approximately fit the mapping function of the image pairs well. In our method, we get the transformation function based on the inlier and then use this transformation function to verify whether the correspondence is an inlier. The process of our method can be summarized as follows: 1) establish a putative correspondence set based on rough matching; 2) determine inliers according to the transformation function; 3) update the transformation function according to inliers; and 4) iterate second and third steps until the algorithm convergence.

As shown in Fig. 1, the first row shows the correct matches on the images, the left is the reference image, and the right is the sensed one. For visuality, only 50 randomly selected pairs of correspondences are shown. The second and third rows are the feature points extracted from the images, where the black points have putative correspondences whose corresponding point has been marked by the blue line; the red points on the reference image are the points that cannot find the corresponding points in the sensed image. The difference between the second and third rows is that the reference image of the third row has gone through the transformation. Thus, the reference image and the sensed image are in the same coordinate system; that is the reason why the blue lines in the third row are parallel. We can observe from point distributions of the second and third rows that both the global features and the local details of the point distribution are preserved under the transformation, which was estimated by our method.

A. Problem Formulation

To achieve the goal of distinguishing the inliers and outliers, we apply a joint distribution to make the complicated distribution

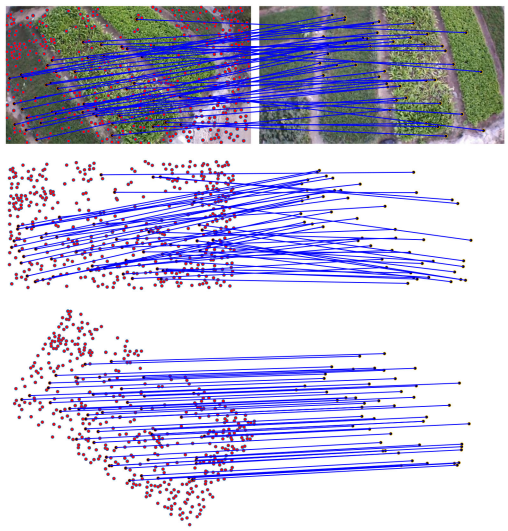


Fig. 1. Overview of the approach. The first row shows the correct matches on the images, the left is the reference image, and the right is the sensed one. The second and third rows are the feature points extracted from the images before and after transformation, respectively, where the black points are putative correspondences whose corresponding point has been labeled by the blue line, and the red points on the reference image are those that cannot find the corresponding points in the sensed image. For visibility, only 50 randomly selected matches are presented in the figure. (Observing from the third row, the blue lines are parallel, which means these feature points are in the same coordinate and keep the same relative distance.)

to be formed from more straightforward and more tractable components. More specifically, we use the Gaussian mixture model. Assuming that we have obtained N pairs of correspondence $S = \{(x_n, y_n)\}_{n=1}^N$,, which contain inliers and outliers and $x_n \in x_i, y_n \in y_j$. We use a latent parameter $z_n \in \{0, 1\}$ to denote the nature of the correspondence $s_n = (x_n, y_n)$, i.e.,

$$z_n = \begin{cases} 1, & \text{if } (x_n, y_n) \text{ is inlier} \\ 0, & \text{if } (x_n, y_n) \text{ is outlier} \end{cases}. \tag{1}$$

As the outliers are disordered, we think they follow the uniform distribution. In contrast, inliers follow the isotropic Gaussian distribution with zero mean and covariance $\sigma^2 I$, and I is the identity matrix. Since inliers have the character that they satisfy the transformation function f, the model can be expressed as

$$\begin{aligned} p(y_n|x_n, \theta) &= \sum_{z_n} p(y_n, z_n|x_n, \theta) \\ &= p(z_n = 1)p(y_n|x_n, \theta, z_n = 1) \\ &\quad + p(z_n = 0)p(y_n|x_n, \theta, z_n = 0) \\ &= \frac{\gamma}{2\pi\sigma^2} e^{\frac{\|y_n - f(x_n)\|^2}{2\sigma^2}} + \frac{1-\gamma}{a} \end{aligned} \tag{2}$$

where γ indicates the probability of inliers, i.e., $p(z_n = 1) = \gamma$, a is the parameter of the uniform distribution, which is decided by the range of y_n, and $\theta = \{f, \sigma^2, \gamma\}$ contains all the unknown parameters that are needed to be solved. We design two $N \times 2$ matrices $X = (x_1, \ldots, x_N)$ and $Y = (y_1, \ldots, y_N)$ to denote the sets of feature points. Then, the image registration turns to solve the problem of probability estimation. Under

the assumption of independent identically distributed (i.i.d.) data distribution, our aim is to solve a likelihood function $p(Y|X,\theta)=\prod_{n=1}^{N}p(y_n|x_n,\theta)$ and obtain the set of parameters $\theta^*=\text{argmax}_\theta p(Y|X,\theta)$. To simplify the calculation, we add the log manipulation on the function and rewrite the energy function

$$E(\theta)=-\ln p(Y|X,\theta)=-\sum_{n=1}^{N}\ln p(y_n|x_n,\theta). \tag{3}$$

The expectation–maximization (EM) algorithm is a general technique to solve this type of problem, which consists of the latent variables and alternates with the expectation step (E-step) and the maximization step (M-step). In the E-step, we fix the parameter set θ to estimate the range of the inlier, and then, in the M-step, we update θ according to the current estimation. Following the standard notation and omitting the terms independent of θ, we can simplify the function and obtain the complete-data log-likelihood function such as

$$\begin{aligned}Q(\theta,\theta^{\text{old}})=&-\frac{1}{2\sigma^2}\sum_{n=1}^{N}p_n\parallel y_n-f(x_n)\parallel^2-\ln\sigma^2\sum_{n=1}^{N}p_n\\&+\ln\gamma\sum_{n=1}^{N}p_n+\ln(1-\gamma)\sum_{n=1}^{N}(1-p_n)\end{aligned} \tag{4}$$

where $p_n=P(z_n=1|x_n,y_n,\theta^{\text{old}})$ is a posterior probability indicating the intensity of (x_n,y_n) being an inlier.

E-step: We set a diagonal matrix P, which is described as $P=\text{diag}(p_1,\ldots,p_N)$, and based on the Bayes rule, we fix the current parameter set θ^{old} to calculate p_n for every correspondences, which is written as

$$p_n=\frac{\gamma e^{-\frac{\|y_n-f(x_n)\|^2}{2\sigma^2}}}{\gamma e^{-\frac{\|y_n-f(x_n)\|^2}{2\sigma^2}}+\frac{2\pi\sigma^2(1-\gamma)}{a}}. \tag{5}$$

M-step: In this step, according to $\theta^{\text{new}}=\text{argmax}_\theta Q(\theta,\theta^{\text{old}})$, we update the set of parameters, taking derivatives of Q with respect to σ^2 and γ, and setting to zero; then, we obtain the function

$$\begin{aligned}\sigma^2&=\frac{\text{tr}((Y-f(X))^TP(Y-f(X)))}{2\cdot\text{tr}(P)}\\\gamma&=\frac{\text{tr}P}{N}.\end{aligned} \tag{6}$$

Until the EM iteration converges, we can distinguish the inlier and outlier through a predefined threshold t, where the set of inliers IS satisfies

$$IS=\{(x_n,y_n):p_n>\tau,n\in\mathbb{Z}\}. \tag{7}$$

B. Manifold Regularization

Observing the procedure of image registration, we can find that the transformation function can have a great influence on image registration. Through applying the geometric constraint, we can obtain additional information about the images to get better results. However, putative correspondences do not contain all the feature points, which will cause loss of information. The manifold regularization is composed of all feature points and can exploit the intrinsic structure of feature points as a geometric constraint. Therefore, an additional regularization term is introduced to the energy function

$$E(f)=\parallel P^{1/2}(Y-f(X))\parallel_F^2+\lambda\phi_M. \tag{8}$$

Graph Laplacian is a discrete analog of the manifold Laplacian. Under the assumption that the input points are drawn i.i.d. from the manifold, we use the weighted neighborhood graph for the data. $\{x_i\}_{i=1}^m$ are the feature points in reference image with edges (x_i,x_j) if and only if $\parallel x_i-x_j\parallel^2<\varepsilon$, assigning to edge (x_i,x_j) the weight

$$G_{ij}=e^{-\frac{1}{\varepsilon}\|x_i-x_j\|^2}. \tag{9}$$

The graph Laplacian is the matrix L which is given by

$$L_{ij}=\Lambda_{ij}-G_{ij} \tag{10}$$

where $\Lambda=\text{diag}(\sum_{j=1}^{m}G_{ij})_{i=1}^{m}$ is the diagonal matrix whose ith entry is the sum of the weights of edges leaving x_i. Let $\Gamma=(f(x_1),\ldots,f(x_m))^T$; the manifold regularization term ϕ_M can be defined as

$$\phi_M(f)=\sum_{i=1}^{n}\sum_{j=1}^{m}G_{ij}(\Gamma_i-\Gamma_j)^2=\text{tr}(\Gamma^TL\Gamma) \tag{11}$$

where tr$(\cdot)$ denotes the trace. Substituting (11) into (8), we can get the following energy function:

$$E(\Gamma)=\parallel P^{1/2}(Y-\Gamma)\parallel_F^2+\lambda\text{tr}(\Gamma^TL\Gamma) \tag{12}$$

where $\lambda>0$ controls the degree of the influence that the manifold regularization have on the final result. There is no doubt that if λ is too large, it will cause a bad result, and if λ is too small, it cannot have obvious improvement in the algorithm.

From the construction of the manifold regularization term, we can find that it depends on the input data, and as a prior term, it will not increase the difficulty of calculation. On the one hand, using the manifold regularization term can add additional information about the images that can greatly improve the accuracy of estimating the parameters in the transformation model. On the other hand, the manifold regularization can control the complexity of the transformation. Applying the manifold regularization in different types of transformation model, we have different solution of parameters. In the next part, we will describe the ways to solve the problems on different models, which contain the rigid model, affine model, and TPS model.

C. Rigid Model Matching

The rigid model is defined as [12]: $\Gamma(x_n)=sRx_n+t$, where R is a 2×2 orthogonal rotation matrix, t is a 2×1 translation vector, and s is a 2×1 arbitrary scaling parameter. The manifold regularization term for the rigid transformation can be written as

$$\begin{aligned}\phi_M&=\sum_{i=1}^{M}\sum_{j=1}^{M}L_{ij}(sRx_i-sRx_j)^2\\&=s^2\sum_{i=1}^{M}\sum_{j=1}^{M}L_{ij}(Rx_i-Rx_j)^2\\&=s^2\text{tr}((RX_M)^TL(RX_M)).\end{aligned} \tag{13}$$

Combining with (12), the error function can be written as

$$E(R,t,s) = \| P^{\frac{1}{2}}(Y - sRX - t) \|^2 + \lambda \text{tr}(s^2 {X_M}^T L X_M)$$
$$\text{s.t. } R^T R = I, \det(R) = 1. \tag{14}$$

Considering the character of R, we cannot directly derive E with R and set to zero to get the closed-form solution. Thus, we use the following theorem.

Theorem 1: Let R be an unknown $D \times D$ rotation matrix and B be a known $D \times D$ real square matrix. Let USV^T be a singular value decomposition of B, where $UU^T = VV^T = I$ and $S = \text{diag}(s_i)$ with $s_1 \leq \cdots \leq s_D \leq 0$. Then, the optimal rotation matrix R that maximizes $\text{tr}(B^T R)$ is $R = UDV^T$, where $D = \text{diag}(1, \ldots, 1, \det(UV^T))$.

To solve the problem, we should eliminate the translation parameter t at first. Taking the derivation of E with respect to t and setting it to zero, we can get

$$t = \frac{1}{\text{tr}(P)} Y^T P 1 - \frac{1}{\text{tr(P)}} sRX^T P 1 = \mu_y - sR\mu_x \tag{15}$$

where μ_x and μ_y are the mean vectors defined as

$$\mu_x = \frac{1}{\text{tr}(P)} X^T P 1$$
$$\mu_y = \frac{1}{\text{tr}(P)} Y^T P 1. \tag{16}$$

By substituting t back into the function (14) and omitting the terms that are independent of R and s, we obtain

$$E(R,s) = \text{tr}(s^2 \hat{X}^T P \hat{X} - 2s\hat{Y}^T P \hat{X} R^T) + \lambda \cdot \text{tr}(s^2 {X_M}^T L X_M) \tag{17}$$

where $\hat{X} = X - 1{\mu_x}^T$ and $\hat{Y} = Y - 1{\mu_x}^T$ are the centered point matrices. Preserving the term related to R, we obtain

$$E(R) = \text{tr}((\hat{Y}^T P \hat{X})^T R). \tag{18}$$

According to Theorem 1, we can get the optimal solution of R

$$R = UDV^T \tag{19}$$

where U and V come from $USV^T = \text{svd}(\hat{Y}^T P \hat{X})$, and $D = \text{diag}(1, \det(UV^T))$. Then, we set (17) to zero and obtain

$$s = \frac{2\text{tr}((\hat{Y}^T P \hat{X})^T R)}{\text{tr}(\hat{X}^T P \hat{X}) + \lambda \text{tr}({X_M}^T L X_M)}. \tag{20}$$

Until now, all the parameters in the M-step have been solved. We summarize our MR algorithm for the rigid model in Algorithm 1.

D. Affine Model Matching

The affine model is defined as [12]: $\Gamma(x_n) = Ax_n + t$, where A is a 2×2 affine matrix, and t is a 2×1 translation vector. Combining the manifold regularization term, we obtain

$$E(A,t) = \| P^{1/2}(Y - AX - t) \|^2 + \lambda \text{tr}((AX_M)^T L (AX_M)). \tag{21}$$

Algorithm 1: The MR Algorithm for Rigid Model.

Input: Correspondences set $S = \{(x_n, y_n : n \in N)\}$, whole feature points set X_M, parameters λ, ε, τ

Output: Inlier set IS

```
Initialize γ, t = 0, P = I_{N×N};
Set a to the volume of the output space;
Initialize σ^2 by Eq.(6);
Compute the L of the reference image by Eq (9);
repeat
    E-step:
    Update P by Eq.(5);
    M-step:
    Compute X̂ and Ŷ accroding to Eq.(16);
    Compute USV^T = svd(ŶPX̂);
    Update R, s, t by Eqs.(18)(19) and (20);
    Update σ^2 and γ by Eq.(6);
until Q converges;;
The consensus set IS is determined by Eq.(7).
```

Algorithm 2: The MR Algorithm for Affine Model.

Input: Correspondences set $S = \{(x_n, y_n : n \in N)\}$, whole feature points set X_M, parameters λ, ε, τ

Output: Inlier set IS

```
Initialize γ, t = 0, P = I_{N×N};
Set a to the volume of the output space;
Initialize σ^2 by Eq.(6);
Compute the L of the reference image by Eq (9);
repeat
    E-step:
    Update P by Eq.(5);
    M-step:
    Compute X̂ and Ŷ accroding to Eq.(16)
    Update A, t by Eq.(22);
    Update σ^2 and γ by Eq.(6);
until Q converges;;
The consensus set IS is determined by Eq.(7).
```

The solution of t is similar to the rigid model, and the solution of A can be obtained directly by taking a partial derivative of E and setting it to zero. Then, t and A take the following form:

$$t = \mu_y - A\mu_x$$
$$A = (\hat{Y}^T P \hat{X})(\hat{X}^T P \hat{X} + \lambda {X_M}^T L X_M)^{-1}. \tag{22}$$

Until now, all the parameters of affine transformation in the M-step have been solved. We summarize our MR algorithm for affine matching in Algorithm 2.

E. TPS Model Matching

TPS is a spline-based technique for smoothing and has a closed-form solution; thus, we introduce the TPS model to approximate the mapping function well. It is a nonrigid model that contains a global affine transformation and a local bending

function and has the following form [52]:

$$f(x) = A\bar{x} + g_x(x)$$
$$g_x(x) = \sum_{i=1}^{m} K(x, x_i)\bar{c_i} \tag{23}$$

where A is a 3×3 affine matrix, and $\bar{x}$ is a homogeneous coordinate matrix defined as $(x^T, 1)$. The TPS has a natural representation in terms of radial basis functions; thus, we define K (a radial basis function) as a TPS kernel matrix

$$\phi(r) = r^2 \log r$$
$$K(x, x_i) = \| x - x_i \|^2 \log \| x - x_i \|. \tag{24}$$

We design $C = (\bar{c_1}, \ldots, \bar{c_m})^T$, which is an $M \times 3$ bending coefficient matrix. As shown in [48], we place the TPS model in the RKHS—$\mathcal{H}$; thus, the regularization term ϕ has the following form:

$$\phi(f) = \| f \|_{\mathcal{H}} \tag{25}$$

where $\| \cdot \|_{\mathcal{H}}$ denotes the norm of space $\mathcal{H}$. We can get the regularization term

$$\| f_x \|_H^2 = \sum_{i=1}^{M}\sum_{i=1}^{M} < K(x_i, x_j)\bar{c_i}, \bar{c_j} > = \text{tr}(C^T K C). \tag{26}$$

Based on the solution of [48], the affine calculation and bending calculation can be divided into two parts. After combining (12), (23), and the manifold regularization term, the error function can be written as

$$\begin{aligned} E(A, C) = & \| P^{1/2}(\bar{Y} - \bar{X}A - JKC) \|^2 + \lambda_1 \text{tr}(C^T K C) \\ & + \lambda_2 \text{tr}((\bar{X}_M A^T)^T L (\bar{X}_M A^T)) \\ & + \lambda_3 \text{tr}((KC)^T L (KC)) \end{aligned} \tag{27}$$

where $\bar{Y} = (\bar{y}_1, \ldots, \bar{y}_n)^T$, $\bar{X} = (\bar{x}_1, \ldots, \bar{x}_n)^T$, and $\bar{X}_M = (\bar{x}_1, \ldots, \bar{x}_m)^T$. To obtain the solutions of A and C, we set

$$\tilde{Y} = P^{1/2}\bar{Y}$$
$$\tilde{X} = P^{1/2}\bar{X} \tag{28}$$

and we use QR decomposition on matrix $\tilde{X}$:

$$\tilde{X} = [Q_1 | Q_2] \begin{bmatrix} R \\ 0 \end{bmatrix} \tag{29}$$

where Q_1 and Q_2 are the orthogonal matrixes of $N \times N$ and $N \times (N-3)$, respectively. R is a 3×3 upper triangular matrix. Assuming that $C = Q_2 \tilde{C}$, $\tilde{C}$ is an $(N-3) \times 3$ matrix; then, we obtain (30):

$$\begin{aligned} E(A, \tilde{C}) = & \| Q_1^T \tilde{Y} - RA^T - Q_1^T P^{\frac{1}{2}} JKC \|^2 \\ & + \| Q_2^T \tilde{Y} - Q_2^T P^{\frac{1}{2}} JKC \|^2 \\ & + \lambda_1 \text{tr}(\tilde{C}^T Q_2^T K Q_2 \tilde{C}) + \lambda_2 \text{tr}((\bar{X_M} A)^T L (\bar{X_M} A)) \\ & + \lambda_3 \text{tr}((KC)^T L (KC)). \end{aligned} \tag{30}$$

Algorithm 3: The MR Algorithm for TPS Model.

Input: Correspondences set $S = \{(x_n, y_n : n \in N)\}$, whole feature points set X_M, parameters λ, ε, τ
Output: Inlier set IS

```
Initialize γ, C = 0, A = I_{3×3}, P = I_{N×N};
Set a to the volume of the output space;
Initialize σ^2 by Eq.(6);
Construct the Kernel matrix K using the definition of K;
Compute the L of the reference image;
repeat
    E-step:
    Update P by Eq.(5);
    M-step:
    Update A, C by Eqs.(32);
    Update σ^2 and γ by Eq.(6);
until Q converges;;
The consensus set IS is determined by Eq.(7).
```

Minimizing the function (31), the results of $\tilde{C}$ and A are

$$\begin{aligned} C &= Q_2 \tilde{C} = (S^T S + \lambda_1 K + \lambda_3 K L K)^{-1} S^T \tilde{Y} \\ A &= (\tilde{Y} - SC)^T Q_1 R (R^T R + \lambda_2 \bar{X_M}^T L \bar{X_M})^{-T} \end{aligned} \tag{31}$$

where $S = P^{\frac{1}{2}} JKC$. Until now, we summarize the parameters in the M-step of the MR algorithm for the TPS model in Algorithm 3.

F. Implementation Details

The whole feature matching process is based on the feature points that are extracted by the SIFT algorithm, and through comparing the similarity of the descriptors, we can obtain rough matching and a set of putative correspondences, which represent the coordinate system. We use data normalization to control them and use linear rescaling to make both two feature point sets have zero mean and unit variance. Obviously, the constant of uniform distribution in (1) should be set according to the data normalization.

Parameter setting: There are mainly six parameters in our method: ε, λ, λ_1, λ_2, λ_3, τ, and γ. The parameter ε controls the construction of the manifold regularization term for the distance between the points larger than ε; we think there is a line between the points. The parameter λ controls the influence of the geometrical constraint on the transformation Γ of the rigid model and the affine model. Parameters λ_1, λ_2, and λ_3 are the parameters in the TPS model; all of them control the transformation complexity. Parameter τ is a threshold; it controls the intensity of the correspondence to be an inlier. Parameter γ is the initial assumption of the ratio of the inlier and the rough matching.

We set $\varepsilon = 0.08$, $\lambda = 9\,000\,000\,000$, $\lambda_1 = 100\,000$, $\lambda_2 = 0.001$, $\lambda_3 = 100$, $\tau = 0.75$, and $\gamma = 0.9$ throughout this article.

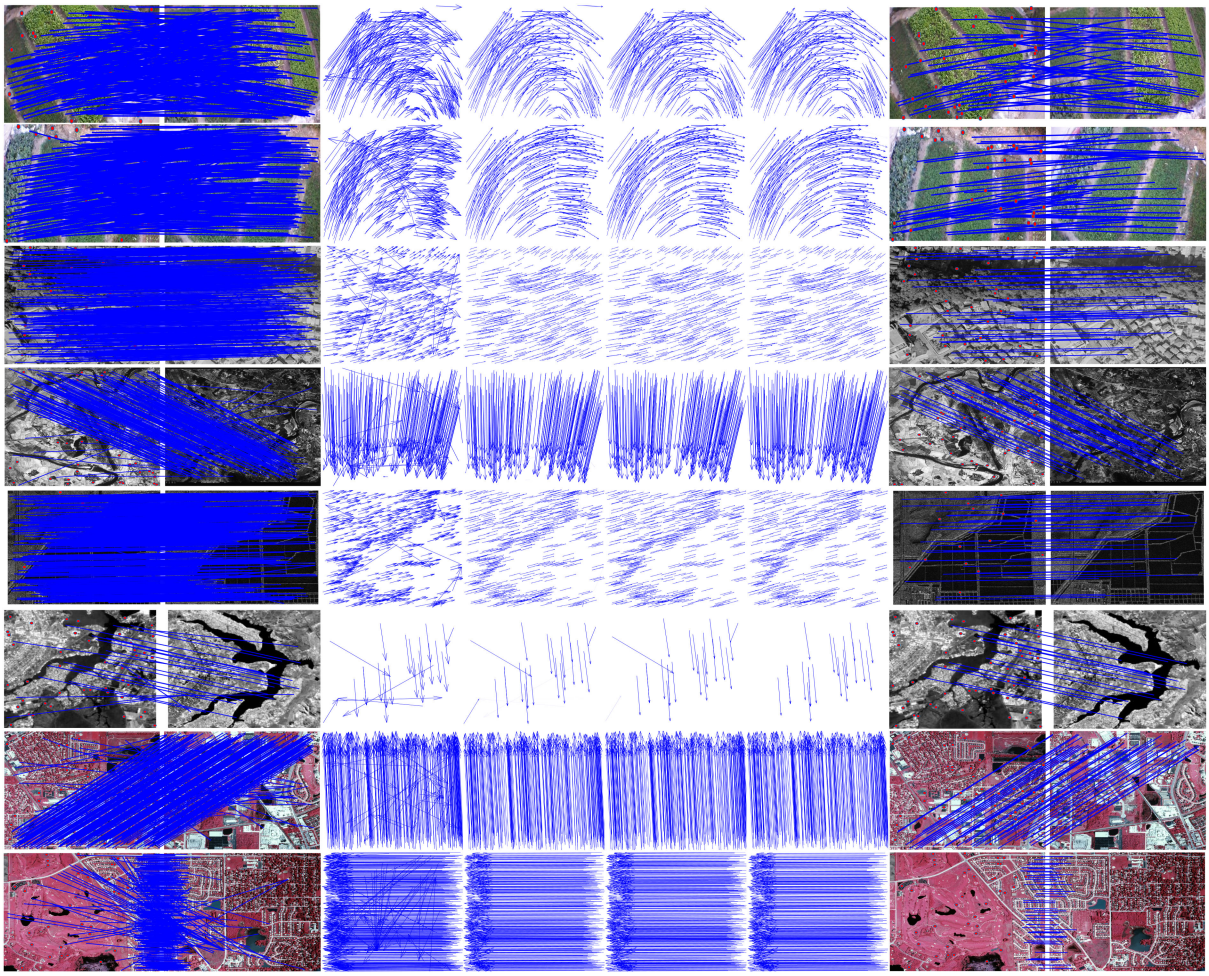


Fig. 2. Qualitative matching results on several typical remote sensing image pairs. The first and second rows are the results of MRTPS on the UAV dataset. The third and fourth rows are the results of MRTPS on the PAN dataset. The fifth and sixth rows are the results of MRA on the SAR dataset. The final two rows are the results of MRR on the CIAP dataset. The columns show the process of the EM algorithm; the intensity of the blue lines indicates the probability of being the inliers. The red points on the first and last columns are the feature points that not participate in the matching but control the transformation of the reference images.

IV. Experiment Results

In this section, we test the performance of our algorithm in sets of images, which contain different types of remote sensing images, and compare our algorithm with the other algorithms applied in feature matching, such as RANSAC [7], ICF [8], GS [53], and VFC [11].

A. Datasets and Settings

We test our algorithm on four different types of remote sensing image pairs, which suffer from different types of transformations and have different problems on feature matching.

1) *UAV:* This dataset contains 35 all-color image pairs, which are captured from a piece of farmland by an unmanned aerial vehicle (UAV) and are used for the automatic crop monitoring task. These are high-quality images, and the observation objection on the images is pure, just containing the grassland. However, intensive planting makes the physical correspondences become not similar at all under a small view change. Most of the images suffer from projection distortion, and their size is 600×337.

2) *PAN:* This dataset contains 21 image pairs captured from Tokyo, Japan, and Wuhan, China. These images are panchromatic (PAN) aerial photographs and show the look of these places at a different time and suffer from a large view changes or ground relief variations such as mountains and buildings. These challenges increase the difficulty of feature matching because the linear transformation function cannot model the transformation of these images. The feature matching task for such image pairs typically arises in change detection. Sizes of the images range from 600×700 to 700×700.

3) *SAR:* This dataset contains 33 SAR image pairs that were captured over Nantong, Jiangsu Province, China. The SAR image pairs come from two different SARs loaded, respectively, on a satellite and a UAV. The difficulty of this dataset is that the SAR imaging process produces noise, which degrades the quality of images. In addition, two different sensors also make some problems with feature matching. These images are used in the positioning and navigating problem, such as solving the matching of a real-time UAV image and corresponding stored satellite image, and obtain an accurate current position. The

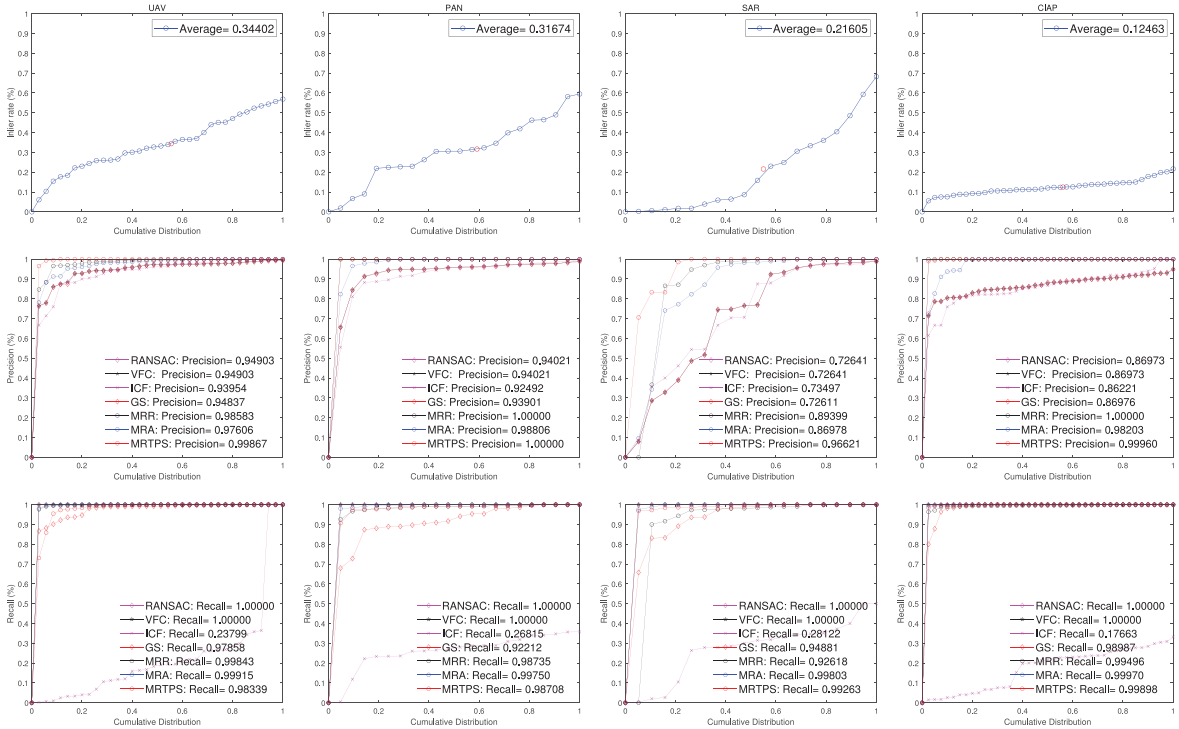


Fig. 3. Quantitative comparisons of RANSAC [7], ICF [8], GS [53], and VFC [11] on the four test datasets: UAV, PAN, SAR, and CIAP. *Top to bottom:* initial inlier ratio, precision, and recall with respect to the cumulative distribution.

images have a size of 800×800, and they, in general, suffer from affine distortions.

4) *CIAP:* This dataset contains 39 CIAP image pairs with small overlap areas. The transformation model in these images is rigid, and all of them have size of 700×700 and have been already orthorectified. The feature matching task for such image pairs typically arises in the image mosaic problem. The images are publicly available (from the Erdas example data), which were captured over eastern Illinois, IL, USA.

We apply the SIFT algorithm, based on the open-source VLFEAT toolbox, to construct the rough matching. We use the precision rate and the recall rate to demonstrate the performance of the algorithms, as follows:

$$\text{Precision} = \frac{\#\text{identified correct matches}}{\#\text{preserved matches}}$$
$$\text{Recall} = \frac{\#\text{identified correct matches}}{\#\text{ground truth}}. \tag{32}$$

We set the ground truth in such a way that we have made a benchmark before conducting any experiments to ensure objectivity; specifically, each putative correspondence in each image pair is checked manually. Experiments are performed on a laptop with 3.4-GHz Intel Core i7 CPU, 16-GB memory, and MATLAB core.

B. Qualitative Results

In this section, we show our result on different types of image data. The first two rows are the UAV image pairs, which suffer from the large view change, resulting in that some physical correspondences are not similar at all. However, with the guidance of the manifold regularization, we can find these correspondences. The third and fourth rows are the PAN image pairs; the ground relief variations in them increase the difficulty of matching. The next two rows are the SAR image pairs; they come from different sensors, and the imaging process produces noise on images. In the last two rows, there are CIAP image pairs, in which the challenge is a small overlap area. All the problems, appearing in matching these image pairs, are typical that need to be solved in feature matching. We use our algorithm to establish accurate feature correspondences.

As shown in Fig. 2, the blue lines drawn in the first column are the rough matches found out by the SIFT algorithm, and the middle columns are the iterative process of our algorithm, using the EM algorithm to calculate the degree of p_n. We use the intensity of lines to indicate the probability of the points to be inliers and the arrow to show the direction of point moving. We can find that our algorithm converges quickly and almost at the fifth iteration, whose results are similar to the final results. The red points, on the first and final columns, are the feature points, which cannot find corresponding points in the sensed image; they work as a guide for N pairs of correspondences. And the blue lines on the last column are the final result of correspondences found out by our algorithm. For the visuality of the matching process, the correspondences are shown in the results.

C. Quantitative Results

We compare our algorithms with other classical feature matching methods, which are RANSAC [7], ICF [8], GS [53], and VFC [11]. They are tested on four different types of remote sensing image data: UAV, PAN, SAR, and CIAP.

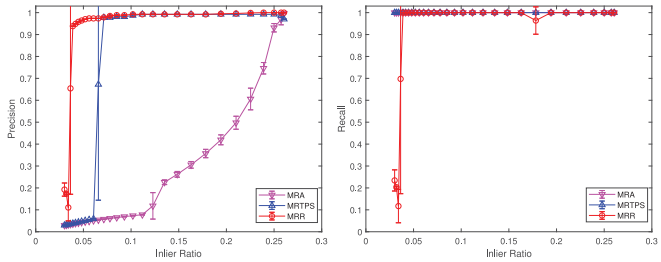


Fig. 4. Robustness tests of our MR algorithm. Each column is a group of results, where (a) shows the precision curves and (b) shows the recall curves. The red curves represent the results of MR with a rigid model, the pink curves represent the results of MR with the affine model, and the blue curves represent the results of MR with the TPS model. In the experiment, we fix the inlier number and vary the inlier ratio. The error bars indicate the precision/recall means and standard deviations over ten trials.

We can observe from Fig. 3 that all of the methods get the worst performances in the SAR dataset. The reason is that the noise on the SAR images degrades the quality of the image, which increases difficulty in matching. Most SAR images suffer from affine distortion, which is a complex transformation. By contrast, the best results are obtained on the CIAP and PAN, even if there are some challenges such as small overlap area and ground relief variation, but they suffer from the rigid transformation, which is a relatively simple transformation.

Observing from the character of each algorithm, ICF tends to get high precision and low recall because ICF is the iterative calculation in the neighbor of the points. RANSAC's results are highly affected by the inlier rate; the higher the inlier, the higher the precision on the RANSAC. Therefore, a low inlier rate cannot provide sufficient information to approximate the model of the image. VFC and GS have a better precision–recall tradeoff compared with ICF, but the matching result is still worse than our algorithms. Although the RANSAC and VFC have a high recall, the low precision means that they cannot distinguish between the inliers and the outliers.

Considering our algorithms, i.e., MRR, MRA, and MRTPS, we can find that MRTPS almost gets the best performance on all the datasets. TPS is a nonrigid transformation, which is composed of a global affine transformation and a local distortion; that is the reason why TPS is more general, and most of the remote sensing images have local distortion. Comparing MRR and MRA, MRA is based on the affine transformation, which is more complex; thus, MRR can get better performance than MRA.

D. Robust Test

We test the robustness of MRA, MRR, and MRTPS. Considering that the inlier is one of the most important factors in performance, we design the following experiment. We fix the inlier number, vary the percentage of the inlier, and then test the performance of our methods.

Fig. 4 shows the results of our methods' robustness. We fix the number of inliers to 259 and then vary the inlier ratio from 0.3 to 0. Fig. 4(a) shows the precision curves, and Fig. 4(b) shows the recall curves. The blue, red, and pink curves represent the results of MRTPS, MRR, and MRA, respectively. From the results, we find that the performance of MRA becomes better and stable as the inlier ratio increases and that MRR and MRTPS are stable when the inlier ratio increases above 0.1.

V. Conclusion

In this article, we proposed a novel mismatch removal method for robust feature matching of remote sensing images. We improved the performance of the algorithm through reliable spatial relationships, the manifold regularization term, which is used as the geometrical constraint containing the global feature and local details. We formulated the feature matching problem as a maximum likelihood estimation and use the EM algorithm to get the closed-form solution. We combined the manifold regularization term with different transformation models and tested on four different types of datasets. The results show that our algorithms are general in the dataset and can get higher precision and recall compared with other methods.

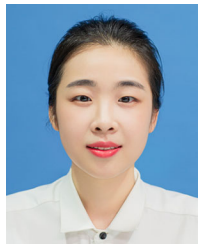
References

[1] B. Zitova and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, 2003.
[2] A. A. Goshtasby, *Image Registration: Principles, Tools and Methods*. Berlin, Germany: Springer, 2012.
[3] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, "Image matching from handcrafted to deep features: A survey," *Int. J. Comput. Vis.*, Aug. 2020, doi: 10.1007/s11263-020-01359-2.
[4] J. Fan, Y. Wu, M. Li, W. Liang, and Q. Zhang, "SAR image registration using multiscale image patch features with sparse representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 4, pp. 1483–1493, Apr. 2016.
[5] K. Yang, L. Karlstrom, L. C. Smith, and M. Li, "Automated high-resolution satellite image registration using supraglacial rivers on the greenland ice sheet," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 3, pp. 845–856, Mar. 2016.
[6] T. Hu, H. Zhang, H. Shen, and L. Zhang, "Robust registration by rank minimization for multiangle hyper/multispectral remotely sensed imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2443–2457, Jun. 2014.
[7] 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, 1981.
[8] X. Li and Z. Hu, "Rejecting mismatches by correspondence function," *Int. J. Comput. Vis.*, vol. 89, no. 1, pp. 1–17, 2010.
[9] S. M. M. Kahaki, M. J. Nordin, A. H. Ashtari, and S. J. Zahra, "Invariant feature matching for image registration application based on new dissimilarity of spatial features," *Plos One*, vol. 11, no. 3, 2016, Art. no. e0149710.
[10] S. M. M. Kahaki, A. Haslina, N. M. Jan, I. Waidah, and Z. D. Capella, "Geometric feature descriptor and dissimilarity-based registration of remotely sensed imagery," *Plos One*, vol. 13, no. 7, 2018, Art. no. e0200676.
[11] 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.
[12] 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.
[13] S. M. M. Kahaki, S.-L. Wang, and A. Stepanyants, "Accurate registration of *in vivo* time-lapse images," *Proc. SPIE*, vol. 10949, 2019, Art. no. 109491D.
[14] Y. Liu, "Automatic range image registration in the Markov chain," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 12–29, Jan. 2009.
[15] C. Yao, X. Bai, and W. Liu, "A unified framework for multioriented text detection and recognition," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4737–4749, Nov. 2014.

[16] J. Ma, J. Jiang, C. Liu, and Y. Li, "Feature guided Gaussian mixture model with semi-supervised em and local geometric constraint for retinal image registration," *Inf. Sci.*, vol. 417, pp. 128–142, 2017.
[17] G. Wang, Z. Wang, Y. Chen, Q. Zhou, and W. Zhao, "Removing mismatches for retinal image registration via multi-attribute-driven regularized mixture model," *Inf. Sci.*, vol. 372, pp. 492–504, 2016.
[18] J. Ma, J. Zhao, J. Jiang, H. Zhou, and X. Guo, "Locality preserving matching," *Int. J. Comput. Vis.*, vol. 127, no. 5, pp. 512–531, 2019.
[19] Y. Gao, J. Ma, and A. L. Yuille, "Semi-supervised sparse representation based classification for face recognition with insufficient labeled samples," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2545–2560, May 2017.
[20] H. Zhou, J. Ma, C. C. Tan, Y. Zhang, and H. Ling, "Cross-weather image alignment via latent generative model with intensity consistency," *IEEE Trans. Image Process.*, vol. 29, pp. 5216–5228, 2020.
[21] 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, 2017, Art. no. 581.
[22] C. Chen, Y. Li, W. Liu, and J. Huang, "SIRF: Simultaneous satellite image registration and fusion in a unified framework," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4213–4224, Nov. 2015.
[23] J. Ma, Y. Ma, and C. Li, "Infrared and visible image fusion methods and applications: A survey," *Inf. Fusion*, vol. 45, pp. 153–178, 2019.
[24] 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.
[25] 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.
[26] 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, 1999.
[27] J. Flusser and T. Suk, "A moment-based approach to registration of images with affine geometric distortion," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 2, pp. 382–387, Mar. 1994.
[28] Y. C. Hsieh, D. M. McKeown, and F. P. Perlant, "Performance evaluation of scene registration and stereo matching for artographic feature extraction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 214–238, Feb. 1992.
[29] S. Moss and E. R. Hancock, "Multiple line-template matching with the EM algorithm," *Pattern Recognit. Lett.*, vol. 18, no. 11-13, pp. 1283–1292, 1997.
[30] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679–698, Nov. 1986.
[31] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. Roy. Soc. London. Ser. B, Biol. Sci.*, vol. 207, no. 1167, pp. 187–217, 1980.
[32] H. P. Moravec "Rover visual obstacle avoidance." in *Proc. 7th Int. Joint Conf. Artif. Intell.*, 1981, pp. 785–790.
[33] C. G. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. Alvey Vis. Conf.*, 1988, pp. 147–151.
[34] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[35] S. Pang, J. Xue, Q. Tian, and N. Zheng, "Exploiting local linear geometric structure for identifying correct matches," *Comput. Vis. Image Understanding*, vol. 128, pp. 51–64, 2014.
[36] 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.
[37] P. J. Besl and N. D. McKay, "Method for registration of 3-D shapes," *Proc. SPIE*, vol. 1611, pp. 586–606, 1992.
[38] 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.
[39] P. H. Torr and A. Zisserman, "MLESAC: A new robust estimator with application to estimating image geometry," *Comput. Vis. Image Understanding*, vol. 78, no. 1, pp. 138–156, 2000.
[40] O. Chum, J. Matas, and J. Kittler, "Locally optimized RANSAC," in *Proc. Joint Pattern Recognit. Symp.*, 2003, pp. 236–243.
[41] O. Chum and J. Matas, "Matching with PROSAC—Progressive sample consensus," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 220–226.
[42] A. N. Tikhonov, "On the solution of ill-posed problems and the method of regularization," *Doklady Akademii Nauk*, vol. 151, pp. 501–504, 1963.
[43] O. Dubovik and M. D. King, "A flexible inversion algorithm for retrieval of aerosol optical properties from sun and sky radiance measurements," *J. Geophys. Res., Atmos.*, vol. 105, no. D16, pp. 20673–20696, 2000.
[44] O. Dubovik *et al.*, "Accuracy assessments of aerosol optical properties retrieved from aerosol robotic network (AERONET) sun and sky radiance measurements," *J. Geophys. Res.: Atmos.*, vol. 105, no. D8, pp. 9791–9806, 2000.
[45] K. Miller, "Least squares methods for ill-posed problems with a prescribed bound," *SIAM J. Math. Anal.*, vol. 1, no. 1, pp. 52–74, 1970.
[46] J. Ma, J. Zhao, and J. Tian, "Nonrigid image deformation using moving regularized least squares," *IEEE Signal Process. Lett.*, vol. 20, no. 10, pp. 988–991, Oct. 2013.
[47] S. Schaefer, T. McPhail, and J. Warren, "Image deformation using moving least squares," *ACM Trans. Graph.*, vol. 25, pp. 533–540, 2006.
[48] H. Zhou *et al.*, "Image deformation with vector-field interpolation based on MRLS-TPS," *IEEE Access*, vol. 6, pp. 75886–75898, 2018.
[49] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learn. Res.*, vol. 7, no. 11, pp. 2399–2434, 2006.
[50] 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.
[51] H. Zhou, J. Ma, Y. Zhang, Z. Yu, S. Ren, and D. Chen, "Feature guided non-rigid image/surface deformation via moving least squares with manifold regularization," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2017, pp. 1063–1068.
[52] C. Yang, M. Zhang, Z. Zhang, L. Wei, R. Chen, and H. Zhou, "Non-rigid point set registration via global and local constraints," *Multimedia Tools Appl.*, vol. 77, no. 24, pp. 31 607–31 625, 2018.
[53] H. Liu and S. Yan, "Common visual pattern discovery via spatially coherent correspondences," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 1609–1616.

Huabing Zhou received the B.S. and M.S. degrees in computer science and technology from the Wuhan Institute of Technology, Wuhan China, in 2005 and 2008, respectively, and the Ph.D. degree in control science and engineering from the Huazhong University of Science and Technology, Wuhan, in 2012.

From 2009 to 2010, he was a Research Intern with the Chinese Academy of Surveying and Mapping. From 2018 to 2019, he was a Visiting Scholar with Temple University, Philadelphia, PA, USA. He is currently an Associate Professor with the School of Computer Science and Engineering, Wuhan Institute of Technology. His research interests include computer vision, remote sensing image analysis, and intelligent robots.

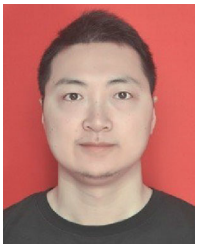


Anna Dai was born in Hubei, China, in 1995. She received the B.S. degree in computer science and technology, in 2018 from the School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan, China, where she is currently working toward the M.S. degree in computer vision, image process, and machine learning.

Tian Tian received the B.S. degree in electronic information engineering and Ph.D. degree in control science and engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2009 and 2015, respectively.

From 2012 to 2014, she visited Oakland University, Rochester, MI, USA, as a Ph.D. student sponsored by the China Scholarship Council. In 2015, she joined the School of Computer Sciences, China University of Geosciences, Wuhan, as a Postdoctoral Lecturer and is currently an Associate Professor with the same school. Her major research interests include remote sensing image processing and computer vision and its applications.

Yulu Tian was born in Hubei, China, in 1995. She received the B.S. degree in computer science and technology, in 2018 from the School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan, China, where she is currently working toward the M.S. degree in computer vision, image process, and machine learning.



Zhenghong Yu received the B.S. and M.S. degree in computer science from the Wuhan Institute of Technology, Wuhan, China, in 2005 and 2008, respectively, and the Ph.D. degree in control science and engineering from the Huazhong University of Science and Technology, Wuhan, in 2014.

He is currently an Associate Professor with the College of Robotics, Guangdong Polytechnic of Science and Technology, Zhuhai, China. Meanwhile, he has been invited as a Guest Professor at the Hubei Provincial Laboratory of Intelligent Robot and a Distinguished Research Fellow at Fujian Agriculture and Forestry University. His research interests include computer vision, intelligent robots, and agriculture automation.

Yuntao Wu received the Ph.D. degree in information and communication engineering from the National Key Laboratory for Radar Signal Processing, Xidian University, Xi'an, China, in 2003.

From April 2004 to September 2006, he was a Postdoctoral Researcher with the Institute of Acoustics, Chinese Academy of Sciences, Beijing, China. From October 2006 to February 2008, he was a Senior Research Fellow with the City University of Hong Kong, Hong Kong. He was a Visiting Researcher with the Faculty of Engineering, Bar-Ilan University, Ramat Gan, Israel, from April 2013 to March 2014. He is currently a Full-Time Professor with the Wuhan Institute of Technology, Wuhan, China, where he is also a Chutian Scholar Project in Hubei Province Distinguish Professor. He has authored or coauthored more than 80 journal and conference papers. His research interests include signal detection, parameter estimation in array signal processing, and source localization for wireless sensor networks, biomedicine signal analysis, etc.

Dr. Wu is a senior member of the Chinese Institute of Electronic Engineers. He is an Associate Editor for *Multidimensional Systems and Signal Processing*.

Yanduo Zhang received the Ph.D. degree in aircraft design from the School of Astronautics, Harbin Institute of Technology, Harbin, China, in 1999.

From 2012 to 2013, he was a Visiting Professor with the Department of Statistics, University of California at Los Angeles, Los Angeles, CA, USA. He is currently a Professor with the Hubei Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan, China. His main research interests include artificial intelligence, intelligent robotics, and computer vision.