

Non-Rigid Image Registration With Dynamic Gaussian Component Density and Space Curvature Preservation

Zhuoqian Yang^{ID}, Yang Yang^{ID}, Member, IEEE, Kun Yang, and Zi-Quan Wei^{ID}

Abstract—Image registration plays an important role in military and civilian applications, such as natural disaster damage assessment, environmental monitoring, ground change detection, and military damage assessment. This paper presents a new feature-based non-rigid image registration method. The main contributions of this paper are: 1) a dynamic Gaussian component density is designed to better exploit available potential image information and provide sufficient inlier pairs for image transformation and 2) a spatial structure preservation, which consists of an image transformation space curvature preservation and a local spatial structure constrain, is proposed to constrain the image transforming cost as well as the local structure of feature points during feature point set registration. The performances of the proposed method in multi-spectral natural images, low-altitude aerial images, and medical images against four types of nine state-of-the-art methods are tested where our method shows the best performances in most scenarios.

Index Terms—Image registration, feature matching, non-rigid, dynamic Gaussian component density, space curvature preservation.

I. INTRODUCTION

IMAGE registration is to find an optimal alignment between two or more images. As a necessary step in comparing or integrating data obtained from the same or different measurements, its applications include target detection [1],

Manuscript received May 23, 2018; revised November 3, 2018 and December 12, 2018; accepted December 13, 2018. Date of publication December 17, 2018; date of current version March 6, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 41661080, in part by the Yunnan Province Ten-Thousand Talents Program, and in part by the National Undergraduate Training Program for Innovation and Entrepreneurship under Grant 201710681017. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Lisiachos P. Kondi. (*Zhuoqian Yang and Yang Yang contributed equally to this work.*) (*Corresponding authors:* Yang Yang; Kun Yang.)

Z. Yang is with the Engineering Research Center of GIS Technology in Western China, Ministry of Education, Yunnan Normal University, Kunming 650500, China, and also with the College of Software, Beihang University, Beijing 100083, China.

Y. Yang is with the School of Information Science and Technology, Yunnan Normal University, Kunming 650500, China, with the Engineering Research Center of GIS Technology in Western China, Ministry of Education, Yunnan Normal University, Kunming 650500, China, and also with the Laboratory of Pattern Recognition and Artificial Intelligence, Yunnan Normal University, Kunming 650500, China (e-mail: yyang_ynu@163.com).

K. Yang is with the School of Information Science and Technology, Yunnan Normal University, Kunming 650500, China, and also with the Engineering Research Center of GIS Technology in Western China, Ministry of Education, Yunnan Normal University, Kunming 650500, China (e-mail: kmcdywu@163.com).

Z.-Q. Wei is with the Laboratory of Pattern Recognition and Artificial Intelligence, Yunnan Normal University, Kunming 650500, China.

Digital Object Identifier 10.1109/TIP.2018.2887204

large-scale 3D reconstruction [2] and multi-modal image processing [3]. It is also a critical prerequisite in environment monitoring [4], [5], image fusion and map updating [6]–[8].

Traditionally, there are several representative classifications for image registration methods, such as spatial vs. frequency domain, global vs. local as well as rigid vs. non-rigid. Since this work focuses on feature matching/registration problems, we introduce the related work along the classification perspective: area-based methods vs. feature-based methods presented in [9].

Area-based methods directly manipulates image intensity values and do not rely on salient features. Its subclasses, in a broad sense, include Correlation-like methods, Fourier methods, and mutual information (MI) methods. Correlation-like methods such as the cross-correlation and its derivations [10] are representative of area-based methods. These methods are less sophisticated due to the flatness of similarity measure in textureless regions and high computational complexity [11]. Fourier methods, exploiting the Fourier frequency-domain representation of images [12], have some advantages in computational efficiency and are also more robust to frequency-dependent noise comparing with correlation-like methods. However, they have some limitations in the case of image pairs having significantly different spectral contents [13]. MI methods provide an attractive metric for maximizing the dependence between two images, and are particularly suitable for multimodal registration [14]. Despite their outstanding performance, MI-based methods cannot provide a global maximum of the entire search space for transformation which will inevitably reduce their robustness [13].

Feature-based methods are also popular in the field of image registration, and typically consist of the following three steps:

- (i) Feature extraction: two sets of feature points are extracted from the sensed image I^s and the reference image I^r , respectively, by a feature detector like scale-invariant feature transform (SIFT) [15] or Harris and Stephens [16].
- (ii) Feature point set matching or registration: applying matching or registering strategy finds correspondences or both correspondences and transformation between two feature point sets.
- (iii) Image transformation: I^s is aligned to I^r by an image transformation T computed by the recovered correspondences in the previous step.

Moreover, these are two kinds of strategies regarding the above (ii) step in current feature-based methods:

- Matching strategy: it only focuses on determining correct correspondences between two point sets using a salient

feature similarity measurement, such as the matching algorithm in SIFT [17] and the speeded up robust feature (SURF) [18].

- Registration strategy: it recovers both correct correspondences and the underlying spatial transformation between two feature point sets through an iterative approach, such as [19]–[24].

These two strategies are methodologically designed for different considerations. For example, matching strategy focuses on developing a robust feature descriptor and its similarity measurement, while registration strategy is committed to estimate the underlying transformation which can reasonably map a point set onto other.

There are many famous methods using the matching strategy such as the SIFT algorithm proposed by Lowe [15], [17] which describes scale-invariant local features of images for object recognition. It has been widely used for different kinds of image registrations [25]–[29]. Various SIFT-like local descriptors such as gradient location and orientation histogram (GLOH) [30] and DAISY [31] draw upon the effectiveness of the SIFT. Bay *et al.* [18] proposed the speeded up robust feature (SURF), a novel scale-and rotation-invariant feature point detector, which likewise has proved its validity for image registration [32]–[34]. Harris corner detector [16] is one of the best corner detection operators [35], and has already been applied in image registration [36]–[38]. Rosten and Tom [39] proposed the feature from accelerated segment test (FAST) corner detector that can perform a high-speed corner detection without necessarily sacrificing quality, is commonly used in scenarios of remote sensing where real-time processing is the priority [40], [41]. Binary robust independent elementary feature (BRIEF) proposed by Calonder *et al.* [42] is another outstanding feature descriptor with low complexity. Oriented FAST and rotated BRIEF (ORB) proposed by Rublee *et al.* [43] not only inherited the speed and low cost of BRIEF descriptor, but also improved its rotational invariance. Aguilera *et al.* [44] proposed the edge oriented histogram (EOH) descriptor, for multi-modal and multi-spectral image registration cases which both allow matching interest points in images via costing method. Ma *et al.* [45] proposed an efficient feature matching approach based on maintaining the local neighborhood structures of potential true matches. It provides a closed-form solution which can accomplish the mismatch removal from thousands of putative correspondences in only a few milliseconds.

Under the registration strategy, the famous iterative two-steps approach was first introduced in the iterate closest point (ICP) algorithm [46]. The first step, namely correspondence estimation, is to establish correspondences between two point sets by a feature similarity measurement, and the second step, namely transformation updating, is to recover the underlying transformation with certain spatial structure preservations [23]. Based on this two-steps framework, Myronenko and Song [20] introduced a probabilistic method named as the coherent point drift (CPD) algorithm. It regards the source and target point sets as the centroid of components and data, respectively, and the registration procedure is interpreted to a maximum likelihood estimation problem. The expectation

maximization (EM) algorithm framework is employed for the parametric estimation, and the transformation updating is constrained with the motion coherence theory (MCT). Jian and Vemuri (GMMREG) [47] then extended the idea of registration from fitting Gaussian mixtures into data to aligning two Gaussian mixture models (GMM). Ma *et al.* [48] employed the robust L_2 minimizing estimation (L_2E) [49] criterion to refine the point-wise transformation. Yang *et al.* [21] (GLMDTPS) proposed a mixture-feature based correspondence estimation method using global and local mixture distance (GLMD), where the global distance is point-to-point Euclidean distance and the local distance is computed by summing the squared Euclidean distance between the i^{th} neighboring points according to the index. Ma *et al.* [22] (PRGLS) obtained a binary corresponding matrix by matching the shape context (SC) descriptors, and it is used to improve CPD through directly assigning the membership probabilities of GMMs to make it close to one if matched, otherwise to zero and updating the transformation under preserving global and local structures. Recently, Ma *et al.* [12], [50] who expanded the spatial structure preservation by combining it with the EM algorithm framework, proposed the vector field consensus (VFC) and the locally linear transforming (LLT), successively. VFC provides a uniform framework for robust feature matching by interpolating a non-parametric function with a slow-and-smooth prior condition, which is very general and can accept various geometric models. LLT computes the correspondence matrix by a binary putative feature correspondence adopting SIFT matching method, then a locally linear transformation is designed. More recently, Yang *et al.* [27] and Wei *et al.* [28] provided two combinations of SIFT feature description and spatial geometric feature which were used in the L_2E criterion and the EM algorithm framework, respectively, to solve remote sensing image registration problems. Zhang *et al.* [51] proposed a nonrigid feature-based method for low-altitude small unmanned aerial vehicle (SUAV) image registration, which is robust to a very large proportion of outliers, even with up to 80% outliers.

Although the two strategies have successfully solved some problems in image registration and have been used in different applications, their initial considerations (or goals) limit image registration accuracy before registration begins. For example:

- In the matching strategy, reliable inlier pairs (i.e., points in a feature point set have their corresponding points in the other) are obtained by thresholding a feature distance, such as in SIFT and SURF. In order to prevent unstable/incorrect inlier matches, the thresholds in such methods are normally fixed at a relatively high similarity measurement. This kind of setting, however, exacts a loss of available potential image information, i.e., losing a part of potential inlier pairs. In other words, losing potential inlier pairs will limit the image registration accuracy since the recovered inlier pairs are used as control points in building image transformation and the more control points, the finer the transformation.
- In the registration strategy, methods focus on recovering the underlying transformation, while rejecting outliers (i.e., points have no counterparts in either set) and reducing their negative impacts on the transformation

using certain constraints. However, these constraints were designed to only preserve feature points (e.g., the famous motion coherent theory in CPD [20]), rather than image itself. In addition, the above problem in the matching strategy also exists in current registration methods due to their “static” feature point extractions (i.e., extracting feature points with a fixed threshold).

In order to solve the above issues, we present a new feature-based non-rigid image registration method in this work. The main contributions of the proposed methods are: (i) a dynamic Gaussian component density is designed to gradually maximize the available potential image information and provide sufficient inlier pairs for image transformation; (ii) a spatial structure preservation, which consists of an image transformation space curvature preservation and a local spatial structure constrain, is proposed to constrain not only the local structure of feature points but also the image transforming cost during feature point set registration.

II. METHODOLOGY

This section describes the proposed feature-based non-rigid image registration method. We first introduce a traditional three-steps process of registration strategy for feature-based image registration in II-A and then present two ideas for improving the non-rigid registration step in II-B and II-C, respectively. We subsequently give the solution of the optimal parameters in II-D, and finally discuss the parameter setting and complexity analysis in II-E.

A. Main Process

The purpose of feature-based image registration is to estimate a transformation for laying a sensed image $I_{X^s \times Y^s}^s$ into the coordinate system of a reference image $I_{X^r \times Y^r}^r$, where I^s and I^r have the resolutions with $X^s \times Y^s$ and $X^r \times Y^r$, respectively, and current registration strategy based methods generally have the three-steps process as follows:

1) *Feature Points Extraction*: In the first step, a target feature point set $\mathbf{T}_{N \times 2} = \{t_1, t_2, \dots, t_N\}^T$ and a source feature point set $\mathbf{S}_{M \times 2} = \{s_1, s_2, \dots, s_M\}^T$ are extracted from I^r and I^s by a same feature extraction method, respectively. $\lambda(t_i)$ and $\lambda(s_i)$ denote the feature descriptor of t_i and s_i , respectively. Coordinates of the two point sets are then rescaled within $[0, 1]$.

2) *Non-Rigid Registration for Feature Points*: In the second step, a point set registration is employed to recover an unknown non-rigid transformation $\mathbf{S}^* = \mathcal{F}(\mathbf{S}, \Gamma)$ for mapping the source points onto the target points, where Γ is a set of unknown parameters. The unknown transformation can be solved by estimating Γ based on the expectation maximization (EM) algorithm [52], [53], which has two iterative steps: (i) correspondence estimation step (E-step) and (ii) transformation estimation step (M-step). The EM algorithm for the point set registration problem can be described as follows.

The source points are considered as Gaussian mixture model (GMM) centroid and the target points are considered as data points onto which the centroid should be fitted. Let t_i be the

i^{th} data and s_j be the centroid of the j^{th} component. The GMM probability density function is then obtained as:

$$p(t_i) = w \frac{1}{N} + \sum_{j=1}^M (1-w) \mathcal{C}_{ji} g(t_i | s_j), \quad (1)$$

where the component density \mathcal{C}_{ji} is a nonnegative quantity having $\sum_{j=1}^M \mathcal{C}_{ji} = 1$, and the Gaussian component $g(t_i | s_j) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{\|t_i - s_j\|^2}{2\sigma^2})$ having the covariance parameter σ^2 . An additional uniform distribution $\frac{1}{N}$ is normally used to deal with outliers by a weighting parameter $0 \leq w \leq 1$.

- For E-step: the posterior probability is computed based on the Bayes rule:

$$p_{ji} = \frac{\mathcal{C}_{ji} g(t_i | s_j)}{\frac{w}{1-w} \frac{1}{N} + \sum_{m=1}^M \mathcal{C}_{mi} g(t_i | s_m)}, \quad (2)$$

The poster probability matrix (i.e., correspondence matrix) is obtained by $\mathbf{P} = \{p_{ji}\}_{j=1, i=1}^{M, N}$, and the putative corresponding target set can be obtained by $\hat{\mathbf{T}} = \mathbf{PT}$. Under the iterative registration strategy, the non-rigid transformation $\mathcal{F}(\mathbf{S}, \Gamma)$ should be reproduced in a special functional space. Thus, we define this transformation as

$$\mathcal{F}(\mathbf{S}, \Gamma) = \mathbf{S} + v, \quad (3)$$

where $v = \mathbf{K}\Gamma$ is the displacement function, and required to lie within a reproducing kernel Hilbert space (RKHS) \mathcal{H} with an unknown coefficient matrix Γ and a $M \times M$ Gaussian kernel matrix \mathbf{K} consisted of element $k_{xy} = \exp(-\frac{1}{2\varepsilon^2} \|s_x - s_y\|^2)$, where ε was the covariance of $\{\|s_x - s_y\|\}_{x=1, y=1}^{M, M}$.

- For M-step: the locations of GMM centroid are reparameterized using the transformation and updated the new value of Γ by maximizing the likelihood, or equivalently minimizing:

$$Q = - \sum_{i=1}^N \sum_{j=1}^M p_{ji} \log [\mathcal{C}_{ji} g(t_i | s_j)] + \mathcal{R}(v), \quad (4)$$

where the first term is the expectation of the negative log-likelihood function of GMM and the last term is the regularization of the displacement v . Eq.4 can be rewritten by ignoring the implementations of Gaussian component density and the regularization as:

$$Q = \frac{1}{2\sigma^2} \sum_{i=1}^N \sum_{j=1}^M p_{ji} \|t_i - \mathcal{F}(s_j, \Gamma)\|^2 + \frac{N_p}{2} \log 2\pi\sigma^2 - \sum_{i=1}^N \sum_{j=1}^M p_{ji} \log \mathcal{C}_{ji} + \mathcal{R}(v), \quad (5)$$

where $N_p = \sum_{i=1}^N \sum_{j=1}^M p_{ji} \leq N$ ($N_p = N$ only if $w = 0$). The unknown parameter set can be obtained by $\Gamma = \arg \min \Gamma Q$. The transformation \mathcal{F} and new locations of \mathbf{S}^* are then updated by the new parameters by $\mathbf{S}^* = \mathcal{F}(\mathbf{S}, \Gamma)$.

The registration iteration can be stopped when any of these conditions is satisfied: (i) the iteration number reaches

the limit, (ii) the object function (e.g., Eq. 5) converges, (iii) σ^2 shrinks below the boundary v . The final locations of \mathbf{S}^* denotes the registration result.

3) Image Transformation: In the third step, the image registration result is obtained by a recovered image transformation model (e.g., the thin plate spline (TPS) [54]) which is computed by the corresponding set $\mathcal{C} = \{\mathbf{S}, \mathbf{S}^*\}$.

In this work, the image transformation space is considered as a point lattice which is defined as $\mathcal{L}(\cdot) = \{\Omega_n(\cdot)\}_{n=1}^{r \times c}$ with $r \times c$ lattice points, where dots form a control point set for updating the transformation. r and c are the number of rows and columns, respectively, and the index is reshaped by row-major order. $\mathcal{L}(\mathbf{S})_{\mathcal{Z} \times 2}$, a regular point lattice, is obtained from the reference image I^r by a pixel-by-pixel indexing process, where $\mathcal{Z} = X^r \times Y^r$. We make $\mathcal{L}(\mathbf{S})$ be the source point lattice using the TPS transformation model under the set \mathcal{C} . The transformed point lattice is obtained by computing $\tilde{\Theta}_{\mathcal{Z} \times 2}^t = \mathcal{L}(\mathbf{S}^*)$ which is defined as follows:

$$\mathcal{L}(\mathbf{S}^*) = \mathcal{W} \mathcal{Y}^{-1} \begin{pmatrix} \mathbf{S}^* \\ \mathbf{0} \end{pmatrix}, \quad \mathcal{W} = (\bar{\Phi} \quad \Xi) \quad \text{and} \quad \mathcal{Y} = \begin{pmatrix} \Phi & \Upsilon \\ \Upsilon^T & \mathbf{0} \end{pmatrix}, \quad (6)$$

where the $X^r \cdot Y^r \times M$ matrix $\bar{\Phi}$ is obtained by $\bar{\phi}_{ij} = \|\Omega_i(\mathbf{S}) - s_j\|^2 \log \|\Omega_i(\mathbf{S}) - s_j\|$ and the $M \times M$ matrix Φ is obtained by $\phi_{ij} = \|s_i - s_j\|^2 \log \|s_i - s_j\|$. Ξ is a $X^r \cdot Y^r \times 3$ matrix with the n^{th} denoting $(1, \Omega_n(\mathbf{S}))$, Υ is a $M \times 3$ matrix with the n^{th} denoting $(1, s_n)$ and $\mathbf{0}$ is a matrix of zeros.

$\hat{\Theta}^t$ within a window which has the same size as the reference image is obtained by

$$\hat{\Theta} = \hat{\Theta}^t \cap \Theta^t. \quad (7)$$

Finally, the transformed image I^* is obtained from the sensed image I^r by re-sampling intensity based on $\hat{\Theta}$, setting the rest of the pixels black. Note that the bi-cubic interpolation is used to improve the smoothness of I^* , to be more precise, the intensity of each pixel in I^* is determined by the summation of the weighted neighbor pixel intensity within a 4×4 window.

B. Dynamic Gaussian Component Density

In order to be unfettered by the fixed threshold such as in SIFT and SURF, while gradually maximizing available potential image information and providing sufficient inlier pairs for the subsequent image transformation step, a dynamic Gaussian component density is designed to dynamically assign different costs to putative outliers and inliers in the aforementioned E-step in II-A2, respectively. This design provides a new solution compared with the traditional “static” probability density used in previous work [19], [20], [22]–[24], [47], and [55].

The Gaussian components density matrix $\mathbf{C}_{M \times N} = \{\mathcal{C}_{ji}\}_{j=1, i=1}^{M, N}$ (defined in Eq. 1) is priori and defined as $M \times N$ equivalent elements, i.e., $\frac{1}{M}$, in several popular methods such as Myronenko and Song [20] and Horaud *et al.* [55]. Nevertheless, this widely used Gaussian components density was based on an assumption that all feature points are consubstantial, e.g. in a circumstance where the point set registration algorithm is agnostic about the features associated with the points.

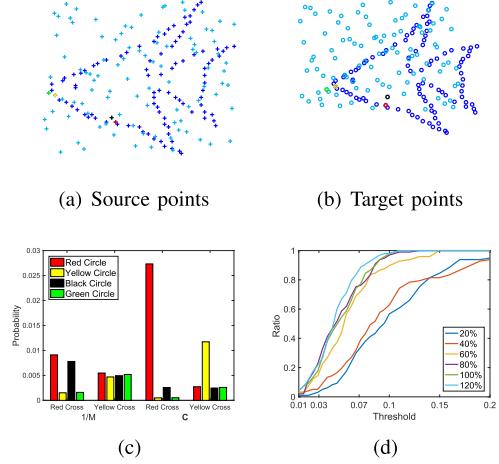


Fig. 1. Illustration of the influence of severe outliers. (a) and (b) are two fish contour point sets with 100% outliers. All points are normalized in $[0, 1]$. The blue, red and yellow points are inliers, the cyan, green and black points are random outliers. The cyan and the green ones are the nearest neighbor to the red and the yellow ones, respectively. Moreover, the red circle is correspondent with the red cross and the yellow circle is correspondent with the yellow cross. (c) demonstrates correspondence probabilities which is obtained by $g(o, o, o, o|+)$ and $g(o, o, o, o|+)$ with their density $\frac{1}{M}$ or \mathcal{C}_{ji} . (d) shows the ratio of distracted inliers under different outlier ratios.

Fig. 1 gives an example of assessing the reliability of a few pairs of potential correspondences. Fig. 1 (a) and (b) shows the positions of the examined points and (c) shows the correspondence probabilities computed using $\frac{1}{M}$ and \mathcal{C}_{ji} , respectively. It can be observed that in the results produced by traditional methods, the factually corresponding pairs have correspondence probabilities that is approximate to or less than that of the uncorrelated pairs. The proposed dynamic Gaussian component density, on the other hand, is able to make a much more unequivocal distinction. Additionally, the ratio of distracted inliers increases with the increase of outlier ratio as shown in (d), where the distracted inliers are considered as outliers existing within the threshold.

The priori probability of each component is determined using intensity information extracted from the images. There are various kinds of optional intensity information, such as the famous SIFT [15] and SURF [18]. For explanatory convenience, a specified feature descriptor is written as X feature, and we then compute the X feature distance between source and target by:

$$\Delta(s_j, t_i) = \|\lambda(s_j) - \lambda(t_i)\|, \quad (8)$$

which is the size of $M \times N$. A soft strategy is designed to compute the dynamic Gaussian components density matrix \mathbf{C} in the aforementioned iterative E-step (see Eq. 2), and consists of the following three steps:

- (i) A set of inlier pairs between the source \mathbf{S} and the target \mathbf{T} is presumed by a pre-matching algorithm $\mathcal{M}(\mathbf{S}, \mathbf{T})$, such as provided in [17]. $\mathcal{M}(\mathbf{S}, \mathbf{T})$ returns an index set \mathcal{I} which indicates the indices of putative corresponding inlier pairs between the two point sets. $\mathcal{M}(\mathbf{S}, \mathbf{T})$ is defined as: point s_j matches to t_i only if the distance $\Delta(s_j, t_i)$ multiplied by a threshold τ is not greater than the distance of $\lambda(s_j)$ from any other target descriptor.

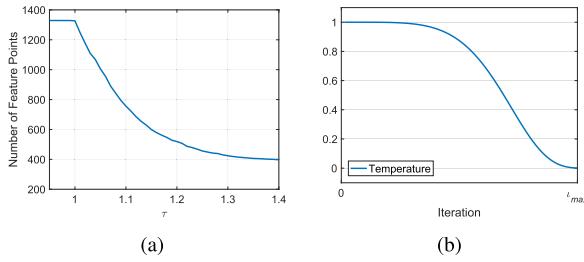


Fig. 2. (a): Number of the extracted feature points with different threshold τ . (b): Changes of the temperature for trade-off parameters.

The threshold τ is updated in the soft strategy process by $\tau \leftarrow \tau - \Delta$ with the step parameter Δ , if $\tau > 1$. The number of the pre-matched feature points with different threshold is exemplified in Fig. 2 (a).

- (ii) Compute two matrices $\mathbf{L}_{M \times N}$ and $\mathbf{D}_{M \times N}$. \mathbf{L} is a shape context chi-square matrix used as a spatial structure feature while combining with Λ :

$$\mathbf{L}(s_j, t_i) = \frac{1}{2} \sum_{k=1}^K \frac{[h_j(k) - h_i(k)]^2}{h_j(k) + h_i(k)}, \quad (9)$$

where $h_j(k)$ and $h_i(k)$ denote the K -bin normalized histogram at s_j and t_i , respectively [56]. \mathbf{D} is a fuzzy outlier distribution cost matrix to weight outliers:

$$\mathbf{D}(s_j, \cdot) = \frac{1}{\sqrt{2\pi\beta^2}} \exp\left(-\frac{\|s_j - s_{n(j)}\|^2}{2\beta^2}\right), \quad (10)$$

where $n(j) = \arg \min_{i \in \mathcal{I}(1,\cdot)} (\|s_i - s_j\|)$ and β is the covariance parameter.

- (iii) The dynamic density matrix \mathbf{C} is then computed by:

$$\mathbf{C}(s_j, t_i) = \begin{cases} \mathbf{L}(s_j, t_i) \cdot \Lambda(s_j, t_i), & \text{case A} \\ \mathbf{L}(s_j, t_i) \cdot (1 - \mathbf{D}(s_j, t_i)\Lambda(s_j, t_i)), & \text{case B,} \end{cases} \quad (11)$$

where case **A** is $j \in \mathcal{I}_{(1,\cdot)}$ or $i \in \mathcal{I}_{(2,\cdot)}$, case **B** is $j \notin \mathcal{I}_{(1,\cdot)}$ and $i \notin \mathcal{I}_{(2,\cdot)}$. Under the pre-matching index set \mathcal{I} , case **A** and **B** are used to assign different costs to putative inliers and outliers, respectively. In case **B**, \mathbf{D} is designed to punish putative outliers. The closer the putative outliers are to putative inliers, the heavier the punishment (i.e., assigning a higher cost in Eq. 11).

In order to improve the efficiency of registration while still maintaining a stable registration (i.e., reduce computation by reducing a certain number of iterations appropriately), the density matrix is not recomputed in every iterations but in every d iterations by solving a linear assignment on the density matrix \mathbf{C} using the Jonker-Volgenant algorithm [57]. After that, the Gaussian components density matrix is processed with the parameter δ :

$$C_{ji} = \begin{cases} \frac{1-\delta}{M}, & \text{if } \mathbf{C}(s_j, t_i) = 0 \\ \delta + \frac{1-\delta}{M}, & \text{if } \mathbf{C}(s_j, t_i) = 1. \end{cases} \quad (12)$$

The reasons of alternating a fuzzy \mathbf{C} and a binary \mathbf{C} with a contiguous, approximate one are that: (i) a fuzzy \mathbf{C} provides a fuzzy correspondence matrix which makes the registration more steadily but requiring more iterations; (ii) a binary \mathbf{C} provides a binary correspondence matrix which makes the registration straightforwardly and only fewer iterations are required, but increases the risk of letting incorrect assignments guide the registration.

Fig. 3 visualizes the proposed image registration with the dynamic Gaussian components density matrix. The proposed soft strategy first combines the intensity feature and the spatial structure feature to enhance the feature descriptions of points, and then applies a decreasing threshold to gradually and steadily increase the number of putative inlier pairs while assigning different costs to putative outliers according to a fuzzy outlier distribution. Based on these, the inlier searching processing (i.e., the proposed dynamic Gaussian component density) is to recover the most reliable inlier pairs at the begin of registration, and then gradually changed to search the maximum number of inlier pairs at the end of registration. This process finally maximizes available potential image information and provides more sufficient and reliable inlier pairs for recovering a better image transformation in the subsequent step.

C. Spatial Structure Preservation

In order to obtain a robust transformation $\mathcal{F}(\mathbf{S}, \Gamma)$ for the source point set \mathbf{S} and a more accurate image transformation $\mathcal{L}(\cdot)$ for the sensed image I^s , a spatial structure preservation is proposed in the aforementioned M-step in II-A2. Compared with the constraints used in current registration strategy based methods, the proposed approach preserves not only the structure of the source point set, but also constrains the image transforming cost during the iterative point set registration.

For these considerations, we estimate the coefficient matrix Γ with the proposed spatial structure preservation, which is represented by a regularization term $\mathcal{R}(v)$ in Eq. 5. $\mathcal{R}(v) = \frac{\alpha}{2}\mathcal{T} + \frac{\eta}{2}\mathcal{L} + \frac{\rho}{2}\mathfrak{X}$, includes the Tikhonov regularization [58], a local structure preservation term, and a space curvature preservation term. In addition, we define a deterministic annealing scheme for three trade-off parameters $\alpha \leftarrow \kappa\alpha$, $\eta \leftarrow \kappa\eta$ and $\rho \leftarrow \kappa\rho$. The annealing parameter is defined as $\kappa = \frac{(t_{max}^4 - t^4 + 1)^{\frac{1}{4}}}{t_{max}}$, where t is the iteration number and t_{max} is the maximum iteration number. The change of κ is shown in Fig. 2 (b). This annealing process implies that constraint $\mathcal{R}(v)$ is released at the end of iterations for achieving the maximization of the point-wise overlap. The simultaneous regularizing for $\mathcal{F}(\mathbf{S}, \Gamma)$ and $\mathcal{L}(\cdot)$ aims to preserve the robustness of the final image transformation.

The existing Tikhonov regularization is a global structure constraint term based on the motion coherent theory (MCT) [59], which is defined as:

$$\mathcal{T} = \|v\|^2. \quad (13)$$

As shown in Figure 4 (a), the regularized function (denoted by black line) is more reasonable than its non-regularized

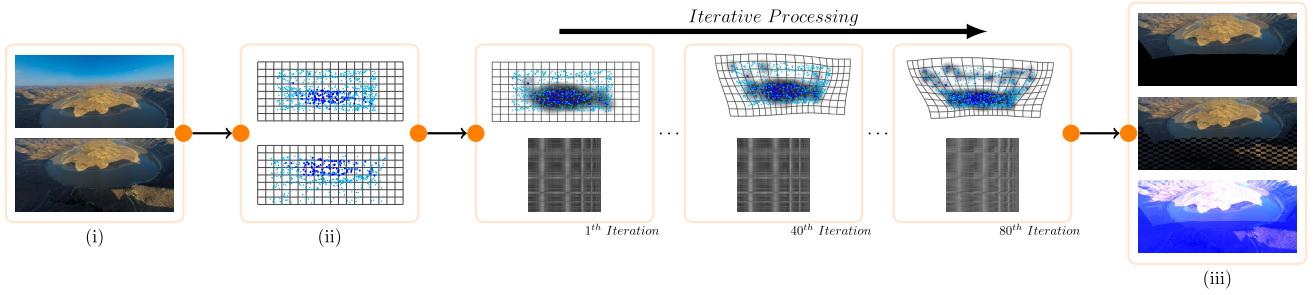


Fig. 3. Illustration of the proposed image registration with the dynamic Gaussian components density matrix. (i): The sensed image is shown in the first row, and the second row is the reference image. (ii): The source feature points with their displacement space are shown in the first row, and the second row is the target. Blue and cyan points denote inliers and outliers, respectively. The grid denotes the displacement space of feature points. (iii): The warped image is shown in the first row, and the second row is a 30×30 checkboard, and the third row shows the overlapping area. In the iterative processing, the gray mist-like region around each inlier visualizes the fuzzy outlier distribution costs, where the greater cost is used to punish the outliers close to putative inliers; the mosaic rectangles in the second row visualize the changes (i.e., “dynamic”) of the density matrix C defined in Eq. 11 during registration. In the iterative processing, greater values are represented by deeper colors.

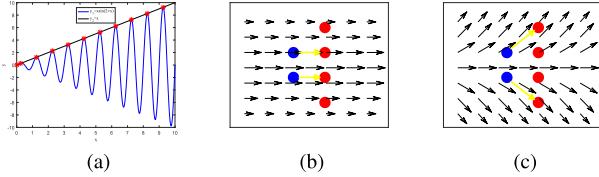


Fig. 4. Two examples for demonstrating the importance of regularization. (a) two different results for a function estimation problem. The black line and blue curve denote the estimated functions with or without being regularized, respectively. In addition, the red asterisks denote 11 data points; (b) point set transformation and its velocity field in regularized scenario; (c) point set transformation and its velocity field in a non-regularized scenario. In (b) and (c), the blue and red points denote the source and target point, respectively.

counterpart (denoted by blue curve). The transformation of an iterative registration process is a procedure that slowly displaces the source point set so that the correspondence estimation is easier and more reliable. In other words, regularizing the transformation is necessary to accomplish the iterative registration. Figure 4 (b) and (c) indicate that the ill-posed problem will exist if the transformation is not regularized. Note that as the number of points increases, the increasing arbitrariness of the transformation will lead to more severe ill-posed problems.

1) *Local Structure Preservation*: The local structure of a source set is preserved by maintaining the local structure similarity between transformed source inlier ($\in S^*$) and putative corresponding target ($\in \hat{T}$). Here we only preserve the local structure similarities of putative inliers to agree with the thought of inlier being the hard skeleton in the whole transformation. The local structure preservation is defined as follows:

$$\mathcal{L} = \sum_{j \in \mathcal{I}_{(1,:)}} \|\boldsymbol{\pi}(\hat{t}_j) - \boldsymbol{\pi}(\mathcal{F}(s_j, \Gamma))\|^2, \quad (14)$$

with a local structure descriptor $\boldsymbol{\pi}(\cdot)$. There is a point set $\mathbf{Z} = \{z_1, z_2, \dots, z_N\}^T$. \mathcal{K} nearest neighbors of z_n are called as $\{z_{nk}\}_{k=1}^{\mathcal{K}}$. We have $\boldsymbol{\pi}(z_n) = \sum_{k=1}^{\mathcal{K}} h_{nk} \bar{z}_n z_{nk}$, where $h_{nk} = \exp(-\frac{\|z_k - z_n\|^2}{\xi_n^2})$ has the variance ξ_n of $\{\|z_k - z_n\|\}_{k=1}^{\mathcal{K}}$. Minimizing the local structure constraint term \mathcal{L} is equivalent

to guiding the transformation to produce a result having the most similar local structure of putative inliers. Finally, we simplified \mathcal{L} in the matrix form:

$$\mathcal{L} = \text{Tr}(\mathbf{U}\mathbf{U}^T), \quad (15)$$

where $\text{Tr}(\cdot)$ denotes the trace operation, $\mathbf{U} = (\mathbf{H}^{\hat{T}} - \mathbf{K}\mathbf{I})\mathbf{P}\mathbf{T} - (\mathbf{H}^S - \mathbf{K}\mathbf{I})(\mathbf{S} + \mathbf{K}\Gamma)$, and \mathbf{H}^Z has non-zero entries h_{ij} in $N \times N$ dimension, if $i \in \mathcal{I}$ and $z_j \in \{z_{ik}\}_{k=1}^{\mathcal{K}}$.

2) *Space Curvature Preservation*: As described in Section II-A3, the image transformation space is defined as $\mathcal{L}(\cdot)$ with a set of control points. Hence, the transformation space in each iteration can be described as $\mathcal{L}(\mathcal{F}(\mathbf{S}, \Gamma))$ using Eq. 6. We constrain the space curvature cost of the image transformation during point set registration, and the proposed preservation helps to improve the accuracy for the subsequent image transformation.

The curvature preservation process is to preserve the smoothness of the transformation space via minimizing the curvature cost in the cross and lengthwise directions. We define the cost as:

$$\mathfrak{X} = \sum_{i=1}^{rc} \chi_i^{r^2} + \sum_{j=1}^{rc} \chi_j^{c^2}, \quad (16)$$

where r and c are the numbers of rows and columns, respectively. $\chi_{rc \times 1}^r$ is the curvature set in the row direction and $\chi_{rc \times 1}^c$ is the curvature set in the column direction.

The i^{th} row and j^{th} column of $\mathcal{L}(\mathcal{F}(\mathbf{S}, \Gamma))$ are denoted by $\boldsymbol{\gamma}^i$ with c lattice points and $\boldsymbol{\epsilon}^j$ with r lattice points, respectively. We compute the curvature of the discrete point set using a finite difference method (FDM). First, two shift operators are defined for a given vector set $\mathbf{Z} = \{z_1, z_2, \dots, z_N\}^T$, left shift $\mathbf{Z}^L = \{z_2, z_3, \dots, z_N, z_N\}^T$ and right shift $\mathbf{Z}^R = \{z_1, z_1, z_2, \dots, z_{N-1}\}^T$. We denote the curvature of points in one row/column by the gradient of curve formed by row/column points. The first-order difference of the n^{th} point in the i^{th} row is obtained by $\boldsymbol{\gamma}_{in}' = [\boldsymbol{\gamma}_{i(n,2)}^R - \boldsymbol{\gamma}_{i(n,2)}^L]/[\boldsymbol{\gamma}_{i(n,1)}^R - \boldsymbol{\gamma}_{i(n,1)}^L]$ and $\boldsymbol{\epsilon}_{m'}^j$. The first-order difference of the m^{th} point in the j^{th} column is obtained in the same manner. $(n, 1)$ and $(n, 2)$ denote the 1st and the 2nd dimension coordinate of the n^{th} point, respectively. Then reshape

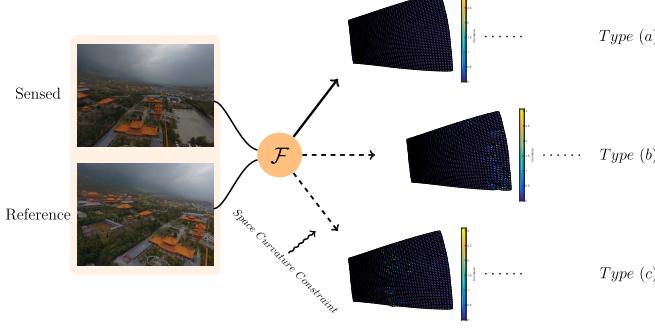


Fig. 5. Illustration of the space curvature constraint. The grid denotes the displacement space and is plotted by the transformed point lattice. Displacement spaces of type (a), (b) and (c) are obtained by three potential transformations. (a) is accepted, (b) and (c) are rejected.

$\{\mathbf{y}_{in}^{r,c}\}_{i=1,n=1}^{rc}$ and $\{\boldsymbol{\epsilon}_m^{j,r}\}_{j=1,m=1}^{c,r}$ in the size of $rc \times 1$ using the same manner as the point lattice one, denoting χ^r and χ^c .

As shown in Fig. 5, minimizing the space curvature constraint term is equivalent to guiding the transformation to produce a result having the most regular displacement space. Finally, we simplify \mathfrak{X} in the matrix form:

$$\mathfrak{X} = \chi^{rT} \chi^r + \chi^{cT} \chi^c. \quad (17)$$

D. Solution of the Optimal Parameters

The coefficient matrix Γ is determined by minimizing Q with the regularization $\mathcal{R}(v)$. The objective function (Eq. 5) can be rewritten in a matrix form as:

$$\begin{aligned} Q = & \frac{1}{2\sigma^2} [Tr(\mathbf{T}^T d(\mathbf{P}^T \mathbf{1}) \mathbf{T}) - 2Tr(\mathbf{S}^T \mathbf{P} \mathbf{T}) + Tr(\mathbf{S}^T d(\mathbf{P} \mathbf{1}) \mathbf{S}) \\ & - 2Tr(\Gamma^T \mathbf{K} \mathbf{P} \mathbf{T}) + 2Tr(\Gamma^T \mathbf{K} d(\mathbf{P} \mathbf{1}) \mathbf{S}) \\ & + Tr(\Gamma^T \mathbf{K} d(\mathbf{P} \mathbf{1}) \mathbf{K} \Gamma)] \\ & + \frac{N_p}{2} log 2\pi \sigma^2 - Tr(\mathbf{P}^T log \mathbf{C}) + \frac{\alpha}{2} Tr(\Gamma^T \mathbf{K} \Gamma) \\ & + \frac{\eta}{2} Tr(\mathbf{U} \mathbf{U}^T) \\ & + \frac{\rho}{2} (\chi^{rT} \chi^r + \chi^{cT} \chi^c), \end{aligned} \quad (18)$$

where $d(\cdot)$ denotes the diagonal of a matrix and $\mathbf{1}$ is a vector of ones.

The coefficient matrix Γ is obtained by acquiring the partial derivative of the objective function Q with respect to Γ , and beginning at the derivation of Q :

$$\begin{aligned} \frac{\partial Q}{\partial \Gamma} = & \frac{1}{2\sigma^2} [-2\mathbf{K} \mathbf{P} \mathbf{T} + 2\mathbf{K} d(\mathbf{P} \mathbf{1}) \mathbf{S} + 2\mathbf{K} d(\mathbf{P} \mathbf{1}) \mathbf{K} \Gamma] \\ & + \alpha \mathbf{K} \Gamma + \frac{\eta}{2} \frac{\partial \mathcal{L}}{\partial \Gamma} + \frac{\rho}{2} \frac{\partial \mathfrak{X}}{\partial \Gamma}, \end{aligned} \quad (19)$$

where the partial derivative of the local structure preservation \mathcal{L} with respect to Γ :

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \Gamma} = & (\frac{\partial \mathbf{U}}{\partial \Gamma})^T \frac{\partial Tr(\mathbf{U} \mathbf{U}^T)}{\partial \mathbf{U}} \\ = & 2(\mathbf{H}^S - \mathcal{K} \mathbf{I})^T [(\mathbf{H}^{\hat{T}} - \mathcal{K} \mathbf{I}) \mathbf{P} \mathbf{T} - (\mathbf{H}^S - \mathcal{K} \mathbf{I})(\mathbf{S} + \mathbf{K} \Gamma)] \\ = & 2\mathbf{K}^T \mathcal{B}^T (\mathcal{A} - \mathcal{B} \mathbf{K} \Gamma) \\ = & 2\mathbf{K}^T \mathcal{B}^T \mathcal{A} - 2\mathbf{K}^T \mathcal{B}^T \mathcal{B} \mathbf{K} \Gamma, \end{aligned} \quad (20)$$

with $\mathcal{A} = (\mathbf{H}^{\hat{T}} - \mathcal{K} \mathbf{I}) \mathbf{P} \mathbf{T} - \mathcal{B} \mathbf{S}$, $\mathcal{B} = \mathbf{H}^S - \mathcal{K} \mathbf{I}$, \mathbf{I} is a identity matrix. And $\mathbf{K}^T = \mathbf{K}$ since the Gaussian kernel \mathbf{K} is a symmetric matrices, hence

$$\frac{\partial \mathcal{L}}{\partial \Gamma} = 2\mathbf{K}(\mathcal{B}^T \mathcal{A} - \mathcal{B}^T \mathcal{B} \mathbf{K} \Gamma) \quad (21)$$

The partial derivative of the space curvature preservation \mathfrak{X} with respect to Γ :

$$\begin{aligned} \frac{\partial \mathfrak{X}}{\partial \Gamma} = & \frac{\partial Tr(\chi^{rT} \chi^r)}{\partial \Gamma} + \frac{\partial Tr(\chi^{cT} \chi^c)}{\partial \Gamma} \\ = & (\frac{\partial \chi^r}{\partial \Gamma})^T \frac{\partial Tr(\chi^{rT} \chi^r)}{\partial \chi^r} + (\frac{\partial \chi^c}{\partial \Gamma})^T \frac{\partial Tr(\chi^{cT} \chi^c)}{\partial \chi^c} \\ = & 2[(\frac{\partial \mathcal{L}(\mathcal{F}(\mathbf{S}, \Gamma))}{\partial \Gamma})^T \frac{\partial \chi^r}{\partial \mathcal{L}(\mathcal{F}(\mathbf{S}, \Gamma))}]^T \chi^r \\ & + 2[(\frac{\partial \mathcal{L}(\mathcal{F}(\mathbf{S}, \Gamma))}{\partial \Gamma})^T \frac{\partial \chi^c}{\partial \mathcal{L}(\mathcal{F}(\mathbf{S}, \Gamma))}]^T \chi^c \\ = & 2(\frac{\partial \chi^r}{\partial \mathcal{L}(\mathcal{F}(\mathbf{S}, \Gamma))})^T \frac{\partial \mathcal{L}(\mathcal{F}(\mathbf{S}, \Gamma))}{\partial \Gamma} \chi^r \\ & + 2(\frac{\partial \chi^c}{\partial \mathcal{L}(\mathcal{F}(\mathbf{S}, \Gamma))})^T \frac{\partial \mathcal{L}(\mathcal{F}(\mathbf{S}, \Gamma))}{\partial \Gamma} \chi^c \\ = & 2(\mathcal{W}^R \mathcal{Y}^{-1} - \mathcal{W}^L \mathcal{Y}^{-1})^T (\mathcal{W}^L \mathcal{Y}^{-1} \bar{\mathbf{K}} \\ & + \mathcal{W}^R \mathcal{Y}^{-1} \bar{\mathbf{K}})^T \mathbf{K}(\bar{\mathbf{S}} + \bar{\mathbf{K}} \Gamma) \\ = & 2\mathbf{K} \mathcal{V} \bar{\mathbf{K}} \Gamma - 2\mathbf{K} \mathcal{V} \bar{\mathbf{S}}, \end{aligned} \quad (22)$$

where $\mathcal{V} = (\mathcal{W}^L \mathcal{Y}^{-1} \bar{\mathbf{K}} + \mathcal{W}^R \mathcal{Y}^{-1} \bar{\mathbf{K}})^T (\mathcal{W}^R - \mathcal{W}^L) \mathcal{Y}^{-1}$, and $\bar{\mathbf{K}} = \begin{pmatrix} \mathbf{K} \\ \mathbf{0} \end{pmatrix}$, $\bar{\mathbf{S}} = \begin{pmatrix} \mathbf{S} \\ \mathbf{0} \end{pmatrix}$.

Therefore, the derivation of Q is formulated as:

$$\begin{aligned} \frac{\partial Q}{\partial \Gamma} = & \frac{1}{2\sigma^2} [-2\mathbf{K} \mathbf{P} \mathbf{T} + 2\mathbf{K} d(\mathbf{P} \mathbf{1}) \mathbf{S} + 2\mathbf{K} d(\mathbf{P} \mathbf{1}) \mathbf{K} \Gamma] \\ & + \alpha \mathbf{K} \Gamma + \eta \mathbf{K}(\mathcal{B}^T \mathcal{A} - \mathcal{B}^T \mathcal{B} \mathbf{K} \Gamma) + \rho \mathbf{K}(\mathcal{V} \bar{\mathbf{K}} \Gamma + \mathcal{V} \bar{\mathbf{S}}). \end{aligned} \quad (23)$$

Let the derivation be equal to zero for obtaining an optimal Γ :

$$\begin{aligned} \frac{\partial Q}{\partial \Gamma} = 0 \\ \Downarrow \\ \frac{1}{2\sigma^2} [-2\mathbf{K} \mathbf{P} \mathbf{T} + 2\mathbf{K} d(\mathbf{P} \mathbf{1}) \mathbf{S} + 2\mathbf{K} d(\mathbf{P} \mathbf{1}) \mathbf{K} \Gamma] \\ + \alpha \mathbf{K} \Gamma + \eta \mathbf{K}(\mathcal{B}^T \mathcal{A} - \mathcal{B}^T \mathcal{B} \mathbf{K} \Gamma) + \rho \mathbf{K}(\mathcal{V} \bar{\mathbf{K}} \Gamma + \mathcal{V} \bar{\mathbf{S}}) = 0 \\ \Downarrow \\ \mathbf{K}[-\mathbf{P} \mathbf{T} + d(\mathbf{P} \mathbf{1}) \mathbf{S} + d(\mathbf{P} \mathbf{1}) \mathbf{K} \Gamma \\ + \sigma^2 \alpha \Gamma + \sigma^2 \eta (\mathcal{B}^T \mathcal{A} - \mathcal{B}^T \mathcal{B} \mathbf{K} \Gamma) + \sigma^2 \rho (\mathcal{V} \bar{\mathbf{K}} \Gamma + \mathcal{V} \bar{\mathbf{S}})] = 0 \\ \Downarrow \\ \mathbf{P} \mathbf{T} - d(\mathbf{P} \mathbf{1}) \mathbf{S} - \sigma^2 \eta \mathcal{B}^T \mathcal{A} - \sigma^2 \rho \mathcal{V} \bar{\mathbf{S}} \\ = d(\mathbf{P} \mathbf{1}) \mathbf{K} \Gamma + \sigma^2 \alpha \Gamma - \sigma^2 \eta \mathcal{B}^T \mathcal{B} \mathbf{K} \Gamma + \sigma^2 \rho \mathcal{V} \bar{\mathbf{K}} \Gamma \\ \Downarrow \\ \Gamma = [d(\mathbf{P} \mathbf{1}) \mathbf{K} + \alpha \sigma^2 \mathbf{I} - \eta \sigma^2 \mathcal{B}^T \mathcal{B} \mathbf{K} + \rho \sigma^2 \mathcal{V} \bar{\mathbf{K}}]^{-1} \\ [\mathbf{P} \mathbf{T} - d(\mathbf{P} \mathbf{1}) \mathbf{S} - \eta \sigma^2 \mathcal{B}^T \mathcal{A} - \rho \sigma^2 \mathcal{V} \bar{\mathbf{S}}] \end{aligned} \quad (24)$$

E. Parameter Setting and Complexity Analysis

1) *Parameter Setting*: There are two types of parameters used in the proposed method: (i) provided by literatures,

and (ii) originally proposed in this work. The default values of originally proposed parameters are set according to an initial trial-and-error experiment using a small set of low-altitude aerial images.

Already provided:

- the outlier weighting parameter w is set to 5, and updated using $w \leftarrow 1 - \frac{N_p}{N}$ [20];
- the covariance $\sigma^2 = \frac{1}{2MN}(M \cdot Tr(\mathbf{S}^T \mathbf{S}) - 2sum(\mathbf{S})^T + N \cdot Tr(\mathbf{T}^T \mathbf{T}))$ as the initialization, and is updated using $\sigma^2 \leftarrow \frac{1}{2N_p}|Tr(\mathbf{T}^T d(\mathbf{P}^T \mathbf{1}) \mathbf{T}) - 2Tr(\mathbf{S}^{*T} \mathbf{P} \mathbf{T}) + Tr(\mathbf{S}^{*T} d(\mathbf{P} \mathbf{1}) \mathbf{S}^*)|$ with $sum(\cdot)$ denotes the sum of each item. We set a lower bound of σ^2 having $\nu = 10^{-6}$ [20];
- each entry in the initialization of the coefficient matrix Γ is equal to zero [20];
- the number of the bins in shape context has the size of 5×12 [56];
- the neighbors in the local structure constraint are detected by the k-dimensional tree (k-d tree) algorithm, and the neighbors' number K is set to 5 [21];
- δ is set to 0.999 for generating the Gaussian components density matrix [22];
- the initialization of trade-off parameters is set as $\alpha = 2$ [20], $\eta = \rho = 5$ [60].

Originally proposed:

- the maximum iteration number t_{max} is set to 80;
- the initial threshold τ in the pre-matching is set as $\tau_{init} = 2.2$, where 2.2 is the largest threshold for detecting the most reliable inlier pairs. The step parameter $\Delta = (1.8 - 1.2)/(t_a/d)$ is set to 0.05 for updating τ in every d iterations, where 1.8 is a good threshold with 95% inlier rate, 1.2 is a suitable threshold for mostly feature point extraction, $t_a \approx 60$ is the average of iterative numbers;
- the cycle parameter d is set to 5;
- the covariance of the Gaussian distance β in the outlier distribution cost matrix is set to 0.5, where 0.5 is to suitably punish stray inliers since the feature point set is ranged in $[0,1]$ after normalization;
- the row and column numbers in the space curvature constraint are set as $r = 100$ and $c = 100$, where 100×100 lattice points is enough to represent the image size of 1000×1000 with an order of magnitude in number. The source point lattice $\mathcal{L}(\mathbf{S})$ used in the TPS model is set as a $r \times c$ regular point lattice.

2) *Complexity Analysis*: \mathcal{K} nearest neighbor searching by the k-d tree has a time complexity $\mathcal{O}((M + \mathcal{K})\log M)$. The shape context chi-square cost matrix takes $\mathcal{O}(MN)$, and the transformed point lattice requires $\mathcal{O}(X^r Y^r M)$. The derivative function Eq. 24 requires $\mathcal{O}(rcM^2)$ due to the existence of $(M + 3) \times (M + 3)$ matrix \mathcal{Y} and $rc \times (M + 3)$ matrix \mathcal{W} . Overall, the computational complexity of the proposed method is $\mathcal{O}(rcM^2)$, and its runtime with different numbers of feature points is shown in Fig. 6.

III. EXPERIMENTAL RESULTS

The performance of our method is compared with four types of nine state-of-the-art methods:

Algorithm 1 Non-Rigid Image Registration With Dynamic Gaussian Component Density and Space Curvature Preservation

```

input :  $I^r$  and  $I^s$ 
output:  $\mathbf{S}^*$ ,  $\mathbf{P}$  and  $I^*$ 

1 Extract  $\mathbf{S}$ ,  $\lambda(\mathbf{S})$  from  $I^s$ , and  $\mathbf{T}$ ,  $\lambda(\mathbf{T})$  from  $I^r$ ;
2 Initialize  $w$ ,  $\sigma^2$ ,  $\Gamma$ ,  $\beta$ ,  $\alpha$ ,  $\eta$ ,  $\rho$ ,  $\tau = \tau_{init}$  and  $\mathcal{L}(\mathbf{S})$ ;
3 Construct the Gaussian kernel  $\mathbf{K}$ ;
4 do
   5     For the 1st and every d times iteration:
   6         Obtain the index set  $\mathcal{I}$  via  $\mathcal{M}(\mathbf{S}, \mathbf{T})$  with  $\tau$ ;
   7         Compute the density matrix  $\mathbf{C}$  using Eq. 11
   8         with the linear assignment;
   9         Update  $\tau \leftarrow \tau - \Delta$ ;
  10     end
  11     E-Step:
  12         Compute the posterior probability matrix  $\mathbf{P}$ 
  13         using Eq. 2;
  14         Compute the putative corresponding target set
  15          $\hat{\mathbf{T}} = \mathbf{P} \mathbf{T}$ ;
  16     end
  17     M-Step:
  18         Update the coefficient matrix  $\Gamma$  using Eq. 24;
  19         Compute the transformed point set
  20          $\mathbf{S}^* = \mathcal{F}(\mathbf{S}, \Gamma)$ ;
  21         Update  $\sigma^2$  and  $w$ ;
  22     end
  23     Annealing trade-off parameters  $\alpha$ ,  $\eta$  and  $\rho$ 
24 while 5 is not convergent and  $\sigma^2 \leq \nu$  and  $\tau \leq \tau_{max}$ ;
25 Compute the transformed image  $I^*$  using Eq. 7.

```

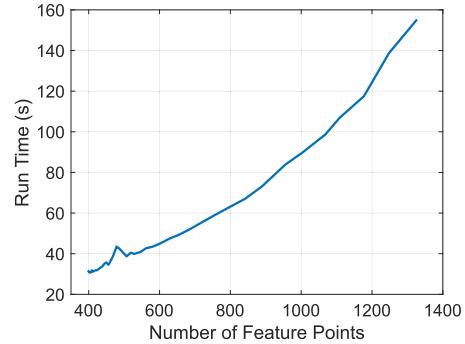


Fig. 6. Runtime of the proposed method with different numbers of feature points under a laptop with a 2.60GHz Intel Core CPU and a 16GB RAM.

- three point set registration methods: CPD [20], GLMDTPS [21] and PRGLS [22];
- one feature matching method: VFC [50];
- one feature-based image registration method: LLT [13];
- four area-based image registration methods: Hermosillo *et al.* [61], SIFT Flow [62], Lin *et al.* [63], and Shen *et al.* [3].

All experiments are implemented in MATLAB R2016a on a laptop with a 2.60GHz Intel Core CPU and a 16GB RAM. The experimental datasets consist of three types:

- (A) 14 pairs of multi-spectral natural images with a low viewpoint change, and each image pair has a set of human labeled corresponding points as the ground truth provided by Shen *et al.* [3];
- (B) 36 pairs of low-altitude aerial images captured by a DJI Phantom 4 Pro UAV with a relatively large viewpoint change, and image resolutions range from 800×400 to 1600×800 pixels. We label a set of corresponding points as the ground truth for each image pair, where at least 20 pairs are well-distributed and selected on easily identifiable places around interest areas;
- (C) Fundus Image Registration Dataset (FIRE): This dataset consists of 129 retinal images forming 134 image pairs, and these images were acquired with a Nidek AFC-210 fundus camera, which acquires images with a resolution of 2912x2912 pixels and a FOV of 45° both in the x and y dimensions. The ground truth for this dataset is provided by Hernandez-Matas *et al.* [64].

A. Evaluation Setting

Two common experimental evaluation methods are employed to access the performances of our method on two types of experiments: (I) image registration and (II) feature matching, respectively.

For type (I), the root mean square error (RMSE) and the maximum error (ME) are used to measure the image registration error, and defined as follows:

$$RMSE = \sqrt{\frac{1}{N^G} \sum_{n=1}^{N^G} \|a_n^G - b_n^G\|^2},$$

$$ME = \max(\{\sqrt{\|a_n^G - b_n^G\|^2}\}_{n=1}^{N^G}), \quad (25)$$

where a_n^G and b_n^G are the n^{th} pair human labeled corresponding points in images, N^G is the count of ground truths, and $\max(\cdot)$ denotes the maximum of a set. In addition, the failed registration case is determined by whether the image structure in the transformed image is significantly destroyed.

For type (II), we follow the same assessment criteria in PRGLS [22], LLT [13] and VFC [50], namely recall, precision, and the F_1 measure, where the F_1 measure is used to evaluate the balance between recall and precision. Their definitions are given as follows:

$$Recall = \frac{TP}{TP + FN},$$

$$Precision = \frac{TP}{TP + FP},$$

$$F_1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}, \quad (26)$$

where TP denotes the true positive (the true inlier matching), FP denotes false positive (the false inlier matching), and FN denotes the false negative (the false negative of outliers). To establish the ground truth, we first apply SIFT for a pre-matching, and then confirm the matching correctness manually. The estimated inlier index set \mathcal{I} is obtained by $\mathcal{I} = \{i : (\mathbf{P}^T \mathbf{1}) > 0.75\}$ for our method, CPD, and PRGLS. For GLMDTPS, it is obtained by its final binary matrix.

TABLE I
REGISTRATION ERRORS (UNIT: PIXELS) OF OUR METHOD AND STATE-OF-THE-ART METHODS ON DATASET (A)

Method	RMSE	ME	Failed count
Ours (SIFT)	3.20	5.67	1
Ours (SURF)	3.53	6.33	1
Ours (PSO-SIFT)	2.99	3.13	0
Ours (SAR-SIFT)	4.39	14.89	0
SIFT	23.21	92.72	1
SURF	38.35	102.91	1
PSO-SIFT	7.38	77.57	0
SAR-SIFT	22.55	187.54	0
SIFT Flow	11.47	18.32	0
Hermosillo <i>et al.</i>	13.81	16.57	0
Lin <i>et al.</i>	5.59	7.33	0
Shen <i>et al.</i>	2.95	4.56	0
VFC	16.15	45.57	0
LLT	32.28	54.39	4

TABLE II
REGISTRATION ERRORS (UNIT: PIXELS) OF OUR METHOD AND STATE-OF-THE-ART METHODS ON DATASET (B)

Method	RMSE	ME	Failed count
Ours (SIFT)	5.37	6.57	0
Ours (SURF)	8.35	26.64	2
Ours (PSO-SIFT)	9.04	30.01	7
Ours (SAR-SIFT)	23.07	47.33	4
Ours (SIFT)- V₁	7.9276	9.9170	3
Ours (SIFT)- V₂	11.9917	14.2304	10
SIFT	83.92	301.83	4
SURF	121.82	278.26	2
PSO-SIFT	43.76	171.57	1
SAR-SIFT	120.76	163.43	1
SIFT Flow	80.55	215.39	21
VFC	35.34	256.77	13
LLT	39.98	251.66	12

For LLT and VFC, it is outputted by their default setting. The true matching is determined by a manual sifting.

B. Image Registration Experiments

In this section, we first evaluate the performances of our method with four different feature extraction methods on image registration as well as compare with six state-of-the-art methods, and then test the role of the proposed spatial structure preservation.

1) *Experimental Comparison on Image Registration*: At first, in order to test the generality of the proposed method in different feature extractions, four feature extraction methods: SIFT [17], SURF [18], PSO-SIFT [65], and SAR-SIFT [66], are implemented as the X feature described in section II-B, respectively. The performances of the four implementations are evaluated on the above datasets (A), (B) and (C). Fig. 7 demonstrates five representative registration examples for the four different implementations and their original methods, and registration errors of these implementations and their original methods for datasets (A), (B) and (C) are listed in the first eight rows of Table I, Table II, and Table III, respectively.

We subsequently further compare with the other six methods, where VFC and LLT compared on all datasets, and SIFT Flow, Hermosillo *et al.* [61], Lin *et al.* [63], and

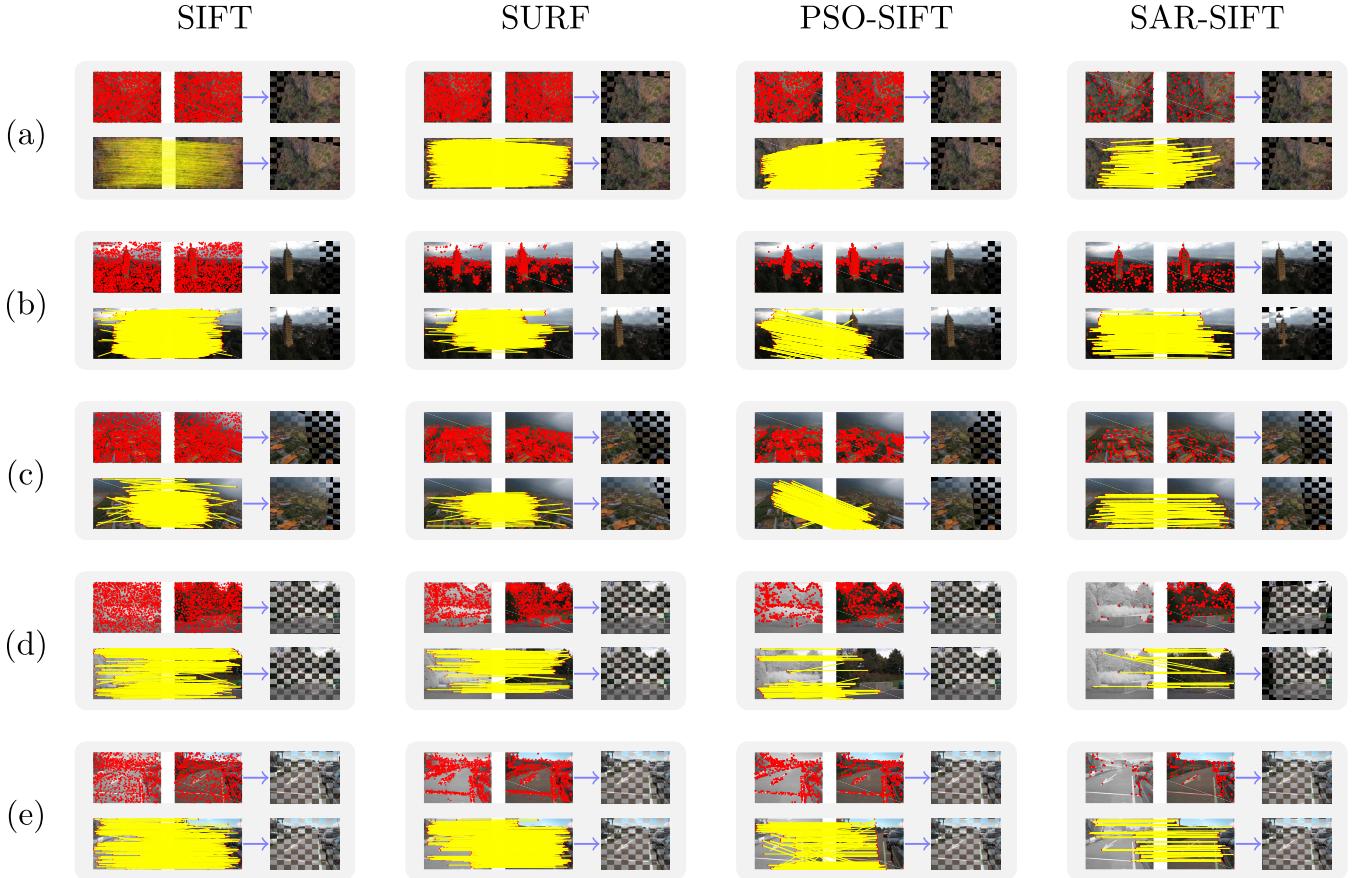


Fig. 7. Representative registration examples of the four different implementations and their original methods (SIFT [17], SURF [18], PSO-SIFT [65], and SAR-SIFT [66]). In each entry, feature extraction, feature matching results and 10×10 registration checkboard (formed by the transformed image and the reference image) between the sensed image and the reference image are demonstrated, respectively. For registration checkboard, from the top row to the bottom row show the results performed by our implementation and its original method. Red circles denote the extracted feature points, and yellow lines denote the matching results determined by its original method. Additionally, image pairs (a), (b), and (c) belong to dataset (B), and (d), (e) belong to dataset (A).

TABLE III
REGISTRATION ERRORS (UNIT: PIXELS) OF OUR METHOD AND
STATE-OF-THE-ART METHODS ON DATASET (C)

Method	RMSE	ME	Failed count
Ours (SIFT)	22.034	143.8923	0
Ours (SURF)	44.6570	348.3849	0
Ours (PSO-SIFT)	9.04	30.01	7
Ours (SAR-SIFT)	23.07	47.33	4
SIFT	109.5909	223.7559	0
SURF	58.1241	286.8473	0
PSO-SIFT	119.7061	327.9719	66
SAR-SIFT	87.7262	285.3459	0
SIFT Flow	30.2384	38.6891	0
VFC	31.5814	360.3299	27
LLT	29.1717	360.5341	36

Shen *et al.* [3], compared on dataset (A). The registration errors of Hermosillo *et al.* [61], Lin *et al.* [63], and Shen *et al.* [3], on the dataset (A) are provided by Su *et al.* [67], and the registration errors of SIFT Flow, VFC, and LLT on all datasets (A), (B) and (C) are obtained based on running their published codes with their default settings. The performances of the six methods are also listed in Table I, Table II, and Table III, respectively, and seven representative registration examples are given in Fig. 8.

Based on these results, our method implemented with PSO-SIFT gives the best performance on the dataset (A), and the proposed method implemented with SIFT provides the best performance on the dataset (B) and (C). SIFT Flow, Hermosillo *et al.* [61], Lin *et al.* [63], Shen *et al.* [3], and VFC register image pairs with no failure on dataset (A). However, the area-based image registration method SIFT Flow gives a relatively lower performance on dataset (B) (with large viewpoint change) than dataset (A) (with spectral change but low viewpoint change), especially 21 failed registration in dataset (B). SIFT Flow has the second best performance on dataset (C). However, the issues of SIFT Flow are that the non-overlap image information is abandoned and the resolution of registered image only has a 25% resolution of the original image.

For the four different implementations as shown in Fig. 7, SIFT extracted the maximum numbers of feature points in both dataset (A) and (B), where such feature points involve large available potential image information for recovering a more accurate image transformation, but also include a large set of outliers, especially when image pairs have a large viewpoint change or a low overlap rate. PSO-SIFT extracted more reliable and sufficient feature points on dataset (A) so that the recovered image transformations are more robust.

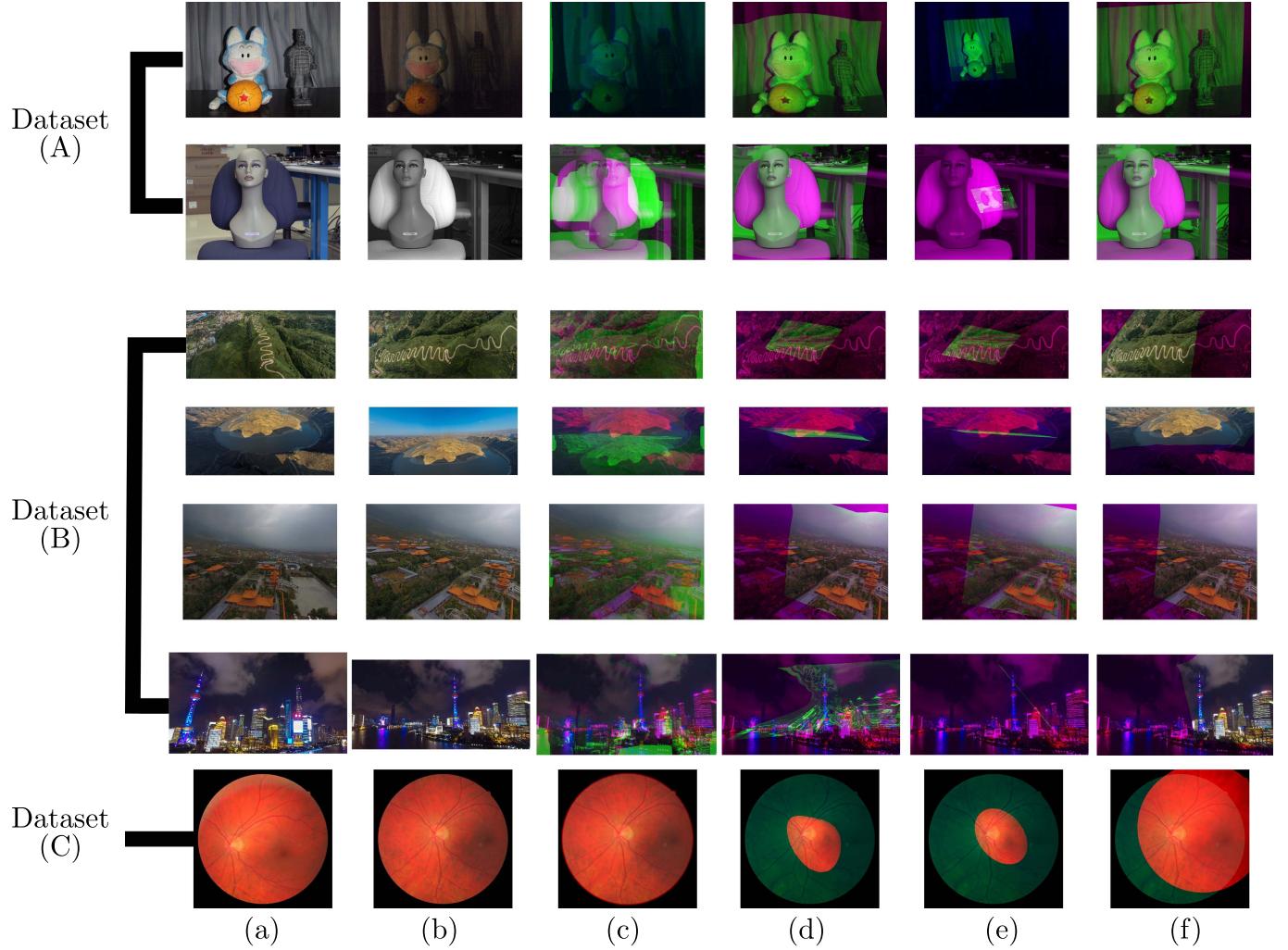


Fig. 8. Registration examples on datasets (A), (B) and (C). (a) the sensed images, (b) the reference images, (c) results of SIFT Flow [62], (d) results of VFC [50]. (e) results of LLT [13], and (f) results of our method implemented with SIFT [17], PSO-SIFT [65] and SIFT [17] on datasets (A), (B) and (C), respectively.

2) The Performance of the Proposed Spatial Structure Preservation: We test the performance of the proposed spatial structure preservation, i.e., the second contribution in this work, by using two imperfect implementations denoted by \mathbf{V}_1 , and \mathbf{V}_2 . These implementations have the same main process, but \mathbf{V}_1 and \mathbf{V}_2 work without the spatial structure preservation and the Tikhonov regularization, respectively. The performances of the proposed method and the two imperfect implementations using SIFT are tested on dataset (B). Fig. 9 shows three representative registration examples and registration errors are given in Table II. As results show, the proposed method has the best performance, and \mathbf{V}_1 gives wrong local structures and contorted space grids, and \mathbf{V}_2 performs more poorly as shown in Fig. 9.

C. Feature Matching Experiments

In this section, we first evaluate the performance of our method on feature matching as well as compare with five state-of-the-art methods, and then test the role of the proposed dynamic Gaussian component density.

1) Experimental Comparison on Feature Matching: In order to evaluate the ability of each method on recovering inliers as many as possible, two scenarios of feature matching experiments are designed: (i) fixing the inlier number as 50 while increasing the inlier ratio from 0.2 to 1.0, with an increment of 0.2, (ii) fixing the inlier ratio as 0.2 while increasing the inlier number from 50 to 250, with an increment of 50. All outliers are added randomly within the boundaries of images. For each degree in (i) and (ii), 50 times random experiments are performed. We followed the same approach to determine the ground truth in [13], [27], and [28], where the matching correctness is determined by first using their own methods to establish rough correspondences and then sifting the results manually.

The proposed method is compared with CPD, GLMDTPS, PRGLS, VFC and LLT on dataset (B) using the aforementioned type (I) evaluation. The experimental comparison results are shown in Fig. 10, and two representative feature matching examples by the three feature matching methods (ours, LLT and VFC) are given in Fig. 11. As results show,

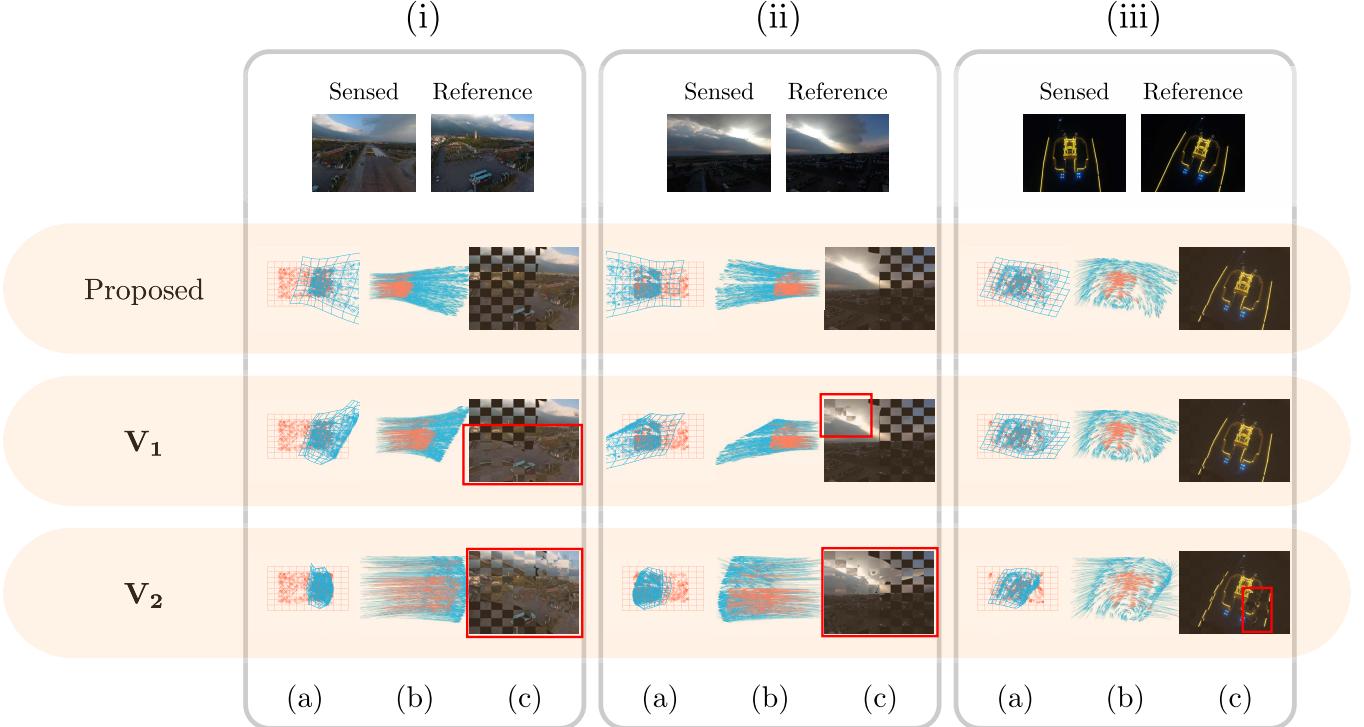


Fig. 9. Three representative registration examples of the proposed method and two imperfect implementations \mathbf{V}_1 , and \mathbf{V}_2 . The results for each method consist of three visualizations. (a): the target grid with target feature points, and the warped grid with transformed feature points, where orange dots, crosses and dotted lines denote target inliers, target outliers and the target grid, respectively, and blue dots, crosses and lines are transformed inliers, transformed outliers and the warped grid, respectively. (b): the vector field for the feature point transformation, starting from source feature points to transformed feature points, where blue and orange arrows denote the transformation of outliers and inliers, respectively. (c): 10 × 10 montages, where the registration errors are highlighted using the red rectangle.

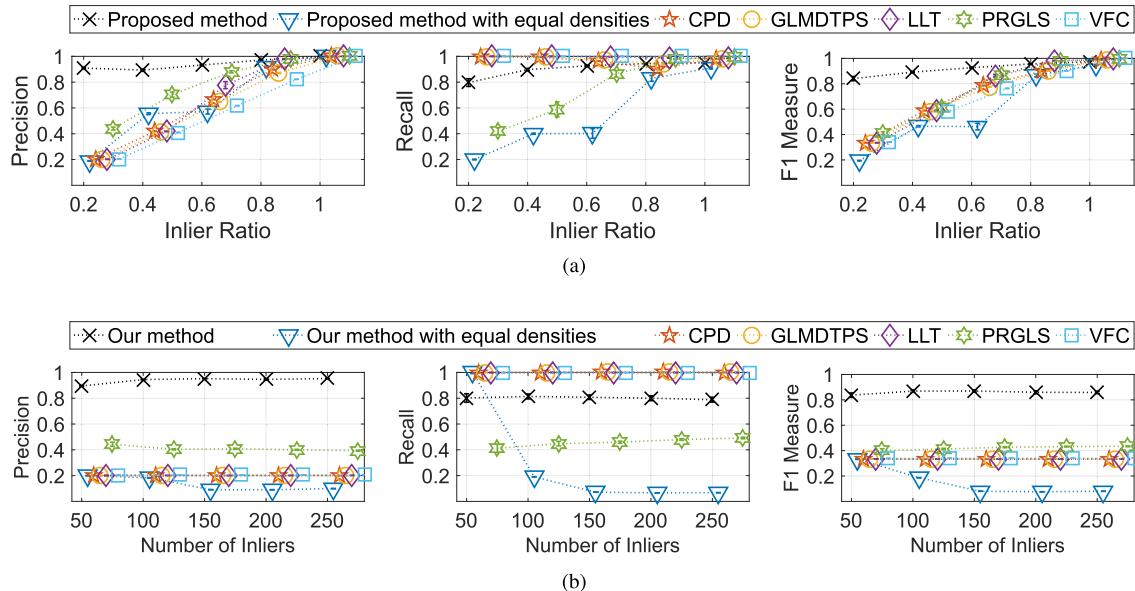


Fig. 10. Quantitative comparison of our method, an imperfect implementation, CPD [20], GLMDTPS [21], LLT [13], PRGLS [22], and VFC [50] on feature matching. (a) is the results of scenario (i), and (b) is the results of scenario (ii). The error bars indicate the standard deviations of the respective averages of 50 random experiments.

our method provides the highest F_1 score and a robust performance at all inlier ratios and inlier numbers in the two scenarios. CPD, GLMDTPS, LLT, and VFC show the high recall scores at all inlier ratios and inlier numbers, but performed poorly in the F_1 measure.

2) The Performance of the Proposed Dynamic Gaussian Component Density: We test the performance of the proposed dynamic Gaussian component density, i.e., the first contribution, by using an imperfect implementation, which has the same main process with the proposed method,

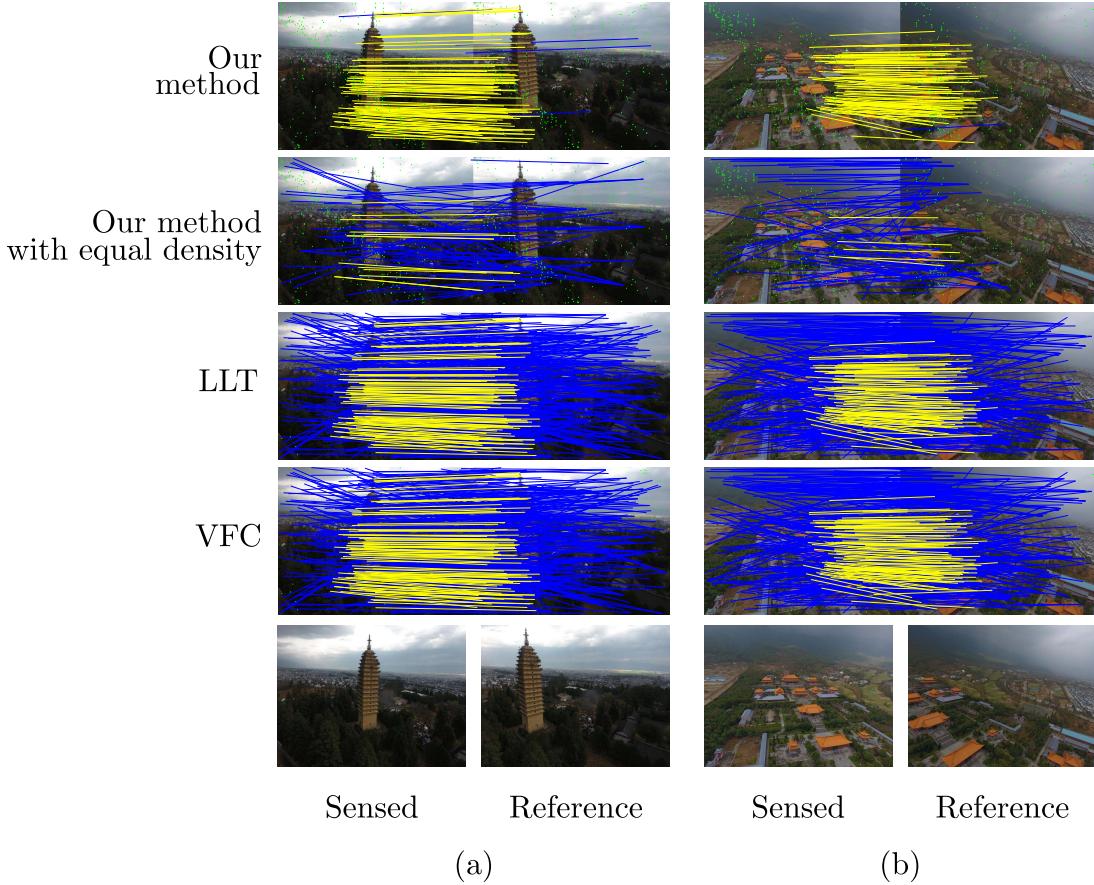


Fig. 11. Two representative feature matching examples of our method, an imperfect implementation, LLT [13], and VFC [50] on dataset (B). For (a) and (b), green dots and yellow lines denote correctly identified outliers and inlier pairs, respectively, and blue lines denote falsely identified inlier pairs. Overall, more yellow lines and green dots give a more accurate matching result, conversely, more blue lines denote a false matching result.

but implemented by a traditional 'static' density (the SIFT threshold is set by the default value 1.5). The experimental comparison results are also given in Fig. 10, and two feature matching examples are shown in the first and second rows of Fig. 11. The proposed method with a static density shows acceptable F_1 scores when the inlier ratio is higher than 0.8, and all failed at low inlier numbers.

IV. CONCLUSION

In this work, we have present a feature-based non-rigid image registration method. In order to maximize the available potential image information and recover the transformations robustly, two ideas were proposed: (i) the dynamic Gaussian component density was designed to recover the most reliable inlier pairs at the begin of registration, and then gradually changed to search the maximum number of inlier pairs at the end of registration; (ii) the spatial structure preservation, which consists of the image transformation space curvature preservation and the local spatial structure constrain, was proposed to constrain not only the local structure of feature points, but also the image transforming cost during point set registration. Experiments on image registration and feature matching were performed to show the performance of the proposed method. Our method, implemented by four different popular feature

extraction methods, compared with four types of nine state-of-the-art methods and shows the best performance in most scenarios.

Although the proposed method outperforms the other methods, it still has a relatively high computational complexity. How to reduce the computational complexity and improving the runtime by parallel computing will be the focus of future work. Moreover, the proposed method can deal with partial matching/registration problems, but it might fail when a relatively large part of image content in the overlapping area is missing or changed in one of the images.

ACKNOWLEDGMENT

The authors wish to thank David G. Lowe, Herbert Bay, Andriy Myronenko, Wenping Ma, Flora Dellinger, Xiaoyong Shen, Ce Liu, Jiayi Ma and Gehua Yang for providing their implementation source codes and experimental datasets, which facilitate the comparison experiments greatly.

REFERENCES

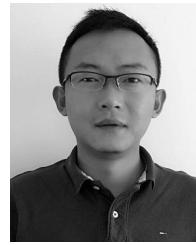
- [1] P. Zhou, G. Cheng, Z. Liu, S. Bu, and X. Hu, "Weakly supervised target detection in remote sensing images based on transferred deep features and negative bootstrapping," *Multidimensional Syst. Signal Process.*, vol. 27, no. 4, pp. 925–944, 2016.

- [2] A. Sameer, S. Noah, S. Ian, S. M. Seitz, and R. Szeliski, "Building Rome in a day," in *Proc. Int. Conf. Comput. Vis.*, 2009, pp. 72–79.
- [3] X. Shen, L. Xu, Q. Zhang, and J. Jia, "Multi-modal and multi-spectral registration for natural images," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 309–324.
- [4] F. Song, M. Li, Y. Yang, K. Yang, X. Gao, and D. Ting, "Small UAV based multi-viewpoint image registration for monitoring cultivated land changes in mountainous terrain," *Int. J. Remote Sens.*, vol. 39, no. 21, pp. 7201–7224, 2018.
- [5] F. Song *et al.*, "Multi-scale feature based land cover change detection in mountainous terrain using multi-temporal and multi-sensor remote sensing images," *IEEE Access*, to be published, doi: [10.1109/ACCESS.2018.2883254](https://doi.org/10.1109/ACCESS.2018.2883254).
- [6] K. Yang, Z. Yu, Y. Luo, Y. Yang, L. Zhao, and X. Zhou, "Spatial and temporal variations in the relationship between lake water surface temperatures and water quality—A case study of Dianchi lake," *Sci. Total Environ.*, vol. 624, pp. 859–871, May 2018.
- [7] J. Ma, Y. Ma, and C. Li, "Infrared and visible image fusion methods and applications: A survey," *Inf. Fusion*, vol. 45, pp. 153–178, Jan. 2019.
- [8] 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.
- [9] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, Oct. 2003.
- [10] R. Gonzalez and P. Wintz, *Digital Image Processing*, 2nd ed. Reading, MA, USA: Addison-Wesley, 1987.
- [11] 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, 2015.
- [12] R. N. Bracewell, *The Fourier Transform and Its Applications*. New York, NY, USA: McGraw-Hill, 1986.
- [13] 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.
- [14] J. Liang, X. Liu, K. Huang, X. Li, D. Wang, and X. Wang, "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.
- [15] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2, Sep. 1999, pp. 1150–1157.
- [16] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. Alvey Vis. Conf.*, Manchester, U.K., vol. 15, no. 50, 1988, pp. 5244–5210.
- [17] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [18] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.
- [19] H. Chui and A. Rangarajan, "A new point matching algorithm for non-rigid registration," *Comput. Vis. Image Understand.*, vol. 89, nos. 2–3, pp. 114–141, Feb. 2003.
- [20] 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.
- [21] Y. Yang, S. H. Ong, and K. W. C. Foong, "A robust global and local mixture distance based non-rigid point set registration," *Pattern Recognit.*, vol. 48, no. 1, pp. 156–173, 2015.
- [22] J. Ma, J. Zhao, and A. L. Yuille, "Non-rigid point set registration by preserving global and local structures," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 53–64, Jan. 2016.
- [23] S. Zhang, Y. Yang, K. Yang, Y. Luo, and S. H. Ong, "Point set registration with global-local correspondence and transformation estimation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2688–2696, doi: [10.1109/ICCV.2017.291](https://doi.org/10.1109/ICCV.2017.291).
- [24] J. Ma, J. Jiang, H. Zhou, J. Zhao, and X. Guo, "Guided locality preserving feature matching for remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4435–4447, Aug. 2018.
- [25] M. Gong, S. Zhao, L. Jiao, D. Tian, and S. Wang, "A novel coarse-to-fine scheme for automatic image registration based on SIFT and mutual information," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4328–4338, Jul. 2014.
- [26] J. Fan, Y. Wu, F. Wang, Q. Zhang, G. Liao, and M. Li, "SAR image registration using phase congruency and nonlinear diffusion-based SIFT," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 562–566, Mar. 2015.
- [27] K. Yang, A. Pan, Y. Yang, S. Zhang, S. H. Ong, and H. Tang, "Remote sensing image registration using multiple image features," *Remote Sens.*, vol. 9, no. 6, p. 581, 2017.
- [28] Z. Wei *et al.*, "A small UAV based multi-temporal image registration for dynamic agricultural terrace monitoring," *Remote Sens.*, vol. 9, no. 9, p. 904, 2017.
- [29] Y. Xiang, F. Wang, and H. You, "OS-SIFT: A robust SIFT-like algorithm for high-resolution optical-to-SAR image registration in suburban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 3078–3090, Jun. 2018.
- [30] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.
- [31] E. Tola, V. Lepetit, and P. Fua, "DAISY: An efficient dense descriptor applied to wide-baseline stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 815–830, May 2010.
- [32] R. Bouchiba and K. Besbes, "Automatic remote-sensing image registration using surf," *Int. J. Comput. Theory Eng.*, vol. 5, no. 1, p. 88, 2013.
- [33] I. Colomina and P. Molina, "Unmanned aerial systems for photogrammetry and remote sensing: A review," *ISPRS J. Photogram. Remote Sens.*, vol. 92, pp. 79–97, Jun. 2014.
- [34] M. S. Patel, N. M. Patel, and M. S. Holia, "Feature based multi-view image registration using SURF" in *Proc. Int. Symp. Adv. Comput. Commun. (ISACC)*, Sep. 2015, pp. 213–218.
- [35] H. Lin, P. Du, W. Zhao, L. Zhang, and H. Sun, "Image registration based on corner detection and affine transformation," in *Proc. 3rd Int. Congr. Image Signal Process. (CISP)*, vol. 5, Oct. 2010, pp. 2184–2188.
- [36] I. Misra, S. M. Moorthi, D. Dhar, and R. Ramakrishnan, "An automatic satellite image registration technique based on Harris corner detection and Random Sample Consensus (RANSAC) outlier rejection model," in *Proc. 1st Int. Conf. Recent Adv. Inf. Technol.(RAIT)*, Mar. 2012, pp. 68–73.
- [37] Y. Ye and J. Shan, "A local descriptor based registration method for multispectral remote sensing images with non-linear intensity differences," *ISPRS J. Photogramm. Remote Sens.*, vol. 90, pp. 83–95, Apr. 2014.
- [38] B. Biswas, A. Chakrabarti, and K. N. Dey, "Image registration method using Harris Corner and modified Hausdorff distance with near set," in *Proc. IEEE 2nd Int. Conf. Recent Trends Inf. Syst. (ReTIS)*, Jul. 2015, pp. 393–398.
- [39] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proc. Eur. Conf. Comput. Vis.*, Berlin, Germany: Springer, 2006, pp. 430–443.
- [40] M. Butenuth *et al.*, "Integrating pedestrian simulation, tracking and event detection for crowd analysis," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Nov. 2011, pp. 150–157.
- [41] B. Sirmacek and P. Reinartz, "Feature analysis for detecting people from remotely sensed images," *J. Appl. Remote Sens.*, vol. 7, no. 1, p. 073594, 2013.
- [42] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary robust independent elementary features," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2010, pp. 778–792.
- [43] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 2564–2571.
- [44] C. Aguilera, F. Barrera, F. Lumbieras, A. D. Sappa, and R. Toledo, "Multispectral image feature points," *Sensors*, vol. 12, no. 9, pp. 12661–12672, 2012. [Online]. Available: <http://www.mdpi.com/1424-8220/12/9/12661>
- [45] J. Ma, J. Zhao, J. Jiang, H. Zhou, and X. Guo, "Locality preserving matching," *Int. J. Comput. Vis.*, to be published, doi: [10.1007/s11263-018-1117-z](https://doi.org/10.1007/s11263-018-1117-z).
- [46] Z. Zhang, "Iterative point matching for registration of free-form curves and surfaces," *Int. J. Comput. Vis.*, vol. 13, no. 2, pp. 119–152, 1994.
- [47] B. Jian and B. Vemuri, "Robust point set registration using Gaussian mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, pp. 1633–1645, Aug. 2011.
- [48] J. Ma, J. Zhao, J. Tian, Z. Tu, and A. L. Yuille, "Robust estimation of nonrigid transformation for point set registration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 2147–2154.
- [49] A. Basu, I. R. Harris, N. L. Hjort, and M. C. Jones, "Robust and efficient estimation by minimising a density power divergence," *Biometrika*, vol. 85, no. 3, pp. 549–559, 1998.

- [50] 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.
- [51] S. Zhang, K. Yang, Y. Yang, and Y. Luo, "Nonrigid image registration for low-altitude SUAV images with large viewpoint changes," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 4, pp. 592–596, Apr. 2018.
- [52] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. B, Methodol.*, vol. 39, no. 1, pp. 1–38, 1977.
- [53] R. M. Neal and G. E. Hinton, "A view of the EM algorithm that justifies incremental, sparse, and other variants," in *Learning in Graphical Models*. Dordrecht, The Netherlands: Springer, 1998, pp. 355–368.
- [54] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 6, pp. 567–585, Jun. 1989.
- [55] 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.
- [56] 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.
- [57] R. Jonker and A. Volgenant, "A shortest augmenting path algorithm for dense and sparse linear assignment problems," *Computing*, vol. 38, no. 4, pp. 325–340, Nov. 1987.
- [58] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*. Washington, DC, USA: V.H. Winston & Sons, 1977.
- [59] A. L. Yuille and N. M. Grzywacz, "A mathematical analysis of the motion coherence theory," *Int. J. Comput. Vis.*, vol. 3, no. 2, pp. 155–175, 1989.
- [60] S. Zhang, K. Yang, Y. Yang, Y. Luo, and Z. Wei, "Non-rigid point set registration using dual-feature finite mixture model and global-local structural preservation," *Pattern Recognit.*, vol. 80, pp. 183–195, Aug. 2018.
- [61] G. Hermosillo, C. Chefd'Hotel, and O. Faugeras, "Variational methods for multimodal image matching," *Int. J. Comput. Vis.*, vol. 50, no. 3, pp. 329–343, 2002.
- [62] C. Liu, J. Yuen, A. Torralba, J. Sivic, and W. T. Freeman, "SIFT flow: Dense correspondence across different scenes," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2008, pp. 28–42.
- [63] H. Yang, W.-Y. Lin, and J. Lu, "DAISY filter flow: A generalized discrete approach to dense correspondences," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3406–3413.
- [64] C. Hernandez-Matas, X. Zabulis, A. Triantafyllou, P. Anyfanti, S. Douma, and A. A. Argyros, "FIRE: Fundus image registration dataset," *J. Model. Ophthalmol.*, vol. 1, no. 4, pp. 16–28, 2017.
- [65] W. Ma *et al.*, "Remote sensing image registration with modified sift and enhanced feature matching," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 1, pp. 3–7, Jan. 2017.
- [66] F. Dellinger, J. Delon, Y. Gousseau, J. Michel, and F. Tupin, "SAR-SIFT: A SIFT-like algorithm for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 453–466, Jan. 2015.
- [67] H.-R. Su and S.-H. Lai, "Non-rigid registration of images with geometric and photometric deformation by using local affine Fourier-moment matching," in *Proc. CVPR*, Jun. 2015, pp. 2874–2882.



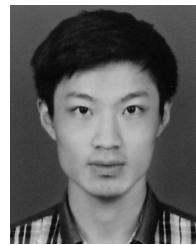
Zhuoqian Yang is currently pursuing the B.Eng. degree majoring in software engineering with Beihang University. He is currently a Visiting Student with the Engineering Research Center of GIS Technology in Western China, Ministry of Education, Yunnan Normal University, under the supervision of Prof. Y. Yang. His research interests include image registration, visual reasoning, and vision-language analysis.



Yang Yang received the M.A.Eng. degree from Waseda University, Japan, in 2007, and the Ph.D. degree from the National University of Singapore, Singapore, in 2013. He is currently an Associate Professor with the School of Information Science and Technology, Yunnan Normal University. His research interests cover computer vision, remote sensing, and medical imaging. He was selected into the Yunnan Province Ten-Thousand Talents Program in 2018.



Kun Yang received the master's degree from The University of New South Wales, Australia, in 1998. He is currently a Professor with the School of Information Science and Technology, Yunnan Normal University. His research interests cover geographic information system and remote sensing image processing.



Zi-Quan Wei is currently pursuing the B.Eng. degree with the School of Information Science and Technology, Yunnan Normal University. His research interests cover point set registration and remote sensing image processing.