Nonrigid Point Set Registration With Robust Transformation Learning Under Manifold Regularization

Jiayi Ma, Jia Wu, Ji Zhao, Junjun Jiang, Huabing Zhou, and Quan Z. Sheng

***Abstract*— This paper solves the problem of nonrigid point set registration by designing a robust transformation learning scheme. The principle is to iteratively establish point correspondences and learn the nonrigid transformation between two given sets of points. In particular, the local feature descriptors are used to search the correspondences and some unknown outliers will be inevitably introduced. To precisely learn the underlying transformation from noisy correspondences, we cast the point set registration into a semisupervised learning problem, where a set of indicator variables is adopted to help distinguish outliers in a mixture model. To exploit the intrinsic structure of a point set, we constrain the transformation with manifold regularization which plays a role of prior knowledge. Moreover, the transformation is modeled in the reproducing kernel Hilbert space, and a sparsity-induced approximation is utilized to boost efficiency. We apply the proposed method to learning motion flows between image pairs of similar scenes for visual homing, which is a specific type of mobile robot navigation. Extensive experiments on several publicly available data sets reveal the superiority of the proposed method over state-of-the-art competitors, particularly in the context of the degenerated data.**

***Index Terms*— Manifold regularization, nonrigid, point set registration, robust estimation, visual homing.**

I. Introduction

POINT set registration is a fundamental problem in computer vision, mobile robotics, and related fields [1]–[5] which is common in a wide variety of real-world tasks, such as structure from motion, panorama mosaic, image fusion, object detection and retrieval, visual homing, to name just a few. One of the reasons these tasks could be commonly casted into solving a point set registration problem is that the point is a simple and generic way of representing objects of interest. In this context, points usually represent the spatial locations of interest points extracted from an image, a shape contour, or an object surface, and the registration problem accordingly reduces to two subproblems, i.e., determining the correct correspondence between two given point sets and learning the underlying spatial transformation.

According to the characteristics of the data in specific applications, point set registration can be categorized into *rigid* or *nonrigid* registration. The former, which involves only a few transformation parameters, is relatively easy to handle (see [1], [2], [6] for a literature survey). By contrast, the latter is challenging because the underlying nonrigid transformations usually cannot be modeled in a simple parametric manner [7]. Despite these challenges, nonrigid registration remains desirable in quite a significant number of applications, such as handwritten character recognition, deformable image registration, visual homing, and so forth [8]. In this paper, we focus on the problem of nonrigid point set registration.

A. Prior Work

The iterated closest point (ICP) approach [2] is arguably one of the most classic approaches for point set registration, which assigns a certain pair of points, a binary indicator based on its nearest neighbors and utilizes the estimated correspondences to refine the transformation. The "hard" binary assignment of ICP can be replaced by soft ones, where a structured correspondence matrix is sought with consideration of parametric or nonparametric constraints [7], [9]. Ma *et al.* [10] introduced a nonrigid registration approach based on Gaussian fields, which was subsequently improved in [11] by using the inner distance shape context (SC) [12] rather than the original SC [13] to construct initial correspondences. More recently, some probabilistic methods [14]–[18] have been developed to accomplish the task. The kernel correlation-based method [14] assumes that the two-point sets can be modeled as two probability distributions, and their dissimilarity is measured based on kernel density estimates. Then, the work [15] representing the point sets using Gaussian mixture models (GMMs) further improves the scheme proposed in [14]. In [16]–[18], GMM is utilized to assign the point correspondence by estimating

Manuscript received November 16, 2017; revised April 8, 2018, August 1, 2018, and September 4, 2018; accepted September 25, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61773295, Grant 61503288, Grant 61501413, and Grant 41501505, in part by the Beijing Advanced Innovation Center for Intelligent Robots and Systems under Grant 2016IRS15, in part by the MQNS under Grant 9201701203, in part by MQ EPS under Grant 9201701455, and in part by the 2018 Collaborative Research Project between Macquarie University and Data61 under Grant 92307766. *(Corresponding author: Jia Wu.)*

J. Ma is with the Electronic Information School, Wuhan University, Wuhan 430072, China, and also with the Beijing Advanced Innovation Center for Intelligent Robots and Systems, Beijing Institute of Technology, Beijing 10081, China (e-mail: jyma2010@gmail.com).

J. Wu and Q. Z. Sheng are with the Faculty of Science and Engineering, Department of Computing, Macquarie University, Sydney, NSW 2109, Australia (e-mail: jia.wu@mq.edu.au; michael.sheng@mq.edu.au).

J. Zhao is with the ReadSense Ltd., Shanghai 200040, China (e-mail: zhaoji84@gmail.com).

J. Jiang is with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China, and also with the Peng Cheng Laboratory, Shenzhen 518055, China (e-mail: jiangjunjun@hit.edu.cn).

H. Zhou is with the School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan 430073, China (e-mail: zhouhuabing@gmail.com).

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

Digital Object Identifier 10.1109/TNNLS.2018.2872528

the parameters of a mixture via the maximum likelihood and expectation maximization (EM) algorithms. Specifically, a method based on coherent point drift (CPD) [16], referred to as global-local topology preservation (GLTP) [19], [20], is customized to cope with highly articulated deformations. The above-mentioned methods have achieved great success in handling both rigid and nonrigid registration. However, they ignore the local structure information among point sets that can be incorporated into feature descriptors. Thus, their performance degrades in complicated registration problems.

Another technical line of point set registration involves two stages: 1) correspondences are first built based on the similarity of local feature descriptors and 2) spatial transformation is then estimated according to the global geometric constraints. Representatives of this strategy include the SC [13] for 2-D cases, fast point feature histograms (FPFHs) for 3-D cases [21], and analogous variants [22], [23]. These methods perform reasonably well when the spatial transformation is not complex. However, they have exposed their limitations if there are errors in the correspondences which frequently occurs in real-world tasks, especially if the transformation is complex and/or the input data are contaminated by outliers (e.g., points in one set do not have corresponding points in the other set). To address this problem, Ma *et al.* [24], [25] introduced the two choices for achieving a robust estimation in the transformation. The first one, i.e., vector field consensus (VFC) [24], builds a complex model with extra (hidden) variables, thereby enabling the identification and rejection of outliers. The second one, i.e., L_2 estimator (L_2E) [25], uses an estimator that is less sensitive to outliers instead of the maximum likelihood estimator, which can be severely biased by outliers. These methods can properly manage complex nonrigid deformations. However, they calculate transformations solely based on the matched putative correspondences instead of fully utilizing the entire input data, which may produce an unsatisfying result if the data degradation is large.

B. Motivations and Contributions

As mentioned earlier, nonrigid registration requires solving the correspondence and transformation between two given point sets. Solving for the two variables simultaneously is difficult. An effective scheme is to solve one variable if the value of the other is given, [7], [13], that is, iteratively use the estimation of correspondence to improve the learning of transformation, and vice versa, until convergence. This paper mainly focuses on how to robustly learn transformation from a putative correspondence in such an iterative procedure.

In the past decades, various methods have been developed to address the problem of robust transformation learning from a given set of initial point correspondences [24], [26]; however, such problem remains a challenging task due to several aspects. First, putative correspondences are likely inaccurate because they are usually established upon local feature descriptors, which are inherently sensitive to noise, occlusion, and similar visual patterns in practice, leading to false correspondences (commonly known as outliers). In such a case, an outlier removal procedure is required to purify the initial correspondences. Second, a certain portion of input points may be excluded from the putative correspondence set due to the similarity constraints imposed on the establishment of correspondences. In fact, these unmatched points may carry useful information about the intrinsic structures of the input data, thereby facilitating transformation learning. Therefore, a desirable way is to involve the entire input point set in the modeling. Third, computational efficiency is a key concern in many scenarios that involve large-scale point sets (e.g., point clouds with tens of thousands of points).

In dealing with these concerns, the matching problem is cast into a mixture model under semisupervised learning. The model brings latent/hidden variables for all matches in the putative set to screen out the outliers, and employs a prior, say a nonparametric constraint to advocate the geometric smoothness on the spatial transformation via manifold regularization on the entire given data [24], [27]. On the one hand, such manifold regularization controls the complexity of the transformation. On the other hand, it can discover the underlying structure of the input data. The maximum *a posteriori* (MAP) estimation technique is a typical option for solving the problem with manifold regularizers, which easily get stuck into bad local points. As a solution, the EM algorithm [28] is adopted to update the variance of position disturbance and distinguish the false correspondences simultaneously, providing a large initial variance. Moreover, a sparsity-induced approximation that is similar to the subset of regressors method [29] is introduced for computational efficiency.

The contributions of this paper include the following four aspects. First, manifold regularization is introduced to the point set registration problem, thereby capturing the intrinsic structure of the given data and helping to learn the transformation. Second, according to the manifold regularization, a novel approach for robust transformation learning, which can learn transformation from point correspondences contaminated by outliers, is proposed. Third, fast implementation is provided for the proposed method using a sparse approximation, which enables the handling of large-scale data, e.g., 3-D point clouds. Fourth, our proposed method is generalized to solve the visual homing problem, which can learn accurate motion flows between image pairs and help improve homing performance significantly.

A preliminary version of this paper appeared in [30]. The primary new contributions include that the model design and the solver are presented in more theoretical details, a more in-depth analysis of the properties and potentials of the proposed method are provided and further generalized to solve a real-word task in the field of mobile robotics, such as visual homing. Extensive experimental comparisons are conducted to verify the advances of our method. The code is released at https://sites.google.com/site/jiayima2013/home to allow comparisons from the community and encourage future work.

C. Organization of This Paper

The remainder of this paper is organized as follows. Section II formalizes the foundational definitions and setup of

the point set registration problem, including correspondence estimation and transformation learning. Section III presents the proposed robust transformation learning algorithm under manifold regularization and its fast implementation, along with some implementation details. Section IV applies the proposed method to solve the visual homing problem. Section V illustrates the use of the proposed method in 2-D shape matching, 3-D point cloud registration, and visual homing on publicly available data sets, with comparisons to other state-of-the-art methods. Section VI summarizes the concluding remarks.

II. Problem Statement

Point set registration aims to determine the correct correspondences and seek the spatial transformation between two given point sets (e.g., a model point set $\{\mathbf{x}_i\}_{i=1}^{M}$ and a target point set $\{\mathbf{y}_j\}_{j=1}^{N}$, where $\mathbf{x}_i$ and $\mathbf{y}_j \in \mathbb{R}^D$ are the point positions (usually $D = 2$ or 3), M and N are, respectively, the numbers of points contained in the two sets) [25]. To solve this problem, an iterative strategy between correspondence construction and transformation learning is considered.

A. Correspondence Construction

A specific shape or object should have a relatively fixed geometric structure at its local parts. Therefore, two-point sets that are generated from two samples of the same shape will generally have similar local geometric structures. By using a shape feature descriptor to incorporate such information (e.g., neighborhood structures), the correspondence can be constructed based on the similarity of corresponding descriptors. For example, two points should be matched only if their descriptors are sufficiently similar. Several well-designed feature descriptors can efficiently establish reliable correspondences between point sets in 2-D and 3-D cases [12], [13], [21], [31].

For the 2-D case, the SC [13] is used to construct descriptor. For two given points (i.e., $\mathbf{x}_i$ and $\mathbf{y}_j$), their corresponding SCs are histograms $\{p_i(k)\}_{k=1}^{H}$ and $\{q_j(k)\}_{k=1}^{H}$, which characterize the distributions of their neighborhood points. Specifically, for a point $\mathbf{x}_i$ from $\{\mathbf{x}_i\}_{i=1}^{M}$, p_i is computed according to the relative spatial positions of the remaining $M-1$ points

$$p_i(k) = \#\{\mathbf{x}_j \neq \mathbf{x}_i : (\mathbf{x}_j - \mathbf{x}_i) \in \text{bin}(k)\}. \tag{1}$$

The bins are uniform in the log-polar space. Consequently, the nearby sample points play more important roles than the points that are farther away. The original SC is not rotation invariant, and a rotation invariant SC could be considered if necessary by using a mass center to compute the positive x-axis for the local coordinate system [22]. After obtaining p_i and q_j, their difference is frequently measured via the χ^2 distance as follows:

$$C_{ij} = C(\mathbf{x}_i, \mathbf{y}_j) = \frac{1}{2}\sum_{k=1}^{H}\frac{[p_i(k) - q_j(k)]^2}{p_i(k) + q_j(k)}. \tag{2}$$

After all the pairwise distances (i.e., $\{C(\mathbf{x}_i, \mathbf{y}_j), i = 1, \ldots, M, \ j = 1, \ldots, N\}$) are obtained, the Hungarian method [32] is applied to determine the correspondences between $\{\mathbf{x}_i\}_{i=1}^{M}$ and $\{\mathbf{y}_j\}_{j=1}^{N}$.

For the 3-D case, the FPFH [21] is considered as the feature descriptor. It is a fast implementation of PFH that captures the underlying surface model properties (e.g., the local geometry, including point coordinates, surface normals, curvatures, and moment invariants) around the k-nearest neighbors of each point. The computational complexity of FPFH is linear with respect to the number of neighbors for each point, and hence beneficial for dealing with large-scale 3-D point clouds. To establish correspondences between two point clouds efficiently, a sample consensus initial alignment method is implemented instead of the original greedy initial alignment in PFH, which tries to preserve the same local geometries of putative correspondences without having to try all possible correspondence pairs.

After using local feature descriptors to establish correspondence, a putative set $S = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{L}$ is obtained with $L \leq \min\{M, N\}$ being the amount of correspondences. In addition, we make an assumption that $\{\mathbf{x}_i\}_{i=1}^{L}$ and $\{\mathbf{y}_j\}_{j=1}^{L}$ in the putative set S correspond to the first L elements of the original point sets $\{\mathbf{x}_i\}_{i=1}^{M}$ and $\{\mathbf{y}_j\}_{j=1}^{N}$, respectively.

B. Transformation Learning

A putative correspondence set, $S = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{L}$, is established from two point sets (i.e., $\{\mathbf{x}_i\}_{i=1}^{M}$ and $\{\mathbf{y}_j\}_{j=1}^{N}$) involving nonrigid deformation. Each input–output pair $(\mathbf{x}_i, \mathbf{y}_i)$ can be considered a random sample drawn from the underlying spatial transformation between the two point sets. This paper aims to learn the transformation $\mathcal{T}$ [e.g., $\mathbf{y}_i = \mathcal{T}(\mathbf{x}_i)$ for any $(\mathbf{x}_i, \mathbf{y}_i)$ in S]. However, for a nonrigid $\mathcal{T}$, its solution will be not unique if no additional constraint is imposed on $\mathcal{T}$. The regularization technique, which usually operates in a reproducing kernel Hilbert space (RKHS) [33] (associated with a particular kernel), can be used to obtain a meaningful solution. In particular, the Tikhonov regularization [34] in the RKHS $\mathcal{H}$ minimizes a regularized risk functional as follows:

$$\mathcal{T}^* = \min_{\mathcal{T}\in\mathcal{H}} \sum_{i=1}^{L} \|\mathbf{y}_i - \mathcal{T}(\mathbf{x}_i)\|^2 + \lambda\|\mathcal{T}\|_{\mathcal{H}}^2 \tag{3}$$

where the first term fitting the input data is the data-fidelity term, or empirical error (risk); the second term is a regularization term, which plays a role of stabilizer and enforces smoothness to the spatial transformation $\mathcal{T}$; $\lambda > 0$ controls the tradeoff; $\|\cdot\|_{\mathcal{H}}$ represents the functional norm of $\mathcal{H}$ (their definitions will be discussed in the Appendix).

The transformation learning here focuses on how to utilize the global geometric constraint on the point sets to obtain a smooth transformation without consideration of the intrinsic geometry involved in a specific shape or object.

III. Method

This section presents the proposed manifold regularization under semisupervised learning for transformation estimation and shows its capability of capturing the underlying intrinsic structure of a point set. Then, a formulation for robust transformation learning from putative correspondences using the

global geometric constraints of manifold regularization is proposed, followed by its fast implementation based on sparsity-induced approximation and some implementation details of the algorithm.

A. Manifold Regularization

The problem of manifold regularization is tied to semisupervised learning. It is a generalization of supervised learning in which regularization is imposed on the labeled and unlabeled data [27]. Manifold learning methods have received considerable research interest in the areas of pattern recognition and machine learning [35]–[37] since the introduction of two pioneering methods, namely, ISOMAP [38] and locally linear embedding method [39]. These methods usually assume that a low-dimensional representation can be used to characterize the underlying intrinsic structure embedded in the high-dimensional data. Based on this assumption, a graph is constructed, and the graph Laplacian matrix is calculated to capture the manifold structure, which is further utilized to conduct various learning tasks, including clustering, dimensionality reduction, and semisupervised learning [40], [41].

Recently, manifold regularization has been widely used in computer vision [42], [43]. Xu *et al.* [44] applied manifold regularization to solve the feature selection problem under a semisupervised learning setting to determine more discriminative features. Zhao *et al.* [45] introduced a compact graph to grasp the geometric structure of a data set and utilized it for manifold ranking in image retrieval. Xiang *et al.* [46] presented a local regression and global alignment approach for graph construction and applied it to image segmentation. A recent work [41] used manifold regularization to manage image classification and visualization and preserve manifold information together with local and global discriminative information. In this paper, manifold regularization is introduced to the point matching problem to utilize the intrinsic geometry of the given point sets supported on a low-dimensional manifold. The proposed method can fully utilize the entire input data, thereby enhancing registration accuracy.

B. Transformation Learning With Manifold Regularization

In the matching problem, the matched points are usually a part of the entire point set (i.e., $L \leq M$) because of the existence of noise, outliers, and occlusions. That is, only L points $\mathbf{x}_1, \ldots, \mathbf{x}_L$ are given labels $\mathbf{y}_1, \ldots, \mathbf{y}_L$ undergoing transformation $\mathcal{T}$, respectively. However, the input data in a point set registration task are typically sampled a specific object that possesses certain "intrinsic geometry." For example, the spatial positions of points comprising a specific shape are not arbitrary and often obey a specific distribution. Thus, the remaining $M - L$ points without labels may contain extra geometric structure information about the input data. Manifold regularization is considered to fully utilize such additional information [27], [47]. It defines an extra regularization term $\|\mathcal{T}\|_{\mathcal{I}}^2$ on the entire input data $\{\mathbf{x}_i\}_{i=1}^M$ to constrain $\mathcal{T}$ in a low-dimensional manifold. Thus, the objective function in (3) becomes

$$\mathcal{T}^* = \min_{\mathcal{T}\in\mathcal{H}} \sum_{i=1}^{L} \|\mathbf{y}_i - \mathcal{T}(\mathbf{x}_i)\|^2 + \lambda_1 \|\mathcal{T}\|_{\mathcal{H}}^2 + \lambda_2 \|\mathcal{T}\|_{\mathcal{I}}^2 \quad (4)$$

where the first regularization term controls the complexity of $\mathcal{T}$, and the second regularization term exploits the intrinsic geometry of the input data. As a manifold is typically located in a subspace, the first regularization term is then necessary. In particular, for those $\mathcal{T} \in \mathcal{H}$ providing the same value on a manifold, the solution that is smoother in the input space is preferred and expected to have better generalization capacity.

The graph Laplacian, which is a discrete analog of the manifold Laplacian, is used to define the manifold regularization term [27]. Suppose the input samples are drawn *i.i.d.* from a manifold, the graph Laplacian then models the manifold according to a weighted neighborhood graph, G. Specifically, G is obtained by constructing the vertex set $V = \{\mathbf{x}_1, \ldots, \mathbf{x}_M\}$ (the matched and unmatched points) with edges $(\mathbf{x}_i, \mathbf{x}_j)$ if and only if $\|\mathbf{x}_i - \mathbf{x}_j\|^2 \leq \epsilon$. The following weight is assigned to edge $(\mathbf{x}_i, \mathbf{x}_j)$:

$$W_{ij} = e^{-\frac{1}{\epsilon}\|\mathbf{x}_i - \mathbf{x}_j\|^2}. \quad (5)$$

According to W, a matrix $\mathbf{A}$ is constructed with each element expressed as follows:

$$A_{ij} = D_{ij} - W_{ij} \quad (6)$$

where $\mathbf{D} = \text{diag}(\sum_{j=1}^{M} W_{ij})_{i=1}^{M}$ (i.e., the diagonal matrix whose ith entry is the sum of the weights of edges leaving $\mathbf{x}_i$). Denote $\mathbf{t} = (\mathcal{T}(\mathbf{x}_1), \ldots, \mathcal{T}(\mathbf{x}_M))^T$. The manifold regularization term is defined as follows:

$$\|\mathcal{T}\|_{\mathcal{I}}^2 = \sum_{i=1}^{M}\sum_{j=1}^{M} W_{ij}(t_i - t_j)^2 = \text{tr}(\mathbf{t}^T\mathbf{A}\mathbf{t}) \quad (7)$$

where $\text{tr}(\cdot)$ indicates the trace of a matrix. Thus, the objective function (4) becomes

$$\mathcal{T}^* = \min_{\mathcal{T}\in\mathcal{H}} \sum_{i=1}^{L} \|\mathbf{y}_i - \mathcal{T}(\mathbf{x}_i)\|^2 + \lambda_1 \|\mathcal{T}\|_{\mathcal{H}}^2 + \lambda_2 \text{tr}(\mathbf{t}^T\mathbf{A}\mathbf{t}). \quad (8)$$

Its solution will be discussed subsequently in Section III-C.

C. Robust Transformation Learning

The transformation can be learned by minimizing the objective function in (4). Nevertheless, the putative set $S = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{L}$ usually involves several unknown false correspondences because it is constructed based on only local neighborhood structures. Therefore, the transformation learning procedure should be resistant to outliers. Next, we introduce an approach for achieving such robust learning based on manifold regularization.

For the inliers, the noise on the components of point positions is assumed to be isotropic Gaussian, i.e., $\mathcal{N}(0, \sigma^2\mathbf{I})$. While for the outliers, the corresponding point could appear anywhere in the output space, leading to a uniform distribution $1/a$, with a denoting the volume of the bounded output space [24]. Then, the ith correspondence is associated with

a latent variable $z_i \in \{0, 1\}$, where $z_i = 1$ and $z_i = 0$ indicate Gaussian and uniform distributions, respectively. Let $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_L)^T$ and $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_L)^T \in \mathbb{R}^{L\times D}$ be the two point sets in the putative set. Thus, the likelihood of matching correctness is a mixture model expressed as follows:

$$p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}) = \prod_{i=1}^{L} \sum_{z_i} p(\mathbf{y}_i, z_i|\mathbf{x}_i, \boldsymbol{\theta}) = \prod_{i=1}^{L} \left(\frac{\gamma}{(2\pi\sigma^2)^{D/2}} e^{-\frac{\|\mathbf{y}_i - \mathcal{T}(\mathbf{x}_i)\|^2}{2\sigma^2}} + \frac{1-\gamma}{a} \right) \quad (9)$$

where $\boldsymbol{\theta} = \{\mathcal{T}, \sigma^2, \gamma\}$ contains the variables to be solved, in which γ specifies the marginal distribution of hidden variable z_i (i.e., $\forall z_i$, $p(z_i = 1) = \gamma$). The nonrigid transformation $\mathcal{T}$ is assumed to lie within the RKHS and simultaneously captures the intrinsic geometry of the input data. These properties can be incorporated into a prior on $\mathcal{T}$ as: $p(\mathcal{T}) \propto e^{-(1/2)(\lambda_1\|\mathcal{T}\|^2_{\mathcal{H}} + \lambda_2\|\mathcal{T}\|^2_{\mathcal{I}})}$. The MAP solution of $\boldsymbol{\theta}$ is estimated via the Bayes rule as follows:

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Y}) = \arg\max_{\boldsymbol{\theta}} p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}) p(\mathcal{T}). \quad (10)$$

The EM algorithm, which is a common technique for addressing the appearance of hidden variables, is considered to optimize the objective function. This paper follows standard notations [48] and omits several terms that are independent of $\boldsymbol{\theta}$. Considering the posterior function in 10, its complete-data log posterior is expressed as follows:

$$\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) = -\frac{1}{2\sigma^2} \sum_{i=1}^{L} p_i \|\mathbf{y}_i - \mathcal{T}(\mathbf{x}_i)\|^2 - \frac{\text{D}L_p}{2} \ln\sigma^2 + L_p \ln\gamma + (L - L_p)\ln(1-\gamma) - \frac{\lambda_1}{2}\|\mathcal{T}\|^2_{\mathcal{H}} - \frac{\lambda_2}{2}\|\mathcal{T}\|^2_{\mathcal{I}} \quad (11)$$

where $p_i = P(z_i = 1|\mathbf{x}_i, \mathbf{y}_i, \boldsymbol{\theta}^{\text{old}})$, $L_p = \sum_{i=1}^{L} p_i$. The EM approach iterates between an E-step and an M-step.

E-Step: This step focuses on computing the posterior distribution of hidden variables (i.e., p_i) according to the current newest parameter value (i.e., $\boldsymbol{\theta}^{\text{old}}$). Let $\mathbf{P} = \text{diag}(p_1, \ldots, p_L)$ denote a diagonal matrix, and it is obtained by using the Bayes rule as follows:

$$p_i = \frac{\gamma\, e^{-\frac{\|\mathbf{y}_i - \mathcal{T}(\mathbf{x}_i)\|^2}{2\sigma^2}}}{\gamma\, e^{-\frac{\|\mathbf{y}_i - \mathbf{f}(\mathbf{x}_i)\|^2}{2\sigma^2}} + (1-\gamma)\frac{(2\pi\sigma^2)^{D/2}}{a}}. \quad (12)$$

The value of p_i ranging from 0 to 1 characterizes the degree of consistency that correspondence $(\mathbf{x}_i, \mathbf{y}_i)$ satisfies the current newest transformation $\mathcal{T}$.

M-Step: This step focuses on determining the updated parameter as $\boldsymbol{\theta}^{\text{new}} = \arg\max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$. Let $\mathcal{T}(\mathbf{X}) = (\mathcal{T}(\mathbf{x}_1), \ldots, \mathcal{T}(\mathbf{x}_L))^T$. As $\mathbf{P}$ is a diagonal matrix, we take the derivatives of $\mathcal{Q}(\boldsymbol{\theta})$ with respect to σ^2 and γ and set them to zero, the following closed-form solutions are obtained:

$$\sigma^2 = \frac{\text{tr}((\mathbf{Y} - \mathcal{T}(\mathbf{X}))^T \mathbf{P}(\mathbf{Y} - \mathcal{T}(\mathbf{X})))}{\text{D}L_p} \quad (13)$$

$$\gamma = \text{tr}(\mathbf{P})/L. \quad (14)$$

Then, the terms of $\mathcal{Q}(\boldsymbol{\theta})$ related to $\mathcal{T}$ are considered, and the following manifold regularized risk functional is obtained [49]:

$$\mathcal{E}(\mathcal{T}) = \frac{1}{2\sigma^2} \sum_{i=1}^{L} p_i \|\mathbf{y}_i - \mathcal{T}(\mathbf{x}_i)\|^2 + \frac{\lambda_1}{2}\|\mathcal{T}\|^2_{\mathcal{H}} + \frac{\lambda_2}{2}\|\mathcal{T}\|^2_{\mathcal{I}}. \quad (15)$$

The transformation $\mathcal{T}$ is modeled in the RKHS $\mathcal{H}$, which is defined uniquely by a matrix-valued kernel $\Gamma : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}^{D\times D}$. For the point set registration problem, a diagonal decomposable kernel is often sufficiently accurate to capture the spatial transformation [24], for example, $\Gamma(\mathbf{x}, \mathbf{x}') = \kappa(\mathbf{x}, \mathbf{x}') \cdot \mathbf{I}$ with $\kappa(\mathbf{x}, \mathbf{x}') = e^{-\beta\|\mathbf{x}-\mathbf{x}'\|^2}$ being a scalar Gaussian kernel, where β determines the range of interactions between different points. Thus, the representer theorem is derived as follows [27], [47] (the proof is provided in the Appendix).

Theorem 1: The optimal solution of the manifold regularized risk functional (15) is expressed as follows:

$$\mathcal{T}^*(\mathbf{x}) = \sum_{i=1}^{M} \Gamma(\mathbf{x}, \mathbf{x}_i)\mathbf{c}_i \quad (16)$$

with the coefficient set $\{\mathbf{c}_i\}_{i=1}^{M}$ determined by the following linear system:

$$(\mathbf{J}^T\mathbf{P}\mathbf{J}\Gamma + \lambda_1\sigma^2\mathbf{I} + \lambda_2\sigma^2\mathbf{A}\Gamma)\mathbf{C} = \mathbf{J}^T\mathbf{P}\mathbf{Y} \quad (17)$$

where $\Gamma \in \mathbb{R}^{M\times M}$ is the Gram matrix and $\Gamma_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j)$, $\mathbf{J} = (\mathbf{I}_{L\times L}, \mathbf{0}_{L\times(M-L)})$ with $\mathbf{I}$ being an identity matrix and $\mathbf{0}$ being a matrix of all zeros, $\mathbf{C} = (\mathbf{c}_1, \ldots, \mathbf{c}_M)^T \in \mathbb{R}^{M\times D}$ is the matrix of coefficients.

Convergence Analysis: Note that the objective function (10) is nonconvex in nature. Although global optimality is difficult to be warranted by existing optimization techniques, a stable local optimum, which is usually sufficient for solving real-world tasks, can be obtained. To this end, the variance σ^2 is initialized with a relatively large value, e.g., obtained by using (13) with $\mathbf{P} = \mathbf{I}$ and $\mathcal{T}(\mathbf{X}) = \mathbf{X}$. The reason for doing so is that for a large value of σ^2, the objective function should be convex in a large range, thereby filtering out many noisy local optimum. The EM algorithm is employed to approach the desired minimum gradually. The objective function changes smoothly as σ^2 decreases; thus, using the minimum from the previous iteration as the initialization for the next round is beneficial to achieving a new better optimum. In other words, as the procedure iterates, a satisfying local optimum can be finally reached. This concept is associated with deterministic annealing [7], which adopts the solution of a relatively easy problem to provide the initializations recursively to increasingly difficult problems.

D. Fast Implementation

In our proposed method, we need to solve transformation $\mathcal{T}$ in (17), which is the most time-consuming procedure. Its time complexity is cubic $O(M^3)$, which is problematic when dealing with large-scale problems. Even when the proposed method is implementable, an alternative which produces suboptimal solutions but is more efficient may be preferred. Next, we give such an approximate and fast solution by using the idea similar to a subset of regressors [29].

Instead of seeking the optimal solution in $\mathcal{H}_M$, a sparsity-induced approximation, which is defined as $\mathcal{H}_K = \{\sum_{i=1}^{K} \Gamma(\cdot, \widetilde{\mathbf{x}}_i)\mathbf{c}_i\}$, is adopted to optimize the problem in a space with less basis functions. In this paper, $K \ll M$ and the point set $\{\widetilde{\mathbf{x}}_i : i \in \mathbb{N}_K\}$ are composed of an arbitrary subset of $\{\mathbf{x}_i : i \in \mathbb{N}_M\}$. This is inspired by Rifkin *et al*. [50] and Ma *et al*. [51] who found that such a simple random selection performs as well as those sophisticated and time-consuming methods. The manifold regularized risk functional over all samples is then minimized. According to sparse approximation, the solution forms the following shape:

$$\mathcal{T}(\mathbf{x}) = \sum_{i=1}^{K} \Gamma(\mathbf{x}, \widetilde{\mathbf{x}}_i)\mathbf{c}_i \tag{18}$$

with the coefficient $\{\mathbf{c}_i\}_{i=1}^{K}$ determined by the following linear system:

$$(\mathbf{U}^T\mathbf{P}\mathbf{U} + \lambda_1\sigma^2\Gamma_s + \lambda_2\sigma^2\mathbf{V}^T\mathbf{A}\mathbf{V})\mathbf{C} = \mathbf{U}^T\mathbf{P}\mathbf{Y} \tag{19}$$

where $\Gamma_s \in \mathbb{R}^{K\times K}$ and $\Gamma_{s,ij} = \kappa(\widetilde{\mathbf{x}}_i, \widetilde{\mathbf{x}}_j)$, $\mathbf{U} \in \mathbb{R}^{L\times K}$ and $\mathbf{U}_{ij} = \kappa(\mathbf{x}_i, \widetilde{\mathbf{x}}_j)$, $\mathbf{V} \in \mathbb{R}^{M\times K}$ and $\mathbf{V}_{ij} = \kappa(\mathbf{x}_i, \widetilde{\mathbf{x}}_j)$. Note that $\mathbf{U}$ equals to the first L rows of $\mathbf{V}$. The details of the derivation of (19), which are similar to that of Theorem 1, are omitted.

Unlike the optimal solution of taking the form of a linear combination of the basis functions $\{\Gamma(\cdot, \mathbf{x}_i) : i \in \mathbb{N}_M\}$, by the representer theorem, the solution based on our sparse approximation only involves K basis functions. In general, this procedure significantly speeds up the processing with a trivial loss in accuracy. Compared with the original algorithm, the fast implementation solves the linear system in (19) rather than that in (17).

Computational Complexity: For the linear system (17), the size of the coefficient matrix $\mathbf{J}^T\mathbf{P}\mathbf{J}\Gamma + \lambda_1\sigma^2\mathbf{I} + \lambda_2\sigma^2\mathbf{A}\Gamma$ is $M \times M$, and thus the time complexity for solving $\mathcal{T}$ is $O(M^3)$. In contrast, the size of the coefficient matrix $\mathbf{U}^T\mathbf{P}\mathbf{U} + \lambda_1\sigma^2\Gamma_s + \lambda_2\sigma^2\mathbf{V}^T\mathbf{A}\mathbf{V}$ in the linear system (19) is only $K \times K$, and thus, the time complexity for solving the transformation $\mathcal{T}$ is reduced to $O(K^3)$. Nevertheless, the time complexity of computing the coefficient matrix, $\mathbf{U}^T\mathbf{P}\mathbf{U} + \lambda_1\sigma^2\Gamma_s + \lambda_2\sigma^2\mathbf{V}^T\mathbf{A}\mathbf{V}$, is $O(KM^2)$ due to the multiplication operation on the $M \times M$ graph Laplacian matrix $\mathbf{A}$. Given that K is a constant and independent of M and $K \ll M$, the total time complexity of solving transformation $\mathcal{T}$ in fast implementation can be written as $O(M^2)$. The space complexity of the proposed method scales as $O(M^2)$ because of the memory requirements for storing the graph Laplacian matrix $\mathbf{A}$.

E. Summarization and Implementation Details

To produce reliable registration results, we iteratively construct the correspondences and learn the transformation. The number of iterations is fixed in this paper (i.e., usually 10 but more when the noise level is high or when the original point sets contain a large percentage of outliers). The proposed robust point matching approach is based on manifold regularization and is thus called robust point matching using manifold regularization (MR-RPM). The MR-MPM method is summarized in Algorithm 1.

Algorithm 1 MR-RPM Algorithm

Input: Model point set $\{\mathbf{x}_i\}_{i=1}^{M}$, target point set $\{\mathbf{y}_j\}_{j=1}^{N}$, parameters ϵ, β, λ_1, λ_2, K

Output: Aligned model point set $\{\hat{\mathbf{x}}_i\}_{i=1}^{M}$

1 Construct descriptors for target point set $\{\mathbf{y}_j\}_{j=1}^{N}$;
2 Calculate the volume of output space and assign it to a;
3 **repeat**
4 Construct descriptors for model point set $\{\mathbf{x}_i\}_{i=1}^{M}$;
5 Construct $S = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{L}$ using descriptors;
6 Compute matrix $\mathbf{A}$ using Eqs. (5) and (6);
7 Compute Γ based on the definition of Γ;
8 Initialize $\mathbf{P} = \mathbf{I}$, γ, $\mathcal{T}(\mathbf{x}_i) = \mathbf{x}_i$, and σ^2 by Eq. (13);
9 **repeat**
10 *E-step:*
11 Compute $\mathbf{P} = \mathrm{diag}(p_1, \ldots, p_L)$ using Eq. (12);
12 *M-step*:
13 Compute σ^2 and γ using Eqs. (13) and (14);
14 Compute $\mathbf{C}$ by solving Eq. (19);
15 **until** $\mathcal{Q}$ *converges*;
16 Transform model set $\{\mathbf{x}_i\}_{i=1}^{M} \leftarrow \{\mathcal{T}(\mathbf{x}_i)\}_{i=1}^{M}$;
17 **until** *achieve the maximum number of iterations*;
18 Output $\{\hat{\mathbf{x}}_i\}_{i=1}^{M}$ as $\{\mathcal{T}(\mathbf{x}_i)\}_{i=1}^{M}$ in the last iteration.

The performance of point set registration is influenced by the coordinate system that expresses the data. To relieve such influence, we normalize the coordinates of the point sets. In particular, we adopt a linear scaling on the coordinates, so that the means of coordinates in the two point sets are both 0 and the variances are both 1. Rather than directly solving transformation $\mathcal{T}$, a displacement function $\mathbf{f}$: $\mathcal{T}(\mathbf{x}) = \mathbf{x} + \mathbf{f}(\mathbf{x})$, which is achieved straightforward by replacing the output with $\mathbf{y} - \mathbf{x}$, is solved. Compared with the original position mapping, the use of the motion field achieves high robustness [16], [24].

Parameter Settings: The MR-RPM algorithm has four main parameters, namely, ϵ, β, λ_1, and λ_2. Parameter ϵ is used to establish the graph Laplacian and compute the weight of its edges. Parameter β characterizes the Gaussian window of the kernel function and determines the range of interactions between different points. The tradeoff between fitting the data and ensuring stability on the transformation is controlled by the two other parameters. In particular, λ_1 regularizes with respect to the entire input space, and λ_2 constrain the transformation to exploit the intrinsic geometry. Generally, the proposed method is robust to parameter changes. The following settings are used throughout this paper: $\epsilon = 0.05$, $\beta = 0.1$, $\lambda_1 = 3$, $\lambda_2 = 0.1$, $K = 15$, which are selected by the exhaustive grid search on one data set and kept unchanged in all the experiments. In addition, as described in Line 8 in Algorithm 1, we should make an initial assumption on the inlier ratio γ, which is fixed as 0.9. The constant a for outlier distribution is assigned to the volume of output space after normalization of the point coordinates.

IV. Application to Visual Homing

In this section, we apply our method to the visual homing problem. Specifically, the MR-RPM is used for robust feature

matching and dense motion flow learning between two images. The focus-of-contraction (FOC) and focus-of-expansion (FOE) are derived accordingly to determine the homing directions.

A. Background

Visual homing, which aims to navigate a robot to a goal or home position from arbitrary starting positions solely based on visual information, has received increasing attention and plays an important role in many applications in the mobile robotics community [3]. The visual homing problem is usually solved based on sparse feature matching. Specifically, it initially matches local features in the two images and subsequently transforms the correspondences into motion flows, which are finally used to determine the homing vector [3], [52], [53].

In early homing methods, the positions and correspondences of landmarks are assumed to be known in advance. Such methods usually demand special settings to achieve a reliable performance, such as placing artificial landmarks in the work environment. This restriction limits its applications in practical visual homing problems. To address this issue, recent correspondence methods use feature points that are automatically extracted from images as landmarks, and the correspondences are built using the descriptors associated with them [3], [52]. For example, in [54], a biologically inspired method, such as average landmark vector, was combined with the feature points detected in panoramic images for visual homing and showed promising results. However, the feature correspondences in these methods are usually established based on only local descriptor information, and hence some unknown false matches will be introduced which degrade the homing performance. The robustness of visual homing methods has been verified to be dominated by the presence and amount of false correspondences [55]. Several heuristic methods are usually adopted to remove false correspondences to remedy the degradation caused by mismatches. In [52], features are assumed to be distributed approximately uniformly, and mismatch removal is not performed. Mismatch removal is explicitly conducted for visual homing using an effective random sample consensus (RANSAC)-like method in [3], but it relies on a parametric model. Usually, the spatial transformation of a panoramic pair cannot be modeled exactly by a parametric model [56]. The current study applies the proposed MR-RPM to establish accurate feature correspondences and learn accurate motion flows to enhance homing performance.

B. Motion Flow Learning

The dense motion flow between an image pair of the same or similar scenes is learned based on sparse feature matching. To this end, a set of putative matches $S = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{L}$ is constructed by considering all possible matches between two feature sets (i.e., $\{\mathbf{x}_i\}_{i=1}^{M}$ and $\{\mathbf{y}_j\}_{j=1}^{N}$) and filtering out matches whose feature descriptors are sufficiently different, which can be fulfiled by several existing well-designed local image feature descriptors (e.g., scale invariant feature transform (SIFT) [57]). The putative match $(\mathbf{x}_i, \mathbf{y}_i)$ can be further converted to a motion flow sample by a transformation $(\mathbf{x}_i, \mathbf{y}_i) \rightarrow (\mathbf{u}_i, \mathbf{v}_i)$, where $\mathbf{u}_i = \mathbf{x}_i$ and $\mathbf{v}_i = \mathbf{y}_i - \mathbf{x}_i$ denote the position and displacement, respectively. Therefore, this paper aims to learn a dense motion flow $\mathbf{f} : \mathbf{v}_i = \mathbf{f}(\mathbf{u}_i)$ from a set of sparse motion flow samples contaminated by several unknown outliers. Clearly, MR-RPM can be used to achieve this goal.

In the visual homing problem, the panoramic image usually achieves a 360° field-of-view horizontally, which is usually called "360 cylindrical panorama." The image plane of this type of image can be considered a cylinder unrolled along with a certain vertical cutting line. Therefore, calculating the displacements between feature matches on the image plane by directly using the Euclidean distance is inappropriate, because the displacement will depend on the cutting line in this case. For example, two nearby matched features on the cylinder will have a large displacement on the image plane if they are located on the two sides of the cutting line. To address this issue, the motion flow vector is defined according to the cylinder distance as follows:

$$\mathbf{u}_i = \mathbf{x}_i \tag{20}$$

$$\mathbf{v}_i = \left(\mathbf{y}_i^h - \mathbf{x}_i^h + \alpha \mathbf{x}_{\max}^h, \mathbf{y}_i^v - \mathbf{x}_i^v\right) \tag{21}$$

where $\mathbf{x}^h$ and $\mathbf{x}^v$ denote the horizontal and vertical coordinates of a feature point $\mathbf{x}$, respectively; $\mathbf{x}_{\max}^h$ is the horizontal width of the image plane; and parameter $\alpha \in \{0, \pm 1\}$ is used to wrap the horizontal displacement to $[-\mathbf{x}_{\max}^h/2, \mathbf{x}_{\max}^h/2]$. After the motion flow sample set $S = \{(\mathbf{u}_i, \mathbf{v}_i)\}_{i=1}^{L}$ is obtained, the motion flow $\mathbf{f}$ can be learned according to Lines 5–15 in Algorithm 1 under manifold regularization.

C. Visual Homing Using Singularities of Motion Flow

Previous work shows that the motion flow of a panoramic image pair has two singularities [58] corresponding to the FOC and FOE. These two singularities are separated by a half horizontal width of the panoramic image. The FOC and FOE have been used in many applications, including 3-D environment reconstruction and estimation of time-to-contact in visual navigation. Specifically, in the visual homing literature, the FOC and FOE have been used to determine the homing direction [52], [53]. A heuristic strategy has been proposed by detecting whether the SIFT features have grown or shrunk with respect to their sizes in the reference home image to localize the two singularities [52].

In our previous work [59], we have introduced a method that uses the dense motion flow to determine the FOC and FOE. Here, we briefly review this method. Generally, the FOC and FOE should lie on the horizontal line $\mathbf{u}^v = \mathbf{u}_{\max}^v/2$ and are separated by $\mathbf{u}_{\max}^h/2$.[1] Therefore, no significant differences in the estimation of these two singularities are observed. The subsequent paragraphs will only focus on the estimation of FOC because the generalization of FOE is straightforward.

After obtaining the motion flow $\mathbf{f}(\mathbf{u})$, deriving the analytical solution of its singularities is impossible or difficult. Instead, several numerical methods can be adopted to derive an approximate solution. Formally, given that FOC lies on the horizontal

[1]The definitions of $\mathbf{u}^h$, $\mathbf{u}^v$, and $\mathbf{u}_{\max}^h$ are the same as $\mathbf{x}^h$, $\mathbf{x}^v$, and $\mathbf{x}_{\max}^h$, respectively, and $\mathbf{u}_{\max}^v$ is the vertical width of the image plane.

line $\mathbf{u}^v = \mathbf{u}^v_{\max}/2$, the following 1-D function can be defined:

$$g(\mathbf{u}^h) \triangleq \mathbf{f}\left(\left[\mathbf{u}^h, \mathbf{u}^v_{\max}/2\right]\right). \tag{22}$$

Clearly, $g(\theta)$ is continuous and differentiable, and the singularities correspond to the points whose left and right local neighborhoods have different signs. The formal definition of FOC is as follows.

Definition (FOC): FOC $\mathbf{u}^h_{\text{FOC}}$ is the point that satisfies 1) $g(\mathbf{u}^h_{\text{FOC}}) = 0$ and 2) $\exists\ \epsilon > 0$ that satisfy $g(\mathbf{u}^h) > 0$ for any $\mathbf{u}^h$ in the left ϵ-neighborhood of $\mathbf{u}^h_{\text{FOC}}$ and $g(\mathbf{u}^h) < 0$ for any $\mathbf{u}^h$ in the right ϵ-neighborhood of $\mathbf{u}^h_{\text{FOC}}$.

A coarse-to-fine grid search strategy is used to find the optimal solution of FOC, which can achieve arbitrary precision. In the visual homing literature, all panoramic images usually have identical compass orientations by preprocessing. By converting the coordinate to the angle, the homing direction can be obtained as follows:

$$\theta_{\text{homing}} = \theta_{\text{FOC}} = \frac{2\pi \cdot \mathbf{u}^h_{\text{FOC}}}{\mathbf{u}^h_{\max}}. \tag{23}$$

With this homing direction, the visual homing task can be accomplished, and a robot can be navigated back to its reference home position.

V. Experimental Results

Experiments on 2-D shape contours and 3-D point clouds are conducted to demonstrate the effectiveness of the proposed MR-RPM. The experimental environment is a laptop which has a 3.0-GHz Intel Core CPU and an 8-GB memory, and the algorithm is implemented with MATLAB code.

A. Results on 2-D Shape Contour

For the registration of the 2-D shape contour, two publicly available shape models (i.e., a *fish* pattern and a *Chinese character* pattern) are used to test the performance of different methods. Following [7] and [22], these two patterns are artificially added with different types of degenerations, such as deformation, noise, outlier, rotation, and occlusion. There are five or six levels of degeneration (each level involves 100 samples) for every degeneration type. Note that the degeneration of outlier is to some extent analogous to the degeneration of occlusion, because in both cases there are some elements only included in one of the two point sets. However, for real-world tasks, the degeneration of occlusion is more common given that the uncommon points originate from an object contour. By contrast, for the degeneration of outlier, the uncommon points are generated randomly on the two patterns in our testing data sets. Moreover, the degeneration of rotation can be easily addressed in our method by adopting a feature descriptor that is rotation invariant. Thus, the proposed method is tested on only three types of degenerations, namely, deformation, noise, and occlusion.

The goal of point set registration is to align two point sets together, such as the model point set marked by "+" and the target point set marked by "○." Fig. 1 schematically illustrates the registration evolution in the *fish* case. Different columns correspond to different stages of registration, while each row is polluted by a different type of degradation. From the results, we can observe that the MR-RPM registration is accurate and robust, and usually converges within 10 iterations.

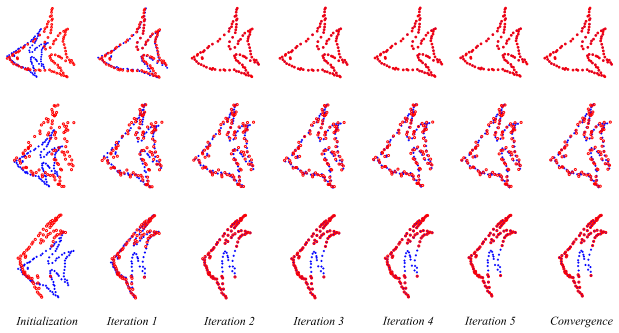


Fig. 1. Schematic of the iterative registration process of our MR-RPM for nonrigid registration. We aim to align the model sets ("+") onto the target sets ("○"). Input data suffer from deformation, noise, and occlusion (from top to bottom).

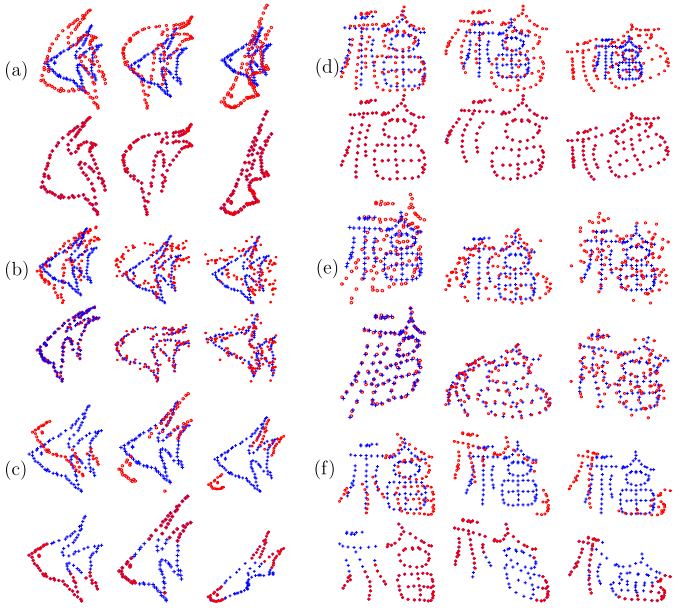


Fig. 2. Qualitative illustration of the MR-RPM on the *fish* (a)–(c) and *Chinese character* (d)–(f) patterns. For each group of results, the top plots are the input data, whereas the bottom plots are the alignment results, with the level of degradation increasing from left to right. Input data suffer from deformation, noise, and occlusion (from top to bottom).

More qualitative results of MR-RPM are offered in Fig. 2. For each group of results, the top figures depict the model ("+") and the target points ("○"), while the bottom ones present the registration results. The results reflect that MR-RPM is capable to handle all of the different degradations. The registration performance declines gradually as the level of degradation goes up. However, even in the case of high degradation, particularly for deformation and occlusion, the results produced by the proposed method are still remarkable. The average elapsed time of our MR-RPM on these two patterns with approximately 100 points is approximately 0.5 s.

The results of seven methods (i.e., SC [13], robust point matching using thin plate spline (TPS-RPM) [7], robust point matching by preserving local neighborhood structures (RPM-LNS) [22], GMM-based registra-

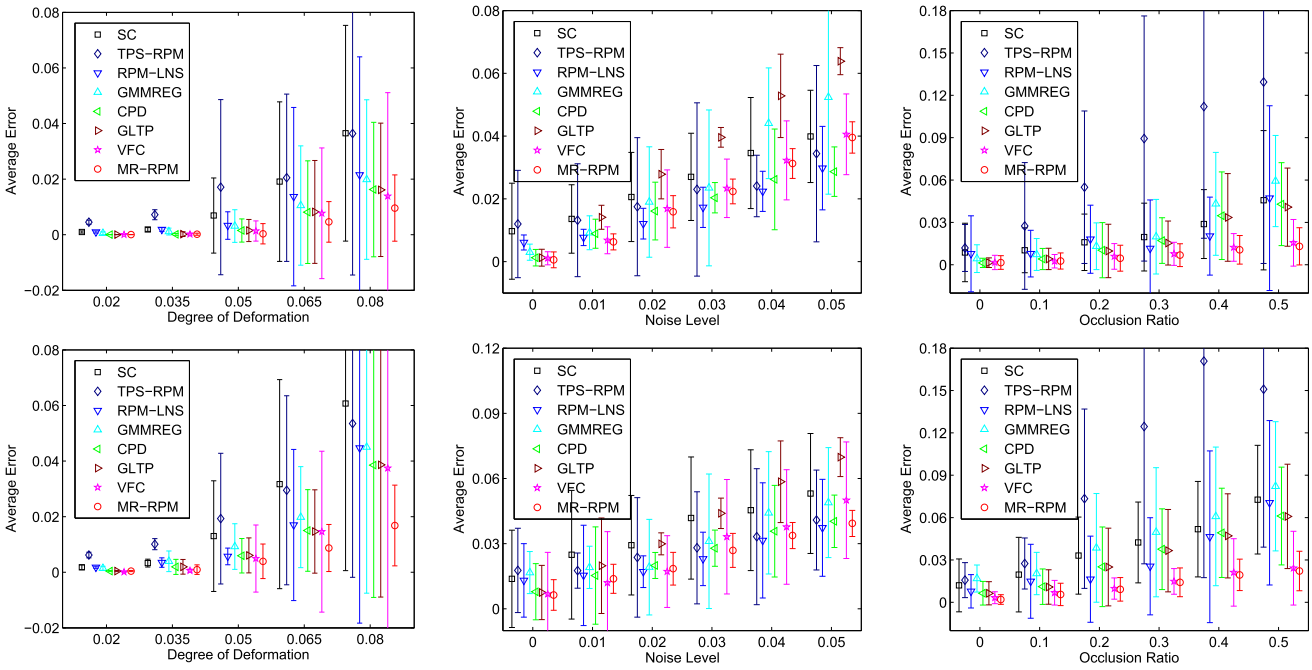


Fig. 3. Comparison of MR-RPM with SC, TPS-RPM, RPM-LNS, GMMREG, CPD, GLTP, and VFC on the *fish* (top) and *Chinese character* (bottom). Error bars: registration error means and standard deviations over 100 trials.

TABLE I

STATISTICS OF THE AVERAGE ERRORS AND RUNTIME (UNIT: SECOND) OF MR-RPM AND ITS SPARSE APPROXIMATION (I.E., MR-RPM-S) ON THE TESTING DATA SETS WITH DEGENERATION OF OCCLUSION. BOLD: BETTER PERFORMANCE

Datsets & Settings		*fish*-Occlusion Ratio						*Chinese character*-Occlusion Ratio					
		0	0.1	0.2	0.3	0.4	0.5	0	0.1	0.2	0.3	0.4	0.5
Ave. Error	MR-RPM	0.0016	0.0027	0.0047	0.0068	**0.0106**	0.0131	**0.0020**	**0.0056**	**0.0092**	0.0142	**0.0195**	**0.0222**
	MR-RPM-S	**0.0015**	**0.0024**	**0.0044**	**0.0061**	0.0109	**0.0123**	0.0023	0.0058	0.0097	**0.0140**	0.0209	0.0225
Ave. Time	MR-RPM	0.0311	0.0159	0.0139	0.0118	0.0059	0.0058	0.0028	0.0030	0.0039	0.0039	0.0045	0.0044
	MR-RPM-S	**0.0018**	**0.0016**	**0.0016**	**0.0016**	**0.0014**	**0.0013**	**0.0008**	**0.0009**	**0.0012**	**0.0012**	**0.0013**	**0.0013**

tion (GMMREG) [15], CPD [16], GLTP [19], and VFC [24]) are reported to provide a quantitative comparison with state-of-the-art methods, as shown in Fig. 3. We implement the seven comparison methods based on publicly available codes and use their default parameter settings because these methods were evaluated on the same data set as that used in this paper. To characterize the registration error, we compute the average Euclidean distance between the warped model points and their ground truth corresponding target points in each pattern pair. Subsequently, to make a quantitative comparison of different methods, we calculate the mean and standard deviation of the registration error on all the 100 samples for each level of degradation in each type of degradation.

The results in Fig. 3 show that SC, GMMRGE, and GLTP are sensitive to noise, whereas TPS-RPM is poorly degraded when the occlusion ratio is high. The alignment performance of RPM-LNS and CPD is relatively good, which declines smoothly when the level of degradation goes up. By contrast, in most cases, the VFC and MR-RPM can achieve the best performance, except in high noise level. MR-RPM almost consistently outperforms VFC for different degradation types and degradation levels on all of the data sets, especially when the data are degraded by a large degree of deformation. Note that the key difference of our MR-RPM and the compared iterative algorithms, especially VFC, is that our MR-RPM adopts an extra manifold regularization term to regularize the transformation. Consistently better results illustrate that manifold regularization does play an important role in improving transformation learning.

We next validate the effectiveness of our sparse approximation and test the accuracy and efficiency of our method with and without sparse approximation on the two shape patterns with occlusion degeneration. The results are reported in Table I. We see that the average registration errors of MR-RPM and MR-RPM-S are quite close and that MR-RPM-S even improves the registration performance in some cases. This result may be because the sparse approximation solution with less basis functions in (18) is "simpler" than the original optimal solution in (16) and it is easier to solve in the context of nonconvex optimization. For the average runtime, we only report in Table I, the time cost of transformation estimation (i.e., without the time cost of correspondence construction), which can directly highlight the advantage of sparse approximation. The results show that sparse approximation can greatly improve registration efficiency, and this advantage is further magnified when the size of the point set goes up, especially in case of 3-D point cloud registration.

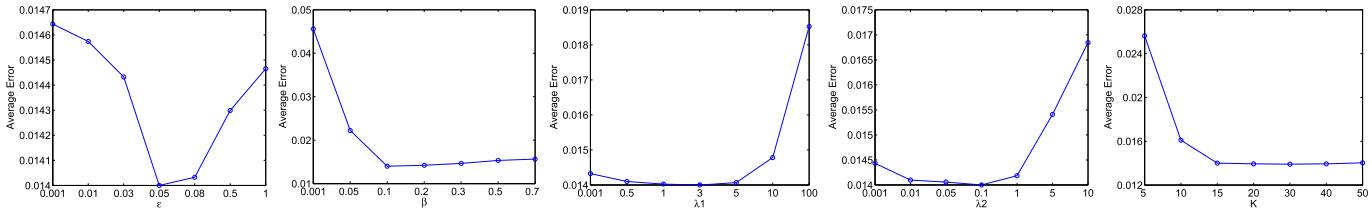


Fig. 4. Illustration of the influence of parameter settings. The testing data are the 100 *fish* pattern pairs with occlusion ratio at 0.3, and we compute the average registration error to characterize the registration performance.

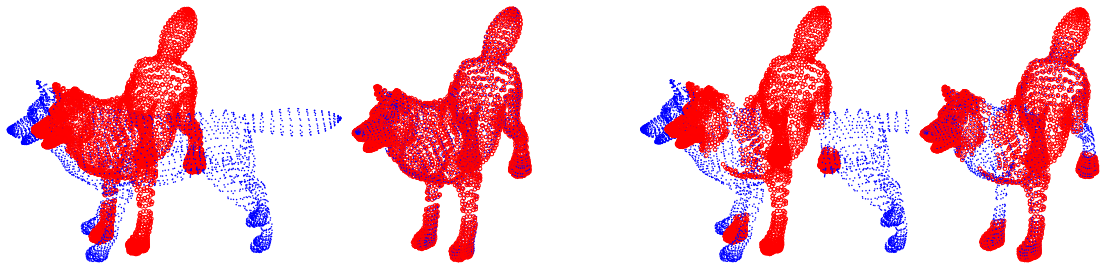


Fig. 5. Qualitative illustration of our MR-RPM on 3-D point clouds (*wolf*) undergoing nonrigid deformation (left two plots) and occlusion (right two plots).

The influence of the parameter settings is also investigated. Specifically, we change the value of one parameter and fix the values of the other four parameters and test the average registration errors on the 100 *fish* pattern pairs, with the occlusion ratio at 0.3 (refer to the bottom right figure in Fig. 3). The results are reported below in Fig. 4. We see that the best performance can be achieved at $\epsilon = 0.05$, $\beta = 0.1$, $\lambda_1 = 3$, and $\lambda_2 = 0.1$. For parameter K, i.e., the number of bases used for sparse approximation, the performance generally improves as K increases. However, the performance is already sufficiently good at $K = 15$. For the sake of efficiency, we set the default value to $K = 15$.

B. Results on 3-D Point Cloud

To test the registration performance of the proposed MR-RPM on 3-D point clouds, we use a public *wolf* shape pattern with approximately 5000 points in different poses for the evaluation [60]. In Fig. 5, the first two columns show the results of the test on nonrigid deformation, and the other two columns show the results of the test on occlusion. Under both conditions, the proposed method consistently produces superior results. An average run time of approximately 47 s is required for this data set.

A quantitative comparison of two typical state-of-the-art methods, namely, CPD and VFC, is conducted. The average alignment errors on the nonrigid deformation and occlusion cases shown in Fig. 5 are 0.82 and 0.72 for CPD, 1.15 and 1.01 for VFC, and 0.78 and 0.53 for MR-RPM, respectively. Clearly, the proposed method exhibits the best results, which indicates that MR-RPM is effective for registration of both the 2-D and 3-D data.

C. Results on Visual Homing

MR-RPM is evaluated using a widely used panoramic image database[2] in the visual homing literature [3], [53]. The panoramic image database contains a collection of omnidirectional and unwrapped images in an indoor environment, together with ground truth for positions where the images were collected. The database includes several scenes, and the sizes of the collected images are 561×81, 583×81, or 295×41. The actual intervals between two nearest positions for image collection are 30 cm. Given that the image resolution is relatively low, the default parameter of SIFT is modified to generate more features. Specifically, the number of layers in each octave is increased from default 3 to 6.

Three types of methods for quantitative comparison, namely, homing in scale space (HiSS) [52], visual servoing-based methods [3], and motion flow interpolation by smoothness prior (MFI-SP) [53], are used to validate the effectiveness of MR-RPM in visual homing. Note that in [3], four variants of homing methods are introduced: 1) bearing-only visual servoing; 2) scale-only visual servoing; 3) scale and bearing visual servoing, and 4) simplified scale-based visual servoing (SSVS). For these four variants, only the results of SSVS are reported because of its superior performance and efficiency compared with the other three methods. Moreover, SSVS is the first choice of the original authors according to their comprehensive evaluation.[3] We implement all the comparison methods and tune all the parameters according to the original papers to find the optimal settings. As in [3] and [52], the total average angular error (TAAE), minimal error (Min), maximal error (Max), and standard variation of error (StdVar) are used to evaluate the homing performance. For all the metrics, small values indicate good results.

1) Sparse Feature Matching on Panoramic Images: The method for sparse feature matching on panoramic images is tested. The ground truth is established by the manual

[2]http://www.ti.uni-bielefeld.de/html/research/avardy/index.html

[3]Among the different feature extraction methods used in this paper, the performances of HiSS [52] and SSVS [3] are not the same as those reported in previous studies. The reimplemented SSVS method in this paper does not contain the mismatch removal introduced in [3].

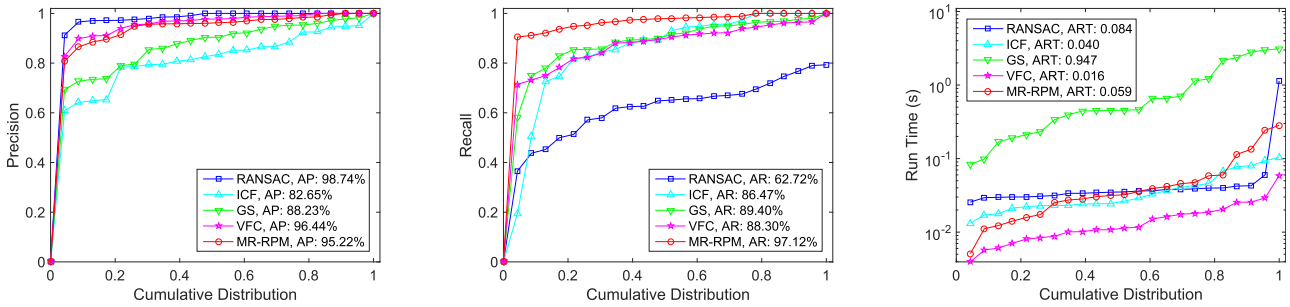
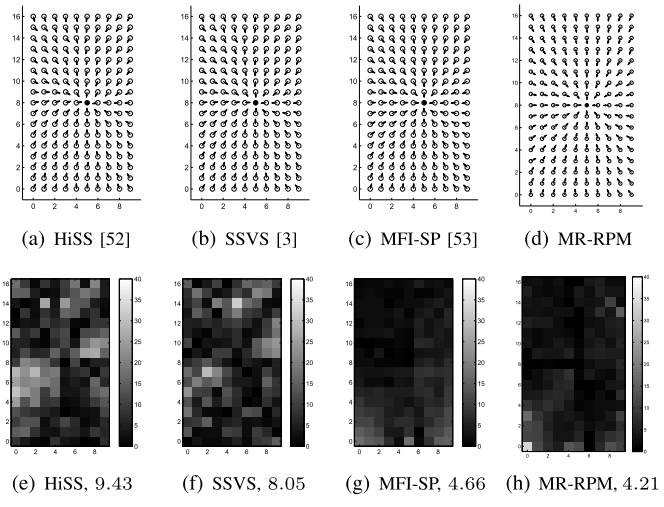


Fig. 6. Precisions (left), recalls (middle), and run times (right) of RANSAC, ICF, GS, VFC, and MR-RPM on a panoramic data set [3].



(a) HiSS [52] (b) SSVS [3] (c) MFI-SP [53] (d) MR-RPM

(e) HiSS, 9.43 (f) SSVS, 8.05 (g) MFI-SP, 4.66 (h) MR-RPM, 4.21

Fig. 7. Homing vectors and error analysis referring to grid position $(5, 8)$ in data set *A1originalH*. (a)–(d) Homing vectors. The solid circle in each figure is the homing position. (e)–(h) Angular errors for each position (unit: degree).

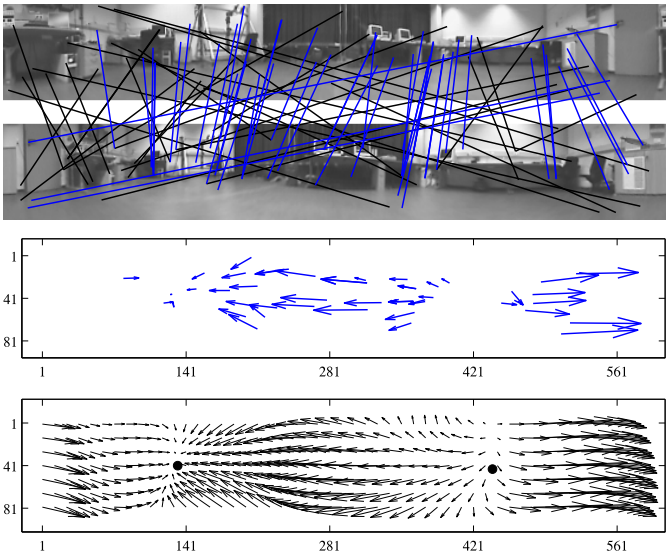


Fig. 8. Schematic of the feature matching and dense motion flow estimation results of MR-RPM. The feature matching result, where the blue and black lines indicate the preserved inliers and removed outliers (top). The corresponding sparse motion flow samples (middle). Dense motion flow estimated based on the preserved matches by MR-RPM, where the black dots indicate the localized FOC and FOE (bottom).

checking of each putative match in each image pair, and only 23 image pairs with large viewpoint changes are selected for quantitative evaluation. This method not only makes the test data challenging but also simplifies the construction of the ground truth. Four state-of-the-art feature matching algorithms, namely, RANSAC [56], identifying correspondence function (ICF) [61], graph shift (GS) [62], and VFC [24], are adopted for quantitative comparison.

The matching results of different approaches are reported in Fig. 6. The average inlier ratio in the putative sets is approximately 76.53%, and the average number of putative matches is approximately 125.2. The results show that MR-RPM clearly has the best precision and recall tradeoff. RANSAC has the best precision but the worst recall because the panoramic pair does not satisfy a parametric model exactly; thus, only a part of the true matches can be identified. The missing matches will inevitably affect the subsequent dense motion field interpolation. GS and ICF have similar performances but have middle-rank precisions and recalls. VFC has slightly better precisions than MR-RPM; however, the recalls of the proposed method are better, which is important for learning accurate dense motion flow in areas with few sparse feature matches. The runtime statistics of the different methods are also provided on the rightmost figure in Fig. 6. The average run time of MR-RPM is approximately 59 ms, which ranks in the middle among run times of the other methods.

2) Visual Homing on Panoramic Images: The method for visual homing is further tested. Fig. 7 provides several intuitive results of the different methods on homing performance. Position $(5, 8)$ of the *A1originalH* data set is considered as the reference home position, and the homing vectors calculated from other images using the four methods are shown in Fig. 7(a)–(d). The corresponding average angular errors for each position of the data set are shown in Fig. 7(e)–(h). The results show that MR-RPM can provide more accurate homing results.

The feature matching result and estimated dense motion flow on a typical image pair are schematically shown in Fig. 8. Clearly, all the inliers and outliers in the putative set are correctly distinguished. The estimated dense motion flow, FOC, and FOE are also consistent with the real motion flow. In this example, the FOC and FOE are approximately $(440, 41)$ and $(133, 41)$, respectively. The proposed method usually takes approximately 9 ms to localize the FOE/FOC.

TABLE II
STATISTICS OF VISUAL HOMING ERROR BY USING DIFFERENT METHODS (UNIT: DEGREE)

Database	HiSS [52]				SSVS [3]				MFI-SP [53]				MR-RPM			
	TAAE	Min	Max	Var	TAAE	Min	Max	Var	TAAE	Min	Max	Var	TAAE	Min	Max	Var
AloriginalH	14.67	8.05	36.40	5.42	12.59	6.50	28.36	**4.17**	7.78	**3.06**	**26.70**	4.75	**7.53**	3.11	26.72	4.57
CHall1H	11.69	8.05	18.84	**1.81**	15.94	10.79	28.84	3.32	7.27	3.67	16.80	2.46	**7.12**	**3.47**	**16.58**	2.31
CHall2H	15.75	10.53	27.59	**2.92**	24.69	12.46	55.16	7.22	13.89	**7.42**	29.24	4.31	**13.39**	7.48	**26.97**	4.35
KitchenH	21.76	12.86	47.62	6.36	24.29	13.75	51.65	7.10	20.10	11.94	42.05	5.90	**19.55**	**10.76**	**38.84**	**5.78**
Roeben1H	28.95	10.96	61.07	12.27	26.70	8.99	64.64	12.82	24.52	8.33	60.07	12.09	**22.87**	**7.79**	**59.28**	**12.01**

The statistics of the homing vector errors of all methods on the test database are reported in Table II. In general, MR-RPM can produce results that are better than or comparable to the results of other competitors.

VI. CONCLUSION

This paper proposes a novel approach named MR-RPM for the nonrigid registration of 2-D shape patterns and 3-D point clouds. Our approach uses manifold regularization to exploit the intrinsic geometrical structures of the given data, resulting in a reliable estimate of the underlying transformation. Fast implementation has also been provided to reduce algorithm complexity from cubic to quadratic. This enhances the suitability of the proposed method for the large-scale data (particularly 3-D point cloud). The superiority of MR-RPM over state-of-the-art alternatives is demonstrated on public 2-D and 3-D data sets qualitatively and quantitatively, particularly in the context of significant nonrigid deformations and/or occlusions. Moreover, the proposed method is applied to a real-world task, that is, visual homing, and it realizes a more advanced performance compared with other state-of-the-art methods.

APPENDIX
PROOF OF THEOREM 1

For any given reproducing kernel Γ, a unique RKHS $\mathcal{H}_M$ can be defined by considering the completion of the space as follows:

$$\mathcal{H}_M = \left\{ \sum_{i=1}^{M} \Gamma(\cdot, \mathbf{x}_i)\mathbf{c}_i : \mathbf{c}_i \in \mathcal{Y} \right\} \tag{24}$$

and its norm is induced by the following inner product:

$$\langle \mathbf{f}, \mathbf{g} \rangle_{\mathcal{H}} = \sum_{i,j=1}^{M} \langle \Gamma(\mathbf{x}_j, \mathbf{x}_i)\mathbf{c}_i, \mathbf{d}_j \rangle \quad \forall \mathbf{f}, \mathbf{g} \in \mathcal{H}_M \tag{25}$$

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

Let $\mathcal{H}_M^{\perp}$ stand for a subspace of $\mathcal{H}$, and it has the form

$$\mathcal{H}_M^{\perp} = \{ \mathcal{T} \in \mathcal{H} : \mathcal{T}(\mathbf{x}_i) = 0, n \in \mathbb{N}_M \}. \tag{26}$$

Considering the reproducing property [24], [49], $\forall \mathcal{T} \in \mathcal{H}_M^{\perp}$, we have the inner product

$$\left\langle \mathcal{T}, \sum_{i=1}^{M} \Gamma(\cdot, \mathbf{x}_i)\mathbf{c}_i \right\rangle_{\mathcal{H}} = \sum_{i=1}^{M} \langle \mathcal{T}(\mathbf{x}_i), \mathbf{c}_i \rangle = 0. \tag{27}$$

That is to say, the two spaces $\mathcal{H}_M^{\perp}$ and $\mathcal{H}_M$ are orthogonal, and each $\mathcal{T} \in \mathcal{H}$ can be decomposed into two orthogonal components: $\mathcal{T} = \mathcal{T}_M + \mathcal{T}_M^{\perp}$, where $\mathcal{T}_M \in \mathcal{H}_M$ and $\mathcal{T}_M^{\perp} \in \mathcal{H}_M^{\perp}$. Moreover, $\mathbf{t}_M = (\mathcal{T}_M(\mathbf{x}_1), \ldots, \mathcal{T}_M(\mathbf{x}_M))^T$. Based on the orthogonality $\|\mathcal{T}_M + \mathcal{T}_M^{\perp}\|_{\mathcal{H}}^2 = \|\mathcal{T}_M\|_{\mathcal{H}}^2 + \|\mathcal{T}_M^{\perp}\|_{\mathcal{H}}^2$ and the reproducing property $\mathcal{T}(\mathbf{x}_i) = \mathcal{T}_M(\mathbf{x}_i)$, the regularized risk functional satisfies the following expression:

$$\begin{aligned}
\mathcal{E}(\mathcal{T}) &= \frac{1}{2\sigma^2} \sum_{i=1}^{L} p_i \|\mathbf{y}_i - \mathcal{T}(\mathbf{x}_i)\|^2 + \frac{\lambda_1}{2} \left\| \mathcal{T}_M + \mathcal{T}_M^{\perp} \right\|_{\mathcal{H}}^2 \\
&\quad + \frac{\lambda_2}{2} \mathrm{tr}(\mathbf{t}^T \mathbf{A} \mathbf{t}) \\
&\geq \frac{1}{2\sigma^2} \sum_{i=1}^{L} p_i \|\mathbf{y}_i - \mathcal{T}_M(\mathbf{x}_i)\|^2 + \frac{\lambda_1}{2} \|\mathcal{T}_M\|_{\mathcal{H}}^2 \\
&\quad + \frac{\lambda_2}{2} \mathrm{tr}\left(\mathbf{t}_M^T \mathbf{A} \mathbf{t}_M\right).
\end{aligned} \tag{28}$$

Thus, the optimal solution of the objective function (15) is derived from the space $\mathcal{H}_M$ and hence can be expressed as (16). By defining the functional $\|\mathbf{f}\|_{\mathcal{H}}$ as $\|\mathbf{f}\|_{\mathcal{H}} = (\langle \mathbf{f}, \mathbf{f} \rangle_{\mathcal{H}})^{1/2}$ and considering the inner product in (25), the manifold regularized risk functional is then converted into the matrix form as follows:

$$\begin{aligned}
\mathcal{E}(\mathcal{T}) = \frac{1}{2\sigma^2} \|\mathbf{P}^{1/2}(\mathbf{Y} - \mathbf{J}\Gamma\mathbf{C})\|_F^2 + \frac{\lambda_1}{2} \mathrm{tr}(\mathbf{C}^T \Gamma \mathbf{C}) \\
+ \frac{\lambda_2}{2} \mathrm{tr}(\mathbf{C}^T \Gamma \mathbf{A} \Gamma \mathbf{C})
\end{aligned} \tag{29}$$

where Γ is an $M \times M$ matrix and $\Gamma_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j)$, $\mathbf{J} = (\mathbf{I}_{L \times L}, \mathbf{0}_{L \times (M-L)})$, and $\mathbf{C} = (\mathbf{c}_1, \ldots, \mathbf{c}_M)^T$ is the $M \times D$ coefficient matrix. The linear system in (17) can be obtained by considering the derivative of (29) with respect to $\mathbf{C}$ and setting it to zero. Therefore, the coefficient set $\{\mathbf{c}_i\}_{i=1}^{M}$ of the optimal solution $\mathcal{T}$ is determined using the linear system (17).

REFERENCES

[1] L. G. Brown, "A survey of image registration techniques," *ACM Comput. Surv.*, vol. 24, no. 4, pp. 325–376, Dec. 1992.
[2] P. J. Besl and D. N. McKay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992.
[3] M. Liu, C. Pradalier, and R. Siegwart, "Visual homing from scale with an uncalibrated omnidirectional camera," *IEEE Trans. Robot.*, vol. 29, no. 6, pp. 1353–1365, Dec. 2013.
[4] 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.

[5] 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.

[6] A. W. Fitzgibbon, "Robust registration of 2D and 3D point sets," *Image Vis. Comput.*, vol. 21, no. 13, pp. 1145–1153, Dec. 2003.

[7] 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.

[8] 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, Nov. 2017.

[9] A. Rangarajan, H. Chui, and F. L. Bookstein, "The softassign procrustes matching algorithm," in *Information Processing in Medical Imaging*. New York, NY, USA: Springer, 1997, pp. 29–42.

[10] 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.

[11] G. Wang, Z. Wang, Y. Chen, Q. Zhou, and W. Zhao, "Context-aware Gaussian fields for non-rigid point set registration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 5811–5819.

[12] H. Ling and D. W. Jacobs, "Shape classification using the inner-distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 2, pp. 286–299, Feb. 2007.

[13] 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.

[14] Y. Tsin and T. Kanade, "A correlation-based approach to robust point set registration," in *Proc. Eur. Conf. Comput. Vis.*, May 2004, pp. 558–569.

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

[16] 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.

[17] 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.

[18] 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.

[19] S. Ge, G. Fan, and M. Ding, "Non-rigid point set registration with global-local topology preservation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 245–251.

[20] S. Ge and G. Fan, "Articulated non-rigid point set registration for human pose estimation from 3D sensors," *Sensors*, vol. 15, no. 7, pp. 15218–15245, 2015.

[21] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3D registration," in *Proc. 26th IEEE Int. Conf. Robot. Autom.*, May 2009, pp. 3212–3217.

[22] Y. Zheng and D. Doermann, "Robust point matching for nonrigid shapes by preserving local neighborhood structures," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 643–649, Apr. 2006.

[23] J. Yang, Z. Cao, and Q. Zhang, "A fast and robust local descriptor for 3D point cloud registration," *Inf. Sci.*, vols. 346–347, pp. 163–179, Jun. 2016.

[24] 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.

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

[26] G. Wang, Z. Wang, Y. Chen, and W. Zhao, "A robust non-rigid point set registration method based on asymmetric Gaussian representation," *Comput. Vis. Image Understand.*, vol. 141, pp. 67–80, Dec. 2015.

[27] 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, pp. 2399–2434, Nov. 2006.

[28] 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.

[29] T. Poggio and F. Girosi, "Networks for approximation and learning," *Proc. IEEE*, vol. 78, no. 9, pp. 1481–1497, Sep. 1990.

[30] J. Ma, J. Zhao, J. Jiang, and H. Zhou, "Non-rigid point set registration with robust transformation estimation under manifold regularization," in *Proc. AAAI Conf. Artif. Intell.*, 2017, pp. 4218–4224.

[31] A. Zaharescu, E. Boyer, K. Varanasi, and R. Horaud, "Surface feature detection and description with applications to mesh matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 373–380.

[32] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*. Chelmsford, MA, USA: Courier Corporation, 1982.

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

[34] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*. Washington, DC, USA: Winston, 1977.

[35] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, 2003.

[36] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using Laplacianfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328–340, Mar. 2005.

[37] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.

[38] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

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

[40] M. Zhao, Z. Zhang, and T. W. S. Chow, "Trace ratio criterion based generalized discriminative learning for semi-supervised dimensionality reduction," *Pattern Recognit.*, vol. 45, no. 4, pp. 1482–1499, 2012.

[41] M. Zhao, T. W. S. Chow, Z. Wu, Z. Zhang, and B. Li, "Learning from normalized local and global discriminative information for semi-supervised regression and dimensionality reduction," *Inf. Sci.*, vol. 324, pp. 286–309, Dec. 2015.

[42] Y. Luo, D. Tao, C. Xu, C. Xu, H. Liu, and Y. Wen, "Multiview vector-valued manifold regularization for multilabel image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 5, pp. 709–722, May 2013.

[43] X. Zhen, M. Yu, A. Islam, M. Bhaduri, I. Chan, and S. Li, "Descriptor learning via supervised manifold regularization for multioutput regression," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 9, pp. 2035–2047, Sep. 2017.

[44] Z. Xu, I. King, M. R.-T. Lyu, and R. Jin, "Discriminative semi-supervised feature selection via manifold regularization," *IEEE Trans. Neural Netw.*, vol. 21, no. 7, pp. 1033–1047, Jul. 2010.

[45] M. Zhao, T. W. S. Chow, Z. Zhang, and B. Li, "Automatic image annotation via compact graph based semi-supervised learning," *Knowl.-Based Syst.*, vol. 76, pp. 148–165, Mar. 2015.

[46] S. Xiang, C. Pan, F. Nie, and C. Zhang, "Turbopixel segmentation using Eigen-images," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 3024–3034, Nov. 2010.

[47] H. Q. Minh and V. Sindhwani, "Vector-valued manifold regularization," in *Proc. Int. Conf. Mach. Learn.*, 2011, pp. 57–64.

[48] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer-Verlag, 2006.

[49] C. A. Micchelli and M. A. Pontil, "On learning vector-valued functions," *Neural Comput.*, vol. 17, no. 1, pp. 177–204, 2005.

[50] R. Rifkin, G. Yeo, and T. Poggio, "Regularized least-squares classification," in *Advances in Learning Theory: Methods, Model and Applications*. Cambridge, MA, USA: MIT Press, 2003.

[51] J. Ma, J. Zhao, J. Tian, X. Bai, and Z. Tu, "Regularized vector field learning with sparse approximation for mismatch removal," *Pattern Recognit.*, vol. 46, no. 12, pp. 3519–3532, 2013.

[52] D. Churchill and A. Vardy, "An orientation invariant visual homing algorithm," *J. Intell. Robot. Syst.*, vol. 71, no. 1, pp. 3–29, 2013.

[53] J. Zhao and J. Ma, "Visual homing by robust interpolation for sparse motion flow," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2017, pp. 1282–1288.

[54] A. Ramisa, A. Goldhoorn, D. Aldavert, R. Toledo, and R. L. de Mantaras, "Combining invariant features and the ALV homing method for autonomous robot navigation based on panoramas," *J. Intell. Robot. Syst.*, vol. 64, nos. 3–4, pp. 625–649, 2011.

[55] D. Schroeter and P. Newman, "On the robustness of visual homing under landmark uncertainty," in *Intelligent Autonomous Systems*. Clifton, VA, USA: IOS Press, 2008, pp. 278–287.

[56] 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.
[57] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[58] R. Möller and A. Vardy, "Local visual homing by matched-filter descent in image distances," *Biol. Cybern.*, vol. 95, no. 5, pp. 413–430, 2006.
[59] J. Ma *et al.*, "Visual homing via guided locality preserving matching," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2018, pp. 7254–7261.
[60] V. G. Kim, Y. Lipman, and T. Funkhouser, "Blended intrinsic maps," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 79:1–79:12, 2011.
[61] X. Li and Z. Hu, "Rejecting mismatches by correspondence function," *Int. J. Comput. Vis.*, vol. 89, no. 1, pp. 1–17, 2010.
[62] H. Liu and S. Yan, "Common visual pattern discovery via spatially coherent correspondences," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 1609–1616.

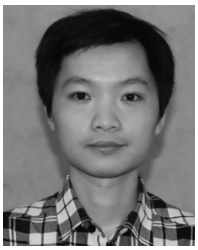
Jiayi Ma received the B.S. degree in mathematics and the Ph.D. degree in automation from the Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2014, respectively.

From 2012 to 2013, he was an Exchange Student with the Department of Statistics, University of California at Los Angeles, Los Angeles, CA, USA. He is currently an Associate Professor with the Electronic Information School, Wuhan University, Wuhan, where he was a Post-Doctoral Researcher from 2014 to 2015.

Dr. Ma is an Associate Editor of the IEEE ACCESS. He has authored or co-authored over 90 refereed journal and conference papers. His current research interests include the areas of computer vision, machine learning, and pattern recognition.

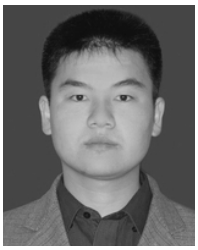
Jia Wu is currently a Lecturer with the Department of Computing, Faculty of Science and Engineering, Macquarie University, Sydney, NSW, Australia. Since 2009, he has authored or co-authored over 60 refereed journal and conference papers, such as the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, the IEEE TRANSACTIONS ON CYBERNETICS, *ACM Transactions on Knowledge Discovery from Data*, *Pattern Recognition*, IJCAI, AAAI, ICDE, ICDM, SDM, CIKM, PAKDD, and IJCNN in these areas. His current research interests include data mining and machine learning.

Mr. Wu is an Associate Editor of *ACM Transactions on Knowledge Discovery from Data* and *Journal of Network and Computer Applications* (Elsevier). He has successfully organized a special session in IJCNN 2016 and IJCNN 2017, respectively. He was a recipient of the IJCNN'17 Best Student Paper Award.



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

From 2012 to 2014, he was a Post-Doctoral Research Associate with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA. He is currently a Research Fellow with the Electronic Information School, Wuhan University, Wuhan. His current research interests include computer vision and machine learning.



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

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

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

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

Quan Z. Sheng received the Ph.D. degree in computer science from the University of New South Wales, Sydney, NSW, Australia.

He was a Post-Doctoral Research Scientist with CSIRO ICT Centre, Canberra, ACT, Australia. He is currently a Full Professor and the Head of the Department of Computing, Macquarie University, Sydney. He has authored or co-authored over 300 publications as edited books and proceedings, refereed book chapters, and refereed technical papers in journals and conferences including *ACM Computing Surveys*, *ACM Transactions on Internet Technology*, *ACM Transactions on Multimedia Computing, Communications, and Applications*, *ACM Transactions on Knowledge Discovery from Data*, *ACM Transactions on Intelligent Systems and Technology*, *VLDB Journal*, Computer (Oxford), the IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, *DAPD*, the IEEE TRANSACTIONS ON SERVICES COMPUTING, WWWJ, *IEEE Computer*, *IEEE Internet Computing*, Commun. ACM, VLDB, ICDE, ICDM, CIKM, EDBT, WWW, ICSE, ICSOC, ICWS, and CAiSE. His current research interests include big data analytics, data mining, Internet of Things, service computing, and Web technologies.

Dr. Sheng is a member of the ACM. He was a recipient of several prestigious awards including the ARC Future Fellowship in 2014, the Chris Wallace Award for Outstanding Research Contribution in 2012, and the Microsoft Research Fellowship in 2003.