Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)

Non-Rigid Point Set Registration with Robust Transformation Estimation under Manifold Regularization

Jiayi Ma,[1] **Ji Zhao,**[1] **Junjun Jiang,**[2] **Huabing Zhou**[3]
[1]Electronic Information School, Wuhan University, Wuhan 430072, China
[2]School of Computer Science, China University of Geosciences, Wuhan 430074, China
[3]Hubei Provincial Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan 430073, China
{jyma2010, zhaoji84, zhouhuabing}@gmail.com, junjun0595@163.com

Abstract

In this paper, we propose a robust transformation estimation method based on manifold regularization for non-rigid point set registration. The method iteratively recovers the point correspondence and estimates the spatial transformation between two point sets. The correspondence is established based on existing local feature descriptors which typically results in a number of outliers. To achieve an accurate estimate of the transformation from such putative point correspondence, we formulate the registration problem by a mixture model with a set of latent variables introduced to identify outliers, and a prior involving manifold regularization is imposed on the transformation to capture the underlying intrinsic geometry of the input data. The non-rigid transformation is specified in a reproducing kernel Hilbert space and a sparse approximation is adopted to achieve a fast implementation. Extensive experiments on both 2D and 3D data demonstrate that our method can yield superior results compared to other state-of-the-arts, especially in case of badly degraded data.

Introduction

Point set registration is a fundamental problem and frequently encountered in computer vision, pattern recognition, medical imaging and remote sensing (Brown 1992). Many tasks in these fields including 3D reconstruction, shape recognition, panoramic stitching, feature-based image registration and content-based image retrieval can be solved by algorithms operating on the point sets (e.g., salient point features) extracted from the input data (Ma et al. 2015a; 2015c; Bai et al. 2017; Zhou et al. 2016). The goal of point set registration is then to determine the right correspondence and/or to recover the spatial transformation between the two point sets (Jian and Vemuri 2011). In this paper, we focus on non-rigid registration where the transformation is characterized by a nonlinear or non-parameterized model (Zhao et al. 2011; Ma et al. 2013; Wang et al. 2015).

The registration problem is typically solved by using an iterative framework, where a set of putative correspondences is established and used to refine the estimate of transformation, and vice versa (Besl and McKay 1992; Chui and Rangarajan 2003). In this process, the most challenging and critical task is to develop an efficient strategy for robust transformation estimation from putative correspondences. First, the putative correspondences are usually established based on only local feature descriptors, where the unavoidable noise, repeated structures and occlusions often lead to a high number of false correspondences. Therefore, a robust procedure of outlier removal is required. Second, to establish reliable correspondence, the putative set usually removes a large part of the original point sets whose feature descriptors are not similar enough. However, the point sets are typically extracted from the contour or surface of a specific object, and hence can provide intrinsic structure information of the input data which is beneficial to the transformation estimation. Therefore, it is desirable to incorporate the whole point sets into the objective function formulation during transformation estimation. Third, for large scale point cloud data, the number of points can reach tens of thousands. This poses a significant burden on typical point registration methods, particularly in the non-rigid case. Therefore, it is of particular advantage to develop a more efficient technique.

To address these issues, we formulate the registration problem by a mixture model with a set of latent variables introduced to identify outliers. We also assume a prior on the geometry involving manifold regularization to impose a non-parametric smoothness constraint on the spatial transformation (Ma et al. 2014; Belkin, Niyogi, and Sindhwani 2006). The manifold regularization defined on the whole input point sets controls the complexity of the transformation and is able to capture the underlying intrinsic geometry of the input data. This leads to a maximum a posteriori (MAP) estimation problem which can be solved by using the Expectation-Maximization (EM) algorithm (Dempster, Laird, and Rubin 1977) to estimate the variance of the prior, while simultaneously estimating the outliers, with the variance given a large initial value. Moreover, a sparse approximation based on a similar idea as the subset of regressors method (Poggio and Girosi 1990) is introduced to improve the computational efficiency.

Our contribution in this paper includes the following three aspects. Firstly, we introduce the manifold regularization to the point set registration problem, which can capture the intrinsic geometry of the input point sets and hence helps to estimate the transformation. Secondly, we propose a new formulation for robust transformation estimation based on manifold regularization, which could estimate transforma-

tion from point correspondences contaminated by outliers. Thirdly, we provide a fast implementation for our method by using sparse approximation, which enables our method to handle large scale datasets such as 3D point clouds.

Related work

The iterated closest point (ICP) algorithm (Besl and McKay 1992) is one of the representative approaches using the iterative framework to solve the registration problem. In (Chui and Rangarajan 2003), Chui and Rangarajan developed a general framework for non-rigid registration called TPS-RPM. Different from ICP, which uses the nearest point strategy in learning the correspondence, TPS-RPM introduces soft assignments and solves it in a continuous optimization framework involving deterministic annealing. Zheng and Doermann (Zheng and Doermann 2006) proposed a method called RPM-LNS which can preserve local neighborhood structures during matching, where the shape context (SC) feature (Belongie, Malik, and Puzicha 2002) is used to initialize the correspondence. Ma et al. (Ma et al. 2015b) introduced a non-rigid registration strategy based on Gaussian fields, which was later improved in (Wang et al. 2016) by using inner distance shape context (Ling and Jacobs 2007) to construct initial correspondences. In the recent past, the point registration is typically solved by probabilistic methods (Jian and Vemuri 2011; Myronenko and Song 2010; Horaud et al. 2011; Ma, Zhao, and Yuille 2016). Specifically, to cope with highly articulated deformation, A global-local topology preservation (GLTP) method (Ge, Fan, and Ding 2014; Ge and Fan 2015) is proposed based on coherent point drift (CPD) (Myronenko and Song 2010). These methods formulated registration as the estimation of a mixture of densities using GMMs, and the problem is solved using the framework of maximum likelihood and the EM algorithm.

Method

Suppose we are given 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 $\mathbf{y}_j$ are D dimensional column vectors denoting the point positions (typically $D = 2$ or 3), M and N are the numbers of points in the two sets, respectively. To solve the registration problem, we consider an iterative strategy which first constructs a putative set of correspondences by using the local geometric structures of points, and then estimates the spatial transformation based on the putative set together with some global geometric constraints. In the following, we start by introducing the correspondence estimation based on local structure information, and then lay out the manifold regularization which could capture the underlying spatial geometry of a point set. We subsequently propose a formulation for robust transformation estimation from putative correspondences based on manifold regularization, and followed by the fast implementation using sparse approximation. Finally, we present the implementation details of the proposed approach.

Correspondence estimation

For two point sets representing similar shapes or objects, their corresponding points will in general have similar local geometric structures (e.g., neighborhood structures) which could be incorporated into a feature descriptor. Therefore, the correspondences could be established by finding for each point in one point set (e.g., the model) the point on the other point set (e.g., the target) that has the most similar feature descriptor. Fortunately, there are several well-designed feature descriptors that can fulfill this task, both in 2D and in 3D cases (Belongie, Malik, and Puzicha 2002; Rusu, Blodow, and Beetz 2009).

For 2D case, the shape context (Belongie, Malik, and Puzicha 2002) has been a widely used feature descriptor. Consider two points $\mathbf{x}_i$ and $\mathbf{y}_j$, their SCs which capture the distributions of their neighborhood points are histograms $\{p_i(k)\}_{k=1}^K$ and $\{q_j(k)\}_{k=1}^K$, respectively. The χ^2 distance is used to measure their difference $C(\mathbf{x}_i, \mathbf{y}_j)$:

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

After we have obtained the distances of all point pairs, i.e., $\{C(\mathbf{x}_i, \mathbf{y}_j),\ i = 1, \cdots, M,\ j = 1, \cdots, N\}$, the Hungarian method (Papadimitriou and Steiglitz 1982) is applied to seek the correspondences between $\{\mathbf{x}_i\}_{i=1}^M$ and $\{\mathbf{y}_j\}_{j=1}^N$.

For 3D case, we consider the fast point feature histograms (FPFH) (Rusu, Blodow, and Beetz 2009) as the feature descriptor. It is a histogram that collects the pairwise pan, tilt and yaw angles between every point and its k-nearest neighbors, followed by a reweighting of the resultant histogram of a point with the neighboring histograms. The computation of the histogram is quite efficient which has linear complexity with respect to the number of surface normals. The matching of FPFH descriptors is performed by a sample consensus initial alignment method.

After using some local feature descriptor to establish correspondence, we obtain a putative set $S = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^L$, where $L \leq \min\{M, N\}$ is the number of correspondence. Without loss of generality, we assume that $\{\mathbf{x}_i\}_{i=1}^L$ and $\{\mathbf{y}_i\}_{i=1}^L$ in the putative set correspond to the first L points in the original model point set $\{\mathbf{x}_i\}_{i=1}^M$ and the first L points in the original target point set $\{\mathbf{y}_j\}_{j=1}^N$, respectively.

Solve transformation with manifold regularization

Given a putative correspondence set $S = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^L$ established from two point sets $\{\mathbf{x}_i\}_{i=1}^M$ and $\{\mathbf{y}_j\}_{j=1}^N$, our purpose is to estimate the underlying spatial transformation $\mathcal{T}$ between them, for example, $\mathbf{y}_i = \mathcal{T}(\mathbf{x}_i)$ for any correspondence $(\mathbf{x}_i, \mathbf{y}_i)$ in S. This problem is in general ill-posed as $\mathcal{T}$ is non-rigid which has an infinite number of solutions. To obtain a meaningful solution, the regularization technique could be used which typically operates in a Reproducing Kernel Hilbert Space (RKHS) (Aronszajn 1950) (associated with a particular kernel). Specifically, the Tikhonov regularization (Tikhonov and Arsenin 1977) in an RKHS $\mathcal{H}$ minimizes a regularized risk functional:

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

where the first term enforces closeness to the data, the second term controls complexity of the transformation $\mathcal{T}$, λ is

a regularization parameter controlling the trade-off between these two terms, and $\|\cdot\|_{\mathcal{H}}$ denotes the norm of $\mathcal{H}$ (we will discuss the detailed forms of $\mathcal{T}$ and $\|\cdot\|_{\mathcal{H}}$ later).

Recall that in our problem, due to the existence of noise, outliers, occlusions, etc., the number of matched points is typically less than the whole point set, i.e., $L \leq M$. That is to say, only L points $\mathbf{x}_1, \ldots, \mathbf{x}_L$ are given labels $\mathbf{y}_1, \ldots, \mathbf{y}_L$ drawn from the spatial transformation $\mathcal{T}$. However, for point set registration, the points we wish to match are usually extracted from a shape contour or an object surface which possess some sort of "intrinsic geometry". Fore instance, the point positions for each type of shapes or objects are not arbitrary, which often form a specific distribution. Therefore, the $M-L$ unlabeled points can provide additional information about the characteristic of the point set. To make full use of such additional information, we consider the manifold regularization (Belkin, Niyogi, and Sindhwani 2006; Minh and Sindhwani 2011; Zhao et al. 2015). It introduces an additional regularizer $\|\mathcal{T}\|_{\mathcal{I}}^2$ to penalize $\mathcal{T}$ along a low-dimensional manifold, which is defined on the whole input set $\{\mathbf{x}_i\}_{i=1}^M$. Thus the regularized risk functional becomes:

$$\mathcal{T}^* = \min_{\mathcal{T}\in\mathcal{H}} \sum_{i=1}^{L} \|\mathbf{y}_i - \mathcal{T}(\mathbf{x}_i)\|^2 + \lambda_1 \|\mathcal{T}\|_{\mathcal{H}}^2 + \lambda_2 \|\mathcal{T}\|_{\mathcal{I}}^2, \quad (3)$$

where the parameter λ_1 controls the complexity of the transformation in the input space, and λ_2 regularizes with respect to the intrinsic geometry. The term λ_1 is necessary since the manifold is a strict subset of the input space; among many $\mathcal{T} \in \mathcal{H}$ which give the same value on the manifold, we prefer a solution which is smooth in the input space.

To define the manifold regularization term, we use the graph Laplacian which is a discrete analogue of the manifold Laplacian (Belkin, Niyogi, and Sindhwani 2006). It models a manifold using the weighted neighborhood graph for the data based on an assumption that the input points are drawn i.i.d. from the manifold. Consider the weighted neighborhood graph G given by taking the graph on 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 $\|\mathbf{x}_i - \mathbf{x}_j\|^2 \leq \epsilon$, and assigning to edge $(\mathbf{x}_i, \mathbf{x}_j)$ the weight

$$W_{ij} = e^{-\frac{1}{\epsilon}\|\mathbf{x}_i - \mathbf{x}_j\|^2}. \quad (4)$$

The graph Laplacian of G is the matrix $\mathbf{A}$ given by

$$A_{ij} = D_{ij} - W_{ij}, \quad (5)$$

where $\mathbf{D} = \operatorname{diag}\big(\sum_{j=1}^M W_{ij}\big)_{i=1}^M$, i.e., the diagonal matrix whose i-th entry is the sum of the weights of edges leaving $\mathbf{x}_i$. Let $\mathbf{t} = (\mathcal{T}(\mathbf{x}_1), \ldots, \mathcal{T}(\mathbf{x}_M))^{\mathrm{T}}$, then the manifold regularization term can be defined as:

$$\|\mathcal{T}\|_{\mathcal{I}}^2 = \sum_{i=1}^{M}\sum_{j=1}^{M} W_{ij}(t_i - t_j)^2 = \operatorname{tr}(\mathbf{t}^{\mathrm{T}}\mathbf{A}\mathbf{t}), \quad (6)$$

where $\operatorname{tr}(\cdot)$ denotes the trace. Therefore, the regularized risk functional (3) becomes:

$$\mathcal{T}^* = \min_{\mathcal{T}\in\mathcal{H}} \sum_{i=1}^{L} \|\mathbf{y}_i - \mathcal{T}(\mathbf{x}_i)\|^2 + \lambda_1 \|\mathcal{T}\|_{\mathcal{H}}^2 + \lambda_2 \operatorname{tr}(\mathbf{t}^{\mathrm{T}}\mathbf{A}\mathbf{t}). \quad (7)$$

We will discuss the solution of this manifold regularized risk functional in the next section.

Robust transformation estimation

The transformation could be solved by minimizing the regularized risk functional in Eq. (3). However, the putative set $S = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^L$ typically involves some unknown false correspondences, as only local neighborhood structures are considered. Therefore, it is important that the transformation estimation is robust to outliers. In this section, we propose a method for robust transformation estimation from point correspondences by using manifold regularization.

We make the assumption that, for the inliers, the noise of point position is Gaussian on each component (dimension) with zero mean and uniform standard deviation σ; for the outliers, the position of the target point $\mathbf{y}_i$ lies randomly in a bounded region of $\mathbb{R}^D$, and the distribution is assumed to be uniform $1/a$ with a denoting the volume of this region (Ma et al. 2014). We then associate the i-th correspondence with a latent variable $z_i \in \{0, 1\}$, where $z_i = 0$ indicates a uniform distribution and $z_i = 1$ points to a Gaussian distribution. Let $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_L)^{\mathrm{T}}$ and $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_L)^{\mathrm{T}} \in \mathbb{R}^{L\times D}$ be the two sets of points in the putative set. Thus, the likelihood is a mixture model given by

$$\begin{aligned} p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}) &= \prod_{i=1}^{L}\sum_{z_i} p(\mathbf{y}_i, z_i|\mathbf{x}_i, \boldsymbol{\theta}) \\ &= \prod_{i=1}^{L}\left(\frac{\gamma}{(2\pi\sigma^2)^{D/2}} e^{-\frac{\|\mathbf{y}_i - \mathcal{T}(\mathbf{x}_i)\|^2}{2\sigma^2}} + \frac{1-\gamma}{a}\right), \end{aligned} \quad (8)$$

where $\boldsymbol{\theta} = \{\mathcal{T}, \sigma^2, \gamma\}$ includes the unknown parameters, and γ is a mixing coefficient specifying the marginal distribution over the latent variable, i.e., $\forall z_i$, $p(z_i = 1) = \gamma$. We assume the non-rigid transformation $\mathcal{T}$ to lie within an RKHS, and it should also reflect the intrinsic structure of a point set. Thus a slow-and-smooth prior could be applied to $\mathcal{T}$: $p(\mathcal{T}) \propto e^{-\frac{1}{2}(\lambda_1\|\mathcal{T}\|_{\mathcal{H}}^2 + \lambda_2\|\mathcal{T}\|_{\mathcal{I}}^2)}$. Using Bayes rule, we estimate an MAP solution of $\boldsymbol{\theta}$:

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Y}) = \arg\max_{\boldsymbol{\theta}} p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}) p(\mathcal{T}). \quad (9)$$

To optimize this objective function, we consider the EM algorithm, which is a general technique for learning and inference in the context of latent variables. We follow standard notations (Bishop 2006) and omit some terms that are independent of $\boldsymbol{\theta}$. Considering the negative log posterior function, the complete-data log posterior is:

$$\begin{aligned} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\mathrm{old}}) = &-\frac{1}{2\sigma^2}\sum_{i=1}^{L} p_i \|\mathbf{y}_i - \mathcal{T}(\mathbf{x}_i)\|^2 - \frac{DL_p}{2}\ln\sigma^2 + \\ &L_p \ln\gamma + (L - L_p)\ln(1-\gamma) - \frac{\lambda_1}{2}\|\mathcal{T}\|_{\mathcal{H}}^2 - \frac{\lambda_2}{2}\|\mathcal{T}\|_{\mathcal{I}}^2, \end{aligned} \quad (10)$$

where $p_i = P(z_i = 1|\mathbf{x}_i, \mathbf{y}_i, \boldsymbol{\theta}^{\mathrm{old}})$, $L_p = \sum_{i=1}^L p_i$. The EM algorithm alternates with two steps: an expectation step (E-step) and a maximization step (M-step).

E-step: We use the current parameter values $\boldsymbol{\theta}^{\mathrm{old}}$ to find the posterior distribution of the latent variables. Denote $\mathbf{P} = \operatorname{diag}(p_1, \ldots, p_L)$ a diagonal matrix which can be computed

by applying Bayes rule:

$$p_i = \frac{\gamma e^{-\frac{\|\mathbf{y}_i - \mathcal{T}(\mathbf{x}_i)\|^2}{2\sigma^2}}}{\gamma e^{-\frac{\|\mathbf{y}_i - \mathbf{f}(\mathbf{x}_i)\|^2}{2\sigma^2}} + (1-\gamma)\frac{(2\pi\sigma^2)^{D/2}}{a}}. \quad (11)$$

The posterior probability p_l is a soft decision, which indicates to what degree the correspondence $(\mathbf{x}_i, \mathbf{y}_i)$ agrees with the current estimated transformation $\mathcal{T}$.

M-step: We determine the revised parameter estimate $\boldsymbol{\theta}^{\text{new}}$ as: $\boldsymbol{\theta}^{\text{new}} = \arg\max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$. Let $\mathcal{T}(\mathbf{X}) = (\mathcal{T}(\mathbf{x}_1), \ldots, \mathcal{T}(\mathbf{x}_L))^{\mathrm{T}}$. Considering the diagonal matrix $\mathbf{P}$ and taking derivative of $\mathcal{Q}(\boldsymbol{\theta})$ with respect to σ^2 and γ, and setting them to zero, we obtain

$$\sigma^2 = \frac{\text{tr}((\mathbf{Y} - \mathcal{T}(\mathbf{X}))^{\mathrm{T}}\mathbf{P}(\mathbf{Y} - \mathcal{T}(\mathbf{X})))}{DL_p}, \quad (12)$$

$$\gamma = \text{tr}(\mathbf{P})/L. \quad (13)$$

Next we consider the terms of $\mathcal{Q}(\boldsymbol{\theta})$ that are related to $\mathcal{T}$. We obtain a manifold regularized risk functional as (Micchelli and Pontil 2005):

$$\mathcal{E}(\mathcal{T}) = \frac{1}{2\sigma^2}\sum_{i=1}^{L} p_i \|\mathbf{y}_i - \mathcal{T}(\mathbf{x}_i)\|^2 + \frac{\lambda_1}{2}\|\mathcal{T}\|_{\mathcal{H}}^2 + \frac{\lambda_2}{2}\|\mathcal{T}\|_{\mathcal{I}}^2. \quad (14)$$

We define model the transformation $\mathcal{T}$ by requiring it to lie within an RKHS $\mathcal{H}$ defined by a matrix-valued kernel $\Gamma : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}^{D\times D}$. In this paper, we consider a diagonal decomposable kernel $\Gamma(\mathbf{x}, \mathbf{x}') = \kappa(\mathbf{x}, \mathbf{x}') \cdot \mathbf{I}$, where $\kappa(\mathbf{x}, \mathbf{x}') = e^{-\beta\|\mathbf{x}-\mathbf{x}'\|^2}$ is a scalar Gaussian kernel, with β determining the width of the range of interaction between samples. Therefore, we have the following representer theorem (Belkin, Niyogi, and Sindhwani 2006).

Theorem 1. *The optimal solution of the manifold regularized risk functional* (14) *is given by*

$$\mathcal{T}^*(\mathbf{x}) = \sum_{i=1}^{M} \Gamma(\mathbf{x}, \mathbf{x}_i)\mathbf{c}_i, \quad (15)$$

with the coefficients $\{\mathbf{c}_i\}_{i=1}^{M}$ *determined by a linear system*

$$(\mathbf{J}^{\mathrm{T}}\mathbf{P}\mathbf{J}\boldsymbol{\Gamma} + \lambda_1\sigma^2\mathbf{I} + \lambda_2\sigma^2\mathbf{A}\boldsymbol{\Gamma})\mathbf{C} = \mathbf{J}^{\mathrm{T}}\mathbf{P}\mathbf{Y}, \quad (16)$$

where $\boldsymbol{\Gamma} \in \mathbb{R}^{M\times M}$ *is the so-called Gram matrix with the* (i,j)*-th element being* $\kappa(\mathbf{x}_i, \mathbf{x}_j)$*,* $\mathbf{J} = (\mathbf{I}_{L\times L}, \mathbf{0}_{L\times(M-L)})$ *with* $\mathbf{I}$ *being an identity matrix and* $\mathbf{0}$ *being a matrix of all zeros,* $\mathbf{C} = (\mathbf{c}_1, \ldots, \mathbf{c}_M)^{\mathrm{T}} \in \mathbb{R}^{M\times D}$ *is a coefficient matrix.*

Convergence analysis. The objective function (9) is not convex, and hence it is unlikely that any optimization technique can find its global minimum. However, for many practical applications a stable local minimum is often enough. To achieve this goal, we use a large value to initialize the variance σ^2 so that the objective function becomes convex in a large region. In this situation, a lot of unstable shallow local minima can be filtered and a good minimum could be achieved. As the EM iteration proceeds, the value of σ^2 gradually decreases and the objective function tends to approach the true curve smoothly. This makes it likely that a better minimum could be reached by using the old minimum as the initial value, and finally converges to a stable local minimum. A similar concept has been introduced in deterministic annealing (Chui and Rangarajan 2003), where the solution of an easy problem is used to recursively give initial conditions to increasingly harder problems.

Fast implementation

The most time consuming step of our proposed algorithm is to solve the transformation $\mathcal{T}$ using linear system (16), which requires $O(M^3)$ time complexity and may pose a serious problem for large values of M. Even when it is implementable, a suboptimal but faster method may be a better choice. In this section, we provide a fast implementation based on a similar kind of idea as the subset of regressors method (Poggio and Girosi 1990).

Rather than searching for the optimal solution in the space of $\mathcal{H}_M = \left\{\sum_{i=1}^{M} \Gamma(\cdot, \mathbf{x}_i)\mathbf{c}_i\right\}$, we use a sparse approximation and search a suboptimal solution in a space with much less basis functions defined as $\mathcal{H}_K = \left\{\sum_{i=1}^{K} \Gamma(\cdot, \widetilde{\mathbf{x}}_i)\mathbf{c}_i\right\}$, and then minimize the manifold regularized risk functional over all the sample data. Here $K \ll M$ and we choose the point set $\{\widetilde{\mathbf{x}}_i : i \in \mathbb{N}_K\}$ as a random subset of $\{\mathbf{x}_i : i \in \mathbb{N}_M\}$ according to (Rifkin, Yeo, and Poggio 2003), who found that simply selecting an arbitrary subset of the training inputs performs no worse than those more sophisticated and time-consuming methods. Therefore, we search a solution with the form

$$\mathcal{T}(\mathbf{x}) = \sum_{i=1}^{K} \Gamma(\mathbf{x}, \widetilde{\mathbf{x}}_i)\mathbf{c}_i, \quad (17)$$

with the coefficients $\{\mathbf{c}_i\}_{i=1}^{K}$ determined by a linear system

$$(\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U} + \lambda_1\sigma^2\boldsymbol{\Gamma}_s + \lambda_2\sigma^2\mathbf{V}^{\mathrm{T}}\mathbf{A}\mathbf{V})\mathbf{C} = \mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{Y}, \quad (18)$$

where the Gram matrix $\boldsymbol{\Gamma}_s \in \mathbb{R}^{K\times K}$ with the (i,j)-th element being $\kappa(\widetilde{\mathbf{x}}_i, \widetilde{\mathbf{x}}_j)$, $\mathbf{U} \in \mathbb{R}^{L\times K}$ and $\mathbf{V} \in \mathbb{R}^{M\times K}$ with the (i,j)-th element being $\kappa(\mathbf{x}_i, \widetilde{\mathbf{x}}_j)$. Note that the matrix $\mathbf{U}$ is composed of the first L rows of the matrix $\mathbf{V}$. The derivation of Eq. (18) is similar to that of Theorem 1. Compared with the original method, the difference of the fast version is that it solves a linear system in Eq. (18) rather than Eq. (16).

Algorithm summary & computational complexity

The two steps of estimating correspondence and transformation are iterated to obtain a reliable result. In this paper, we use a fixed number of iterations, typically 10 but more in case of badly degraded data, for example, large degree of deformation, high level of noise or large percentage of outliers in the point sets. As our Robust Point Matching algorithm is based on Manifold Regularization, we name it *MR-RPM*. We summarize the MR-RPM method in Algorithm 1.

For the linear system (16), the matrix $\mathbf{J}^{\mathrm{T}}\mathbf{P}\mathbf{J}\boldsymbol{\Gamma} + \lambda_1\sigma^2\mathbf{I} + \lambda_2\sigma^2\mathbf{A}\boldsymbol{\Gamma}$ is of size $M \times M$, and hence it requires $O(M^3)$ time complexity to solve the transformation $\mathcal{T}$. However, for the linear system (18), the matrix $\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U} + \lambda_1\sigma^2\boldsymbol{\Gamma}_s + \lambda_2\sigma^2\mathbf{V}^{\mathrm{T}}\mathbf{A}\mathbf{V}$ is of size $K \times K$, and hence the time complexity for solving the linear system reduces to $O(K^3)$. Nevertheless, the time complexity of compute the matrix $\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U} + \lambda_1\sigma^2\boldsymbol{\Gamma}_s + \lambda_2\sigma^2\mathbf{V}^{\mathrm{T}}\mathbf{A}\mathbf{V}$ is $O(KM^2)$, due to the

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

1 Compute feature descriptors for target set $\{\mathbf{y}_j\}_{j=1}^N$;
2 Set a to the volume of the output space;
3 **repeat**
4 Compute feature descriptors for model set $\{\mathbf{x}_i\}_{i=1}^M$;
5 Construct $S = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^L$ using descriptors;
6 Compute graph Laplacian $\mathbf{A}$ by Eqs. (4) and (5);
7 Compute Gram matrix $\mathbf{\Gamma}$ using the definition of Γ;
8 Initialize γ, $\mathbf{P} = \mathbf{I}$, $\mathcal{T}(\mathbf{x}_i) = \mathbf{x}_i$, and σ^2 by Eq. (12);
9 **repeat**
10 *E-step:*
11 Update $\mathbf{P} = \mathrm{diag}(p_1, \ldots, p_L)$ by Eq. (11);
12 *M-step*:
13 Update σ^2 and γ by Eqs. (12) and (13);
14 Update $\mathbf{C}$ by solving linear system (18);
15 **until** $\mathcal{Q}$ *converges*;
16 Update model point set $\{\mathbf{x}_i\}_{i=1}^M \leftarrow \{\mathcal{T}(\mathbf{x}_i)\}_{i=1}^M$;
17 **until** *reach the maximum iteration number*;
18 The aligned model point set $\{\hat{\mathbf{x}}_i\}_{i=1}^M$ is given by $\{\mathcal{T}(\mathbf{x}_i)\}_{i=1}^M$ in the last iteration.

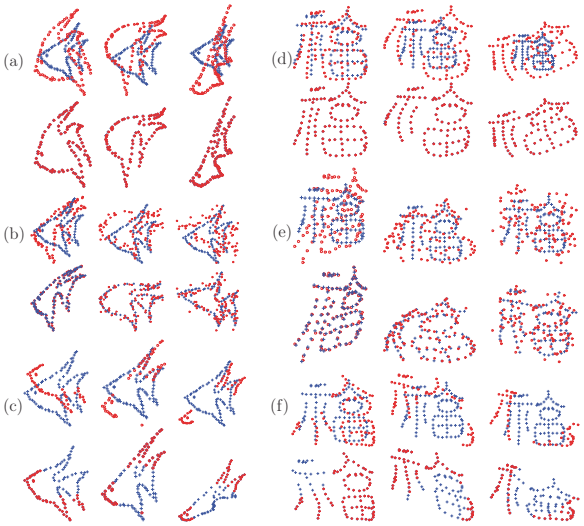

Figure 1: Qualitative results of our MR-GRL algorithm on the fish (a-c) and Chinese character (d-f) shapes shown in every two rows. For each group, the upper figures are the model ('+') and target ('$\circ$') point sets, while the lower figures are our registration results, and the degradation level increases from left to right. From top to bottom, different degradations involving deformation, noise and occlusion.

multiplication operation on the $M \times M$ graph Laplacian matrix $\mathbf{A}$. As K is a constant which is not dependent on M and $K \ll M$, the total time complexity of solving the transformation $\mathcal{T}$ in our fast implementation can be written as $O(M^2)$. The space complexity of our method scales like $O(M^2)$ due to the memory requirements for storing the graph Laplacian matrix $\mathbf{A}$.

Implementation details

Before solving the registration problem, we first normalize the input point sets with a linear scaling, so that they are expressed in the same coordinate system, more specifically, the points in both of the two sets both have zero mean and unit variance. Besides, we solve a displacement function $\mathbf{v}$ defined as $\mathcal{T}(\mathbf{x}) = \mathbf{x} + \mathbf{v}(\mathbf{x})$ rather than directly solving the transformation $\mathcal{T}$, which can be achieved by directly setting the output as $\mathbf{y} - \mathbf{x}$ in our formulation. The use of displacement field guarantees more robustness (Myronenko and Song 2010; Ma et al. 2014).

Parameter settings. There are four main parameters in our MR-RPM: ϵ, β, λ_1 and λ_2. Parameter ϵ is a threshold used to construct the neighborhood graph. Parameter β determines the width of the range of the interaction between samples. The rest two parameters control the trade-off between the closeness to the data and the smoothness of the solution, where λ_1 and λ_2 regularize with respect to the whole input space and the intrinsic geometry, respectively. In general, we found that our method was robust to parameter changes. We set $\epsilon = 0.05, \beta = 0.1, \lambda_1 = 3, \lambda_2 = 0.05$ throughout this paper. The inlier percentage parameter γ needs an initial assumption, as shown in Line 8 in Algorithm 1, here we fix it to 0.9. Moreover, to use the fast implementation, we set the solution base number K to 15 in the 2D case and 50 in the 3D case; the uniform distribution parameter a is set to be the volume of the bounding box of the data.

Experimental results

In order to evaluate the performance of our MR-RPM, we conduct experiments on both 2D shape contour and 3D point cloud. The experiments were performed on a laptop with 3.0 GHz Intel Core CPU, 8 GB memory and Matlab Code.

Results on 2D shape contour

We use the same synthesized data as that in (Chui and Rangarajan 2003) and (Zheng and Doermann 2006), which consists of two shape patterns (i.e., a fish and a Chinese character) with different kinds of degenerations including deformation, noise, outlier, rotation and occlusion. For each kind of degeneration, it involves several different degeneration levels and each degeneration level contains 100 samples. Note that the outlier is somewhat similar to the occlusion, as in both cases one point set contains some points not contained in the other point set. But for the test data, the occlusion is more practical as the non-common points come from the shape contour while in the outlier case they are randomly spread over the shape patterns. Besides, the rotation could be well addressed by using rotation invariant feature descriptors. Therefore, we only test our method on three kinds of degenerations such as the deformation, noise and occlusion.

We first provide some qualitative illustrations of our MR-RPM on the two shape patterns, as shown in Fig. 1. For each

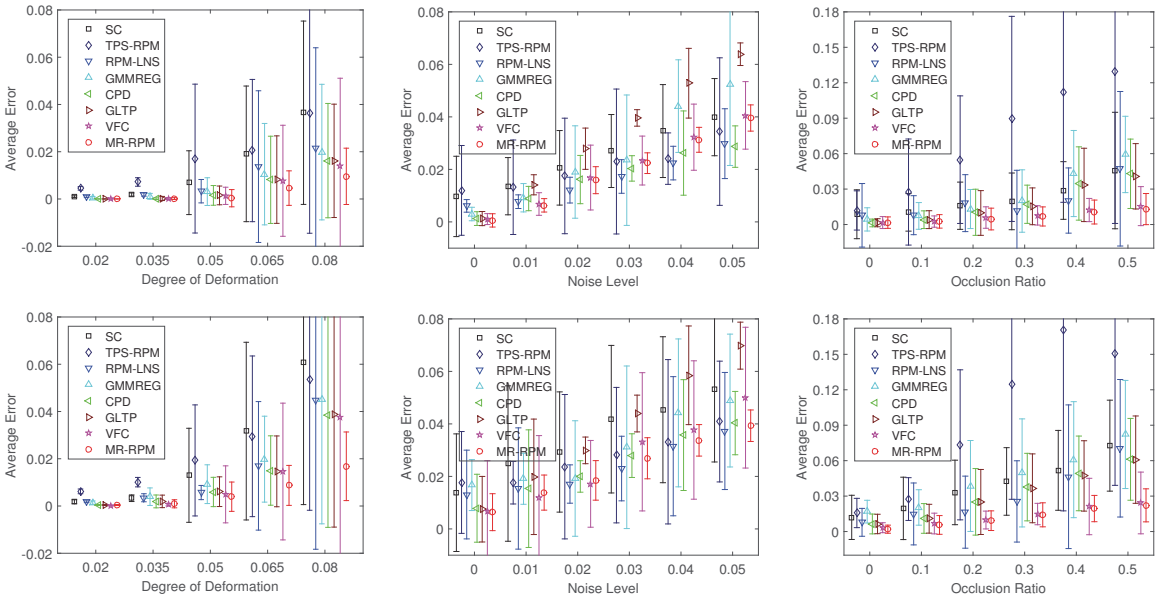

Figure 2: Comparison of MR-RPM with SC, TPS-RPM, RPM-LNS, GMMREG, CPD and VFC on the fish (top) and Chinese character (bottom). The error bars indicate the registration error means and standard deviations over 100 trials.

group of results, our goal is to align the model points ('+') onto the target points ('○') which are both presented in the upper figures, and the registration results are given in the lower figures. From the results, we see that our MR-RPM can well address all the different degradations, and the registration accuracy decreases gradually and gracefully as the degradation level increases. It is interesting that even in case of extreme degradation level, especially for the deformation and occlusion, our method can still generate satisfactory results. The average run time of our MR-RPM on this dataset with about 100 points for each shape pattern is about 0.5s.

To provide a quantitative comparison to the state-of-the-arts, we report the results of six methods such as SC (Belongie, Malik, and Puzicha 2002), TPS-RPM (Chui and Rangarajan 2003), RPM-LNS (Zheng and Doermann 2006), GMMREG (Jian and Vemuri 2011), CPD (Myronenko and Song 2010), GLTP (Ge, Fan, and Ding 2014), and VFC (Ma et al. 2014), as shown in Fig. 2. For each pair of shapes, the registration error is characterized by the average Euclidean distance between the warped model set and its corresponding target set. Then the mean and standard deviation of the registration error on all 100 samples for each degradation level and degradation type are computed for performance comparison. From the results, we see that SC, GMMRGE and GLTP are not robust to noise, while TPS-RPM degrades badly in case of occlusion. The registration accuracies of RPM-LNS and CPD are satisfactory which decrease gracefully as the degradation level increases. In contrast, VFC and our MR-RPM have the best results in most case expect for large noise level, and our MR-RPM almost consistently outperforms VFC for both different degradation type and degradation level on all the dataset. Note that a major difference of our MR-RPM and other iterative methods such as VFC is that we introduce an additional manifold regularization term; our consistently better results demonstrate that the manifold regularization does play an important role for improving the transformation estimation.

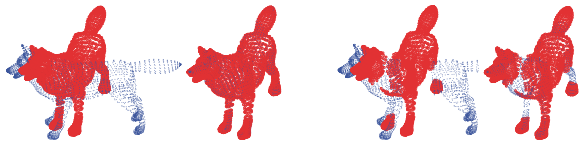

Figure 3: Qualitative results of our method on 3D wolf point clouds involving non-rigid deformation (left two) and occlusion (right two). For each group, the left figure is the model ('·') and target ('○') point sets, and the right is our result.

Results on 3D point cloud

We next test our MR-RPM for registration of 3D point cloud data, where a wolf pattern with about $5,000$ points in different poses is used for evaluation. The results are presented in Fig. 3, where the left two figures test the non-rigid deformation and the right two figures test the occlusion. We see that our method can produce almost perfect alignments for both point cloud pair, even both the model and target sets suffer from degradations, as shown in occlusion test. The average runtime of our MR-RPM on this dataset is about 47s.

We also conduct a quantitative comparison with respect to two representative state-of-the-arts such as CPD and VFC. The average registration errors on the two point cloud pairs shown in Fig. 3 is CPD (0.82, 0.72), VFC (1.15, 1.01), and MR-RPM (0.78, 0.53), respectively. Clearly, our method has the best performance, which means that our MR-RPM is ef-

fective for both 2D and 3D point set registration.

Conclusion

Within this paper, we presented a new method called MR-RPM for non-rigid registration of both 2D shapes and 3D point clouds. A key characteristic of our approach is the using of manifold regularization to capture the underlying intrinsic geometry of the input data, leading to a better estimate of the transformation. We also provide a fast implementation to reduce the algorithm complexity from cubic to quadratic, so that large scale data (especially 3D point cloud) could be addressed. The qualitative and quantitative results on both 2D and 3D public available datasets demonstrate that our MR-RPM outperforms the state-of-the-art methods in most cases, especially when there are significant non-rigid deformations and or occlusions in the data.

Acknowledgments

The authors gratefully acknowledge the financial supports from the National Natural Science Foundation of China under Grant nos. 61503288, 61501413 and 41501505.

References

Aronszajn, N. 1950. Theory of reproducing kernels. *Trans. Amer. Math. Soc.* 68(3):337–404.

Bai, S.; Bai, X.; Tian, Q.; and Latecki, L. J. 2017. Regularized diffusion process for visual retrieval. In *AAAI*.

Belkin, M.; Niyogi, P.; and Sindhwani, V. 2006. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *J. Mach. Learn. Res.* 7:2399–2434.

Belongie, S.; Malik, J.; and Puzicha, J. 2002. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.* 24(24):509–522.

Besl, P. J., and McKay, N. D. 1992. A method for registration of 3-d shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* 14(2):239–256.

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

Brown, L. G. 1992. A survey of image registration techniques. *ACM Comput. Surv.* 24(4):325–376.

Chui, H., and Rangarajan, A. 2003. A new point matching algorithm for non-rigid registration. *Comput. Vis. Image Understand.* 89:114–141.

Dempster, A.; Laird, N.; and Rubin, D. B. 1977. Maximum likelihood from incomplete data via the em algorithm. *J. R. Statist. Soc. Series B* 39(1):1–38.

Ge, S., and Fan, G. 2015. Articulated non-rigid point set registration for human pose estimation from 3d sensors. *Sensors* 15(7):15218–15245.

Ge, S.; Fan, G.; and Ding, M. 2014. Non-rigid point set registration with global-local topology preservation. In *CVPRW*, 245–251.

Horaud, R.; Forbes, F.; Yguel, M.; Dewaele, G.; and Zhang, J. 2011. Rigid and articulated point registration with expectation conditional maximization. *IEEE Trans. Pattern Anal. Mach. Intell.* 33(3):587–602.

Jian, B., and Vemuri, B. C. 2011. Robust point set registration using gaussian mixture models. *IEEE Trans. Pattern Anal. Mach. Intell.* 33(8):1633–1645.

Ling, H., and Jacobs, D. W. 2007. Shape classification using the inner-distance. *IEEE Trans. Pattern Anal. Mach. Intell.* 29(2):286–299.

Ma, J.; Zhao, J.; Tian, J.; Tu, Z.; and Yuille, A. 2013. Robust estimation of nonrigid transformation for point set registration. In *CVPR*, 2147–2154.

Ma, J.; Zhao, J.; Tian, J.; Yuille, A. L.; and Tu, Z. 2014. Robust point matching via vector field consensus. *IEEE Trans. Image Process.* 23(4):1706–1721.

Ma, J.; Qiu, W.; Zhao, J.; Ma, Y.; Yuille, A. L.; and Tu, Z. 2015a. Robust L_2E estimation of transformation for non-rigid registration. *IEEE Trans. Signal Process.* 63(5):1115–1129.

Ma, J.; Zhao, J.; Ma, Y.; and Tian, J. 2015b. Non-rigid visible and infrared face registration via regularized gaussian fields criterion. *Pattern Recognit.* 48(3):772–784.

Ma, J.; Zhou, H.; Zhao, J.; Gao, Y.; Jiang, J.; and Tian, J. 2015c. Robust feature matching for remote sensing image registration via locally linear transforming. *IEEE Trans. Geosci. Remote Sens.* 53(12):6469–6481.

Ma, J.; Zhao, J.; and Yuille, A. L. 2016. Non-rigid point set registration by preserving global and local structures. *IEEE Trans. Image Process.* 25(1):53–64.

Micchelli, C. A., and Pontil, M. 2005. On learning vector-valued functions. *Neural Comput.* 17(1):177–204.

Minh, H. Q., and Sindhwani, V. 2011. Vector-valued manifold regularization. In *ICML*, 57–64.

Myronenko, A., and Song, X. 2010. Point set registration: Coherent point drift. *IEEE Trans. Pattern Anal. Mach. Intell.* 32(12):2262–2275.

Papadimitriou, C. H., and Steiglitz, K. 1982. *Combinatorial optimization: algorithms and complexity*. Courier Corporation.

Poggio, T., and Girosi, F. 1990. Networks for approximation and learning. *Proc. IEEE* 78(9):1481–1497.

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

Rusu, R. B.; Blodow, N.; and Beetz, M. 2009. Fast point feature histograms (FPFH) for 3d registration. In *ICRA*, 3212–3217.

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

Wang, G.; Wang, Z.; Chen, Y.; and Zhao, W. 2015. A robust non-rigid point set registration method based on asymmetric gaussian representation. *Comput. Vis. Image Understand.* 141:67–80.

Wang, G.; Wang, Z.; Chen, Y.; Zhou, Q.; and Zhao, W. 2016. Context-aware gaussian fields for non-rigid point set registration. In *CVPR*, 5811–5819.

Zhao, J.; Ma, J.; Tian, J.; Ma, J.; and Zhang, D. 2011. A robust method for vector field learning with application to mismatch removing. In *CVPR*, 2977–2984.

Zhao, M.; Chow, T. W.; Wu, Z.; Zhang, Z.; and Li, B. 2015. Learning from normalized local and global discriminative information for semi-supervised regression and dimensionality reduction. *Inf. Sci.* 324:286–309.

Zheng, Y., and Doermann, D. 2006. Robust point matching for nonrigid shapes by preserving local neighborhood structures. *IEEE Trans. Pattern Anal. Mach. Intell.* 28(4):643–649.

Zhou, Y.; Bai, X.; Liu, W.; and Latecki, L. J. 2016. Similarity fusion for visual tracking. *Int. J. Comput. Vis.* 1–27.