

Non-rigid point set registration via global and local constraints

Changcai Yang¹ · Meifang Zhang² · Zejun Zhang³ ·
Lifang Wei³ · Riqing Chen³ · Huabing Zhou⁴

Received: 16 September 2017 / Revised: 26 April 2018 / Accepted: 23 May 2018 /
Published online: 9 June 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract Non-rigid point set registration is often encountered in medical image processing, pattern recognition, and computer vision. This paper presents a new method for non-rigid point set registration that can be used to recover the underlying coherent spatial mapping (CSM). Firstly, putative correspondences between two point sets are established by using feature descriptors. Secondly, each point is expressed as a weighted sum of several nearest neighbors and the same relation holds after the transformation. Then, this local geometrical constraint is combined with the global model, and the transformation problem is solved

✉ Riqing Chen
riqing.chen@fafu.edu.cn

Changcai Yang
changcaiyang@gmail.com

Meifang Zhang
zjzhang_fafu@163.com

Zejun Zhang
mfzhang85@163.com

Lifang Wei
weilifang0028@fafu.edu.cn

Huabing Zhou
zhouhuabing@gmail.com

¹ Digital Fujian Research Institute of Big Data for Agriculture and Forestry, College of Computer and Information Sciences, Fujian Agriculture and Forestry University, Fuzhou 350002, China

² Fujian Health College, Fuzhou 350101, China

³ College of Computer and Information Sciences, Fujian Agriculture and Forestry University, Fuzhou 350002, China

⁴ Wuhan Institute of Technology, Wuhan 430073, China

by minimizing an error function. These two steps of recovering point correspondences and transformation are performed iteratively to obtain a promising result. Extensive experiments on various synthetic and real data demonstrate that the proposed approach is robust and outperforms the state-of-the-art methods.

Keywords Point set registration · Coherent spatial mapping · Local geometrical constraint

1 Introduction

Point set registration is a fundamental and challenging problem which has a wide range of application in pattern recognition, medical image analysis, and computer vision [2, 4, 8, 10, 25, 27, 28]. Many tasks in these fields, including image fusing, shape matching, motion tracking, stereo correspondence, and image mosaic, can be typically formulated as point set registration problem, since point representations is simplest form of salient structure [4]. With a shape contour represented by the edge points, or the point feature in a image represented by the locations of interest points, the registration problem is to find the correct correspondence between two sets of points extracted from the input data and recover the underlying spatial mapping.

In order to match two point set accurately a small target, a variety of approaches have been proposed for point set registration in the past few decades. These methods can be classified into rigid or non-rigid registration, and the choice of them depends on the application and different characteristics of the data. For rigid transformation includes only for scaling, rotation, and translation, rigid registration is relatively easy and has reached some maturity [2, 4, 11]. By contrast, solving the non-rigid registration problem is quite difficult, because the underlying non-rigid spatial mappings, which allow anisotropic skews and scaling, are usually complex, unknown, and hard to model [8]. However, non-rigid registration is an issue of critical importance since it have been used in many real world tasks including: medical image mosaic, shape recognition, hand-written character recognition, and deformable motion tracking.

In this paper, we present a effective algorithm for non-rigid case. To solve this problem, we have to solve two unknown variables: the correspondence and the underlying spatial mapping. Because solving for either variable without information regarding the other tends to be difficult to handle, an iterated optimization method can be used [6, 35]. In this iterative process, the estimated transformation is used to refine the estimated correspondence, and vice versa. Their performance degenerates quickly if there are errors in the correspondence which occurs in many applications particularly if there are a significant number of outliers in the data and/or the transformation is large [6, 20].

Many algorithms have been proposed to tackle the problem. The most popular point registration approach is the iterated closest point (ICP) algorithm [2] proposed by Besl and McKay. They iteratively assign a binary correspondence based on the nearest-neighbor relationships and refine the rigid transformation by using the estimated correspondence. The main drawback of this approach is that the correspondence is binary makes a lot of local minimum. Its performance will degrade badly in the presence of outliers. Chui and Rangarajan [8] proposed a general framework for non-rigid point matching based on the optimal correspondences between two points sets found by with soft-assignment and the non-rigid spatial mapping modeled as thin-plate spline (TPS). The deterministic annealing technique

is used to jointly solve the soft-assignment and the non-rigid spatial mapping. This method is quite similar to the EM algorithm and perform better than ICP. In related work, a rich descriptor called shape context (SC) for point set registration was proposed by Belongie et al. [1]. They attach a descriptor, which captures the distribution of the remaining points relative to it, to each point. The correspondence problem is solved as a binary linear assignment problem. But these methods ignore robustness when they recover the transformation from the correspondence.

To address the robust issue, Lian and Zhang [17] proposed to reduce the energy function of the robust point matching (RPM) method [8] to a concave function after eliminating the transformation variables and use concave optimization technique to minimize the resulting energy function. Ma et al. [19, 21] proposed to solve for correspondence by interpolating a vector field between the two point sets. Alternatively, the coherence point drift (CPD) algorithm [29] uses probability density estimation to find the alignment of two point sets, and it was shown to be robust and accurate performance with respect to outliers, noise and missing points. Other related work includes the multi-Layer motion based method [22], Gaussian mixture model based method [15], as well as Locally Linear Transforming based method [23, 24].

The main contribution of our approach is to robustly estimate the transformations from the contaminated correspondences. More precisely, the underlying transformation is recovered by using the global and local geometrical constraint between a set of point correspondences. Our approach jointly estimates the point correspondences and transformation between two point sets. Firstly, we use feature descriptors such as shape context to find correspondence between two point sets. Secondly, we estimate a coherent spatial mapping fitting for the inlier as the transformation. To this end, we preserve the local geometry of inlier point after transformation. Moreover, the spatial transformation is parameterized by the thin-plate spline (TPS) [30].

The rest of the paper is organized as follows. Section 2 describes the proposed GLC method for estimating TPS from point correspondences with global and local geometry constraints. In Section 3, we demonstrate our experimental results on various data. Section 4 concludes this paper.

2 Estimating coherent spatial mapping from correspondence

In this section, we consider the global spatial relationship and local geometrical constraint between a set of point correspondences, and propose an efficient point matching method which is able to establish accurate correspondences without hurting the correct matches.

2.1 Problem formulation

Given a set of point correspondences $S = \{(\mathbf{x}_n, \mathbf{y}_n)\}_{n=1}^N$ which are typically perturbed by noise and by outliers which undergo different transformations, our goal is to distinguish inliers from the outliers by estimating a coherent spatial mapping \mathbf{f} for the underlying inliers, i.e., for an inlier correspondence $(\mathbf{x}_n, \mathbf{y}_n)$, $\mathbf{y}_n = \mathbf{f}(\mathbf{x}_n)$.

Due to the existence of noise and outliers, it is desirable to have a robust estimation of \mathbf{f} . To make this problem tractable, we assume that, the noise on inliers is isotropic Gaussian

with zero mean and uniform standard deviation σ ; the outlier distribution is to be uniform $\frac{1}{a}$ with being a constant. Suppose there are two data in \mathfrak{D} , the observed input set $\mathbf{X} = \{\mathbf{x}_n, n = 1, \dots, N\}$ with point $\mathbf{x}_n = (\mathbf{x}_n^x, \mathbf{x}_n^y, 1)$, and the observed output set $\mathbf{y} = \{\mathbf{y}_n, n = 1, \dots, N\}$ with point $\mathbf{y}_n = (\mathbf{y}_n^x, \mathbf{y}_n^y, 1)$. Thus, the likelihood is a mixture model given by

$$\begin{aligned}
 p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}) &= \prod_{n=1}^N \sum_{z_n} p(\mathbf{y}_n, z_n | \mathbf{x}_n, \boldsymbol{\theta}) \\
 &= \prod_{n=1}^N \left(\frac{\gamma}{(2\pi\sigma^2)^{D/2}} e^{-\frac{\|\mathbf{y}_n - \mathbf{f}(\mathbf{x}_n)\|^2}{2\sigma^2}} + \frac{1 - \gamma}{a} \right), \tag{1}
 \end{aligned}$$

where $\boldsymbol{\theta} = \{\mathbf{f}, \sigma^2, \gamma\}$ includes a set of unknown parameters, and γ is the mixing coefficient which specifies the marginal distribution over the latent variable, i.e., $\forall z_n, p(z_n = 1) = \gamma$. $z_n \in \{0, 1\}$ is a latent variable of the n -th sample, where $z_n = 1$ indicates a Gaussian distribution and $z_n = 0$ points to a uniform distribution.

It is worthy nothing that true parameter set $\boldsymbol{\theta}$ can be estimated by maximizing likelihood (1). A maximum likelihood estimation of $\boldsymbol{\theta}$, i.e. $\boldsymbol{\theta}^* = \arg \max_{\boldsymbol{\theta}} p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta})$, can be obtained by seeking the minimal energy

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

The coherent spatial mapping \mathbf{f} will be directly obtained from the optimal solution $\boldsymbol{\theta}^*$.

In this work, the well known EM algorithm [9] is used to address the problem. It alternates with two steps: an expectation step (E-step) and a maximization step (M-step). The EM algorithm jointly estimates the responsibility indicating to what degree a correspondence belonging to inlier under the given transformation \mathbf{f} and updates \mathbf{f} based on the current estimate of the responsibility. Following standard notations [3], we omit some terms that are independent of $\boldsymbol{\theta}$. Considering the negative log likelihood function, i.e. (2), the complete-data log likelihood takes the following form:

$$\begin{aligned}
 \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) &= -\frac{1}{2\sigma^2} \sum_{n=1}^N p_n \|\mathbf{y}_n - \mathbf{f}(\mathbf{x}_n)\|^2 - \ln \sigma^2 \sum_{n=1}^N p_n \\
 &\quad + \ln \gamma \sum_{n=1}^N p_n + \ln(1 - \gamma) \sum_{n=1}^N (1 - p_n). \tag{3}
 \end{aligned}$$

where $p_n = P(z_n = 1 | \mathbf{x}_n, \mathbf{y}_n, \boldsymbol{\theta}^{\text{old}})$ is a soft decision, which indicates to what degree being inlier.

E-step: Denote $\mathbf{P} = \text{diag}(p_1, \dots, p_N)$ as a diagonal matrix. p_n can be obtained based on Bayes rule, i.e.,

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

M-step: Reestimate the parameter set θ^{new} as follows: $\theta^{\text{new}} = \arg \max_{\theta} \mathcal{Q}(\theta, \theta^{\text{old}})$. The minimum solution σ^2 of $\mathcal{Q}(\theta)$ can be obtained by letting $\frac{\partial \mathcal{Q}(\theta)}{\partial \sigma^2} = 0$

$$\sigma^2 = \frac{(\mathbf{Y} - \mathbf{V})^T \mathbf{P} (\mathbf{Y} - \mathbf{V})}{2 \cdot \text{tr}(\mathbf{P})}, \tag{5}$$

Similar to σ^2 , the minimum solution γ is

$$\gamma = \text{tr}(\mathbf{P})/N, \tag{6}$$

where $\mathbf{V} = (\mathbf{f}(\mathbf{x}_1); \dots; \mathbf{f}(\mathbf{x}_N))$, and $\text{tr}(\cdot)$ is the trace.

After the EM iteration converges, a correspondence $(\mathbf{x}_n, \mathbf{y}_n)$ is considered as inlier if $p_n > \tau$, otherwise it is a outlier, where τ is a predefined threshold. It is worth nothing that the smaller the threshold τ , the higher the outlier rate and registration error are.

2.2 Global and local regularizations

Next we consider the minimization with respect to \mathbf{f} . According to the complete negative log-likelihood (3), the mapping \mathbf{f} is estimated by minimizing a weighted empirical error

$$\mathcal{Q}(\mathbf{f}) = \frac{1}{2\sigma^2} \sum_{n=1}^N p_n \|\mathbf{y}_n - \mathbf{f}(\mathbf{x}_n)\|^2. \tag{7}$$

This is in general ill-posed since \mathbf{f} is not unique. To generate a smooth mapping fitting for the point correspondences, we adopt the TPS for parametrization. The TPS is chosen because it is a general purpose spline tool which can be cleanly decomposed into affine and non-affine subspaces controlled by coefficients \mathbf{A} and \mathbf{W} respectively [30]:

$$\mathbf{f}(\mathbf{x}) = \mathbf{x} \cdot \mathbf{A} + \tilde{\mathbf{K}}(\mathbf{x}) \cdot \mathbf{W}, \tag{8}$$

where $\tilde{\mathbf{K}}(\mathbf{x})$ is a $1 \times N$ vector defined by the TPS kernel, i.e. $K(r) = r^2 \log r$, and each entry $\tilde{K}_n(\mathbf{x}) = K(|\mathbf{x} - \mathbf{x}_n|)$. Define the kernel matrix $\mathbf{K}_{N \times N} = \{K_{ij}\}$ where $K_{ij} = K(|\mathbf{x}_i - \mathbf{x}_j|)$.

To have a stable solution, a typical strategy is to impose regularization on f to control the complexity of the hypothesis space. Here, we consider the standard TPS regularization which is given by

$$\phi(\mathbf{W}) = \text{tr}(\mathbf{W}^T \mathbf{K} \mathbf{W}) \tag{9}$$

which is the bending energy, and is independent on the linear component of the coherent spatial mapping.

The TPS regularization is a global constraint, and it is useful to keep the overall spatial connectivity of the point correspondences during registration. However, for point set registration, the local structures among neighboring points are also very strong and stable. This is particularly beneficial when the motion between the two point sets is nonrigid. In this case, a local geometrical constraint is desired to establish accurate correspondences. We preserve the local neighborhood structure during registration.

In our problem, we hope that the local structures in \mathbf{Y} could be preserved after the displacement of \mathbf{X} . Each point in \mathbf{X} can be reconstructed by a weighted sum of the M nearest neighbors

$$\mathbf{x}_i = \sum_{j=1}^N \alpha_{ij} \mathbf{x}_j, \text{ with } \sum_{j=1}^N \alpha_{ij} = 1, \tag{10}$$

where the α_{ij} are homogeneous barycentric coordinates, and enforce $\alpha_{ij} = 0$ if \mathbf{x}_j does not belong to the set of neighbors of \mathbf{x}_i .

The same relation holds after the transformation \mathbf{f} and we can also write

$$\mathbf{f}(\mathbf{x}_i) = \sum_{j=1}^N \alpha_{ij} \mathbf{f}(\mathbf{x}_j). \tag{11}$$

We then write the local geometry term as follows:

$$\mathcal{E}(\alpha) = \sum_{i=1}^N l \|\mathbf{f}(\mathbf{x}_i) - \sum_{j=1}^N \alpha_{ij} \mathbf{f}(\mathbf{x}_j)\|^2 \tag{12}$$

To preserve the local geometry of each inlier point after the transformation \mathbf{f} , we combine TPS regularization term in (9) and the local constraint in (12). This lets us express the mapping \mathbf{f} estimation as the following minimizing problem

$$\Psi(\mathbf{f}) = \frac{1}{2\sigma^2} \sum_{n=1}^N p_n \|\mathbf{y}_n - \mathbf{f}(\mathbf{x}_n)\|^2 + \frac{\lambda^L}{2} \sum_{i=1}^N p_i \|\mathbf{f}(\mathbf{x}_i) - \sum_{j=1}^N \alpha_{ij} \mathbf{f}(\mathbf{x}_j)\|^2 + \frac{\lambda^G}{2} \text{tr}(\mathbf{W}^T \mathbf{K} \mathbf{W}). \tag{13}$$

where the positive real numbers λ^G and λ^L control the tradeoff between data fitting and global/local geometric preserving.

If we substitute (8) into the expression (13), the error function can be written in the form

$$\begin{aligned} \Psi(\mathbf{f}) &= \frac{1}{2\sigma^2} \sum_{n=1}^N p_n \|\mathbf{y}_n - \mathbf{x}_n \mathbf{A} - \tilde{K}(\mathbf{x}_n) \mathbf{W}\|^2 \\ &\quad + \frac{\lambda^L}{2} \sum_{i=1}^N p_i \|\mathbf{x}_i - \sum_{j=1}^N \alpha_{ij} \mathbf{x}_j\|^2 + (\tilde{K}_i - \sum_{j=1}^N \alpha_{ij} \tilde{K}_j) \mathbf{W}\|^2 \\ &\quad + \frac{\lambda^G}{2} \text{tr}(\mathbf{W}^T \mathbf{K} \mathbf{W}) \\ &= \frac{1}{2\sigma^2} \sum_{n=1}^N p_n \|\mathbf{y}_n - \mathbf{x}_n \mathbf{A} - \tilde{K}(\mathbf{x}_n) \mathbf{W}\|^2 + \frac{\lambda^L}{2} \sum_{i=1}^N p_i \|\mathbf{x}_i^L \mathbf{A} + K_i^L \mathbf{W}\|^2 \\ &\quad + \frac{\lambda^G}{2} \text{tr}(\mathbf{W}^T \mathbf{K} \mathbf{W}). \end{aligned} \tag{14}$$

where $\mathbf{x}_i^L = \mathbf{x}_i - \sum_{j=1}^N \alpha_{ij} \mathbf{x}_j$ and $K_i^L = \tilde{K}_i - \sum_{j=1}^N \alpha_{ij} \tilde{K}_j$.

It is worthy noting that the (14) can be rewritten as follows

$$\mathcal{E}(\mathbf{A}, \mathbf{W}) = \frac{1}{2\sigma^2} \|\tilde{\mathbf{Y}} - \tilde{\mathbf{X}} \mathbf{A} - \mathbf{P}^{1/2} \mathbf{K} \mathbf{W}\|^2 + \frac{\lambda^L}{2} \|\tilde{\mathbf{X}}^L \mathbf{A} + \mathbf{P}^{L/2} \mathbf{K}^L \mathbf{W}\|^2 + \frac{\lambda^G}{2} \text{tr}(\mathbf{W}^T \mathbf{K} \mathbf{W}). \tag{15}$$

where $\tilde{\mathbf{X}} = \mathbf{P}^{1/2} \mathbf{X}$, $\tilde{\mathbf{Y}} = \mathbf{P}^{1/2} \mathbf{Y}$, and $\tilde{\mathbf{X}}^L = \mathbf{P}^{L/2} \mathbf{X}^L$. The third smoothness term is the standard TPS regularization term. Thus, the local constraint term in (15) has the form $\frac{\lambda^L}{2} \|\tilde{\mathbf{X}}^L \mathbf{A} + \mathbf{P}^{L/2} \mathbf{K}^L \mathbf{W}\|^2$. As \mathbf{P}^L and \mathbf{K}^L are fixed during transformation estimation, the local constraint term then becomes $\frac{\lambda^L}{2} \|\tilde{\mathbf{X}}^L \mathbf{A} + \tilde{\mathbf{K}}^L \mathbf{W}\|^2$, where $\tilde{\mathbf{K}}^L = \mathbf{P}^{L/2} \mathbf{K}^L$ is an $N \times N$ constant matrix. In this context, the local constraint term plays a role of regularization on the transformation f , which ensures the well-posedness of the error function minimization, and it is controlled by a locally linear constraint (i.e., $\frac{\lambda^L}{2} \|\tilde{\mathbf{X}}^L \mathbf{A} + \tilde{\mathbf{K}}^L \mathbf{W}\|^2$). Figure 1 illustrates the schematic of the local geometrical constraint.

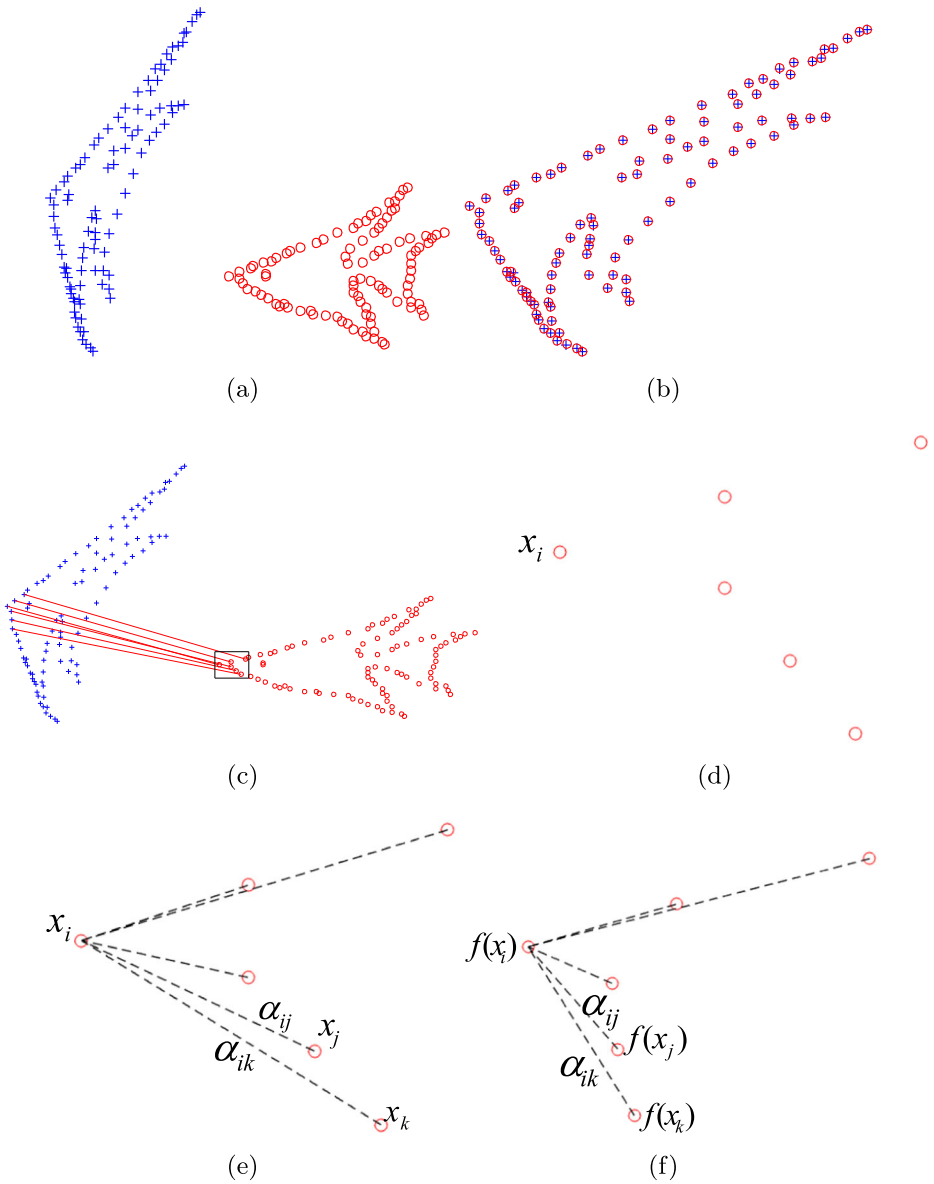


Fig. 1 Schematic of the local geometrical constraint. **a** Model point sets X (red circles) and target point sets Y (blue pluses). The goal is to align X onto Y. **b** Registration result by the proposed method. **c** Set of putative correspondences. **d** Assign neighbors to each point x_i , e.g., the five solid pluses around x_i . **e** Compute the weights W_{ij} that best linearly reconstruct x_i from its neighbors. **f** Optimize the transformation f under the constraint that each point x_i can be reconstructed by its neighbors with weights W_{ij} after the transformation

To solve the TPS parameter pair \mathbf{A} and \mathbf{W} , we use a QR decomposition [8, 30]

$$\tilde{\mathbf{X}} = [\mathbf{Q}_1 \ \mathbf{Q}_2] \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix}, \tilde{\mathbf{X}}^L = [\mathbf{Q}_1^L \ \mathbf{Q}_2^L] \begin{bmatrix} \mathbf{R}^L \\ \mathbf{0} \end{bmatrix}, \tag{16}$$

where \mathbf{Q}_1 and \mathbf{Q}_1^L are orthonormal matrices of sizes $N \times 3$, \mathbf{Q}_2 and \mathbf{Q}_2^L are orthonormal matrices of sizes $N \times (N - 3)$, and \mathbf{R} and \mathbf{R}^L are upper triangular matrices of size 3×3 . With the QR decomposition in place, equation (15) becomes

$$\begin{aligned} \mathcal{E}(\mathbf{A}, \Gamma) &= \frac{1}{2\sigma^2} \left(\|\mathbf{Q}_2^T \tilde{\mathbf{Y}} - \mathbf{Q}_2^T \mathbf{P}^{1/2} \mathbf{K} \mathbf{Q}_2 \Gamma\|^2 + \|\mathbf{Q}_1^T \tilde{\mathbf{Y}} - \mathbf{R} \mathbf{A} - \mathbf{Q}_1^T \mathbf{P}^{1/2} \mathbf{K} \mathbf{Q}_2 \Gamma\|^2 \right) \\ &+ \frac{\lambda^L}{2} \left(\|\mathbf{Q}_2^L \mathbf{P}^{L/2} \mathbf{K}^L \mathbf{Q}_2 \Gamma\|^2 + \|\mathbf{Q}_1^L \mathbf{P}^{L/2} \mathbf{K}^L \mathbf{Q}_2 \Gamma + \mathbf{R}^L \mathbf{A}\|^2 \right) \\ &+ \frac{\lambda^G}{2} \text{tr}(\Gamma^T \mathbf{Q}_2^T \mathbf{K} \mathbf{Q}_2 \Gamma), \end{aligned} \tag{17}$$

where $\mathbf{W} = \mathbf{Q}_2 \Gamma$ and Γ is a matrix of size $(N - 3) \times 3$. Setting $\mathbf{W} = \mathbf{Q}_2 \Gamma$ implies that $\tilde{\mathbf{X}}^T \mathbf{W} = 0$, which enables us to clearly separate the warping into affine and non-affine subspaces [8].

Minimizing the energy function (17) with respect to Γ and \mathbf{A} , we obtain

$$\begin{aligned} \mathbf{W} &= \mathbf{Q}_2 \Gamma \\ &= \mathbf{Q}_2 (\mathbf{S}^T \mathbf{S} + \lambda^L \sigma^2 \mathbf{S}^L \mathbf{T} \mathbf{S}^L + \lambda^G \sigma^2 \mathbf{T} + \tilde{\epsilon} \mathbf{I})^{-1} \mathbf{S}^T \mathbf{Q}_2^T \tilde{\mathbf{Y}}, \end{aligned} \tag{18}$$

$$\begin{aligned} \mathbf{A} &= (\mathbf{R}^T \mathbf{R} + \lambda^L \sigma^2 \mathbf{R}^L \mathbf{T} \mathbf{R}^L)^{-1} \left[\mathbf{R}^T \mathbf{Q}_1^T (\tilde{\mathbf{Y}} - \mathbf{P}^{1/2} \mathbf{K} \mathbf{W}) \right. \\ &\quad \left. - \mathbf{R}^L \mathbf{T} \mathbf{Q}_1^L \mathbf{P}^{L/2} \mathbf{K}^L \mathbf{W} \right]. \end{aligned} \tag{19}$$

where $\mathbf{S} = \mathbf{Q}_2^T \mathbf{P}^{1/2} \mathbf{K} \mathbf{Q}_2$, $\mathbf{S}^L = \mathbf{Q}_2^L \mathbf{P}^{L/2} \mathbf{K}^L \mathbf{Q}_2$, $\mathbf{T} = \mathbf{Q}_2^T \mathbf{K} \mathbf{Q}_2$, and $\tilde{\epsilon} \mathbf{I}$ is used for numerical stability. Thus we obtain the coherent spatial mapping \mathbf{f} in (8).

2.3 Non-rigid point set registration

Point set registration aims to align two point sets $\{\mathbf{x}_m\}_{m=1}^M$ (the model point set) and $\{\mathbf{y}_l\}_{l=1}^L$ (the target point set). Typically, in the non-rigid case, it requires estimating a non-rigid transformation \mathbf{f} which warps the model point set to the target point set. Our approach iteratively recovers the point correspondences and estimates the transformation between two point sets. In the first step of the iteration, feature descriptors are used to establish correspondence. We use the shape context (SC) [1] descriptor proposed by Belongie, which is a 60-dimensional vector, to offer a globally discriminative characterization. The χ^2 distance was used to measure the difference of two points, and then the Hungarian method [5] is applied to find the initial correspondences between $\{\mathbf{x}_m\}_{m=1}^M$ and $\{\mathbf{y}_l\}_{l=1}^L$. In the second step, we adopt the proposed method to estimate a coherent spatial mapping fitting for the inlier as the transformation.

The performance of point matching algorithms depends, typically, on the coordinate system in which points are expressed. We use data normalization to control for this. In our evaluation, we compute two similarity transformations T_x and T_y for the point sets $\{\mathbf{x}_n\}$ and $\{\mathbf{y}_n\}$, i.e. $\hat{\mathbf{x}}_n = T_x \mathbf{x}_n$, which ensure that point sets have zero means and average distance $\sqrt{2}$ to the origin [14]. The regularization parameters λ^G and λ^L control the trade-off between the closeness to the data and the smoothness of the mapping. We set them to 500 in throughout our experiments. Moreover, the uniform distribution parameter a is set to be 5. We summarize the method in Algorithm 1.

Algorithm 1 Non-rigid Point Set Registration

Input: Two point sets $\{\mathbf{x}_m\}_{m=1}^M$, $\{\mathbf{y}_l\}_{l=1}^L$, kernel K , and parameters λ^G , λ^L , a

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

- 1 Compute feature descriptors for the target point set $\{\mathbf{y}_l\}_{l=1}^L$;
- 2 **repeat**
- 3 Compute feature descriptors for the model point set $\{\mathbf{x}_m\}_{m=1}^M$;
- 4 Estimate the Putative correspondence set $S = \{\mathbf{x}_n, \mathbf{y}_n\}_{n=1}^N$ based on the feature descriptors of two point sets;
- 5 Normalization: $\hat{\mathbf{x}}_n = T_x \mathbf{x}_n$, $\hat{\mathbf{y}}_n = T_y \mathbf{y}_n$;
- 6 Construct kernel matrix \mathbf{K} using the definition of K ;
- 7 Initialization: $\gamma = 0.9$, $\mathbf{W} = \mathbf{0}$, $\mathbf{A} = \mathbf{0}$;
- 8 **repeat**
- 9 *E-step:*
- 10 Update the responsibility p_n by (4);
- 11 *M-step:*
- 12 Update mapping \mathbf{f} by using (18) and (19);
- 13 Update σ^2 and γ by (5);
- 14 **until** some stopping criterion is satisfied;
- 15 The spatial mapping \mathbf{f} is determined by (8);
- 16 **until** reach the maximum iteration number;
- 17 The aligned model point set $\{\hat{\mathbf{x}}_m\}_{m=1}^M$ is given by $T_y^{-1}\{\mathbf{f}(\mathbf{x}_m)\}_{m=1}^M$ in the last iteration.

3 Experimental results

In this section we compare our algorithm to other state-of-the-art approaches such as SC [1], TPS-RPM [8], CPD [29], and MR-RPM [26]. The stability and accuracy of registration methods is evaluated using the average Euclidean distance between the transformed model points and their template points.

3.1 Experiments with synthesized data

We first test our method on the synthesized data set by Chui and Rangarajan [8]. The data have two different shape models: a *fish* and a *Chinese character*. The fish consists of 96 points. The Chinese character consists of 108 points. For each model, there are three sets of data designed to measure the robustness of registration algorithms under deformation, occlusion, and rotation. In each test, one of the above distortions is applied to a model set to create a target set, and 100 samples are generated for each degradation level. Figure 2 shows examples from the synthetic data used to evaluate our approach. We use the shape context as the feature descriptor to establish initial correspondences. It is easy to make shape context translation and scale invariant, and in some applications, the rotation invariant shape context is used as in [35] if necessary.

Figure 3 illustrates the registration progress of the proposed method on fish point sets. The columns show the iterative alignment progress, and each row provides a different type of distortion. We align the model point set X (red circles) onto the observed data point set Y (blue pluses). From Fig. 3, we can see that our method is robust and accurate, and it typically converges in several iterations.



Fig. 2 Examples of synthetic data used in the experiments. Top row: the shapes of the fish. Bottom row: the shapes of the Chinese character. The left column: the template point sets. From the second to the right column: examples of the target point sets for the deformation, occlusion, and rotation tests, respectively

Figures 4 and 5 show the registration results of our method with comparison to four state-of-the-art methods: SC [1], TPS-RPM [8], CPD [29], and MR-RPM [26], which are implemented using publicly available codes. In the tests of rotation, we use the rotation

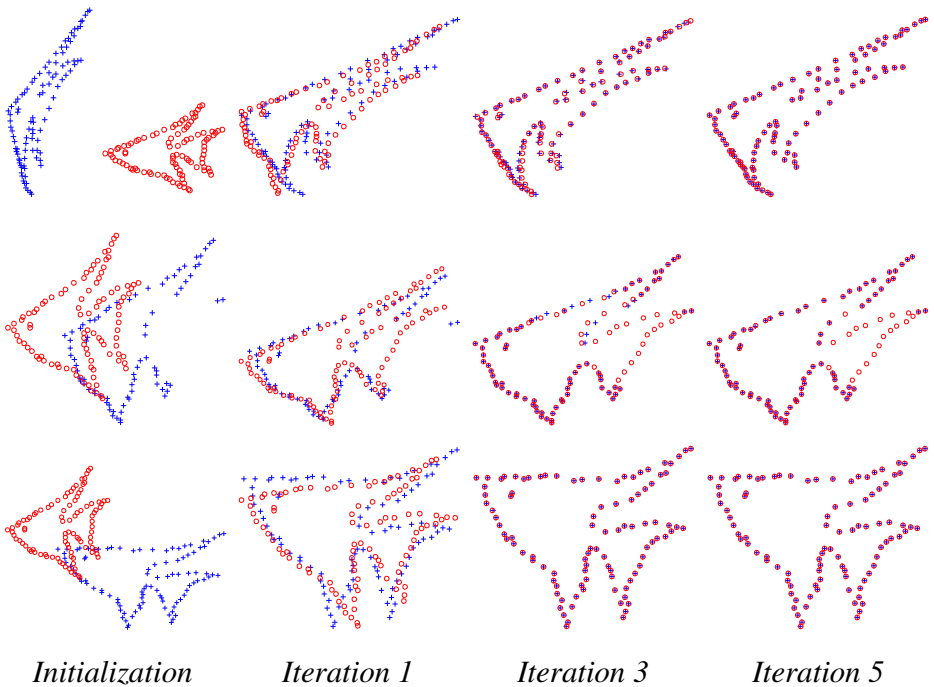


Fig. 3 Illustration of registration progress of our method on fish point sets. The goal is to align X (red circles) onto Y (blue pluses). From top to bottom: results on deformation, occlusion, and rotation tests, respectively. The columns show the iterative alignment progress

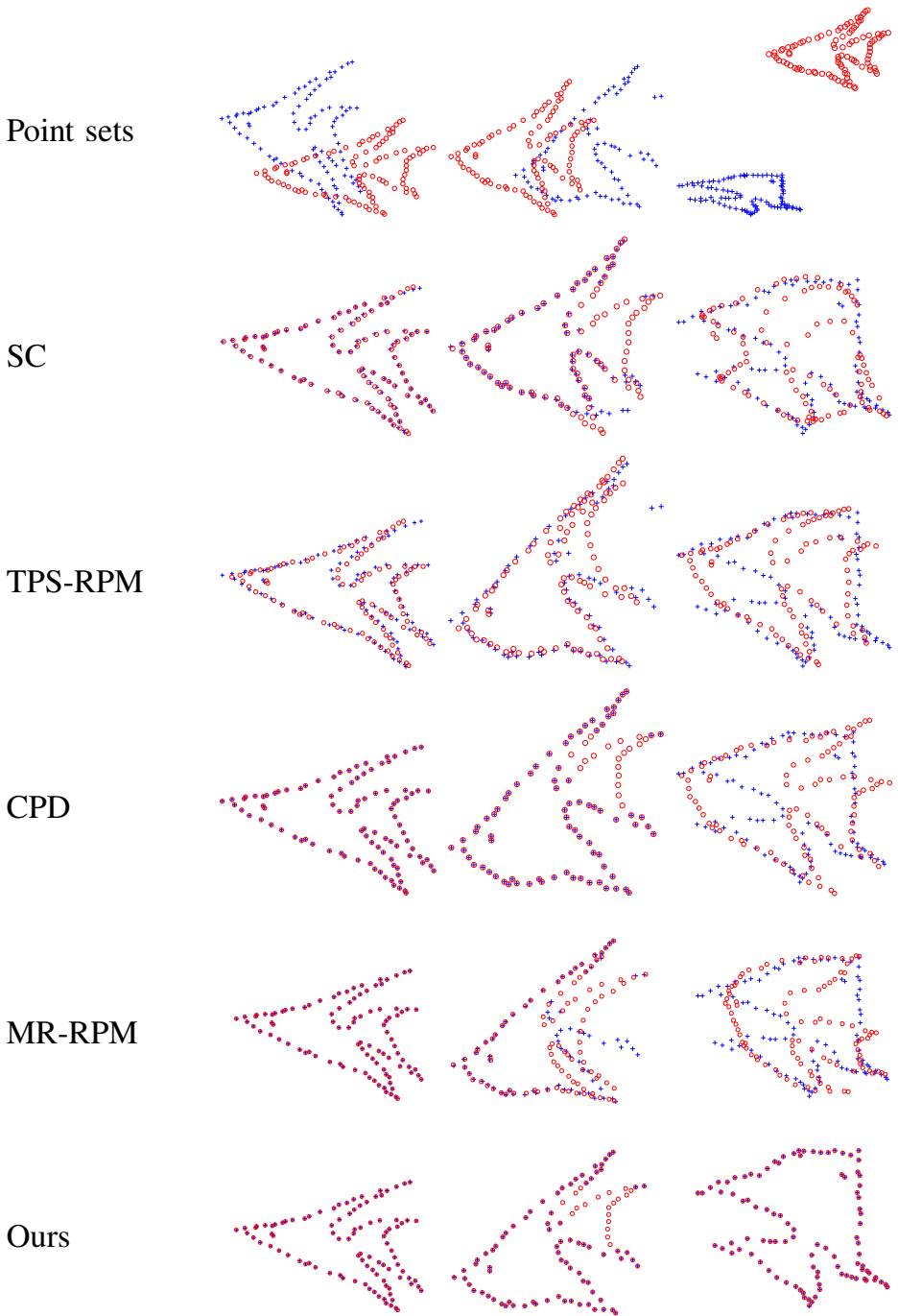


Fig. 4 Registration examples by different methods on the fish shape. The goal is to align the model point set (red circles) onto the target point set (blue pluses). Top row: template and target. From the second row to the bottom row: registration results by SC, TPS-RPM, CPD, MR-RPM, and our method, respectively. From left to right column: the deformation, occlusion, and rotation tests, respectively

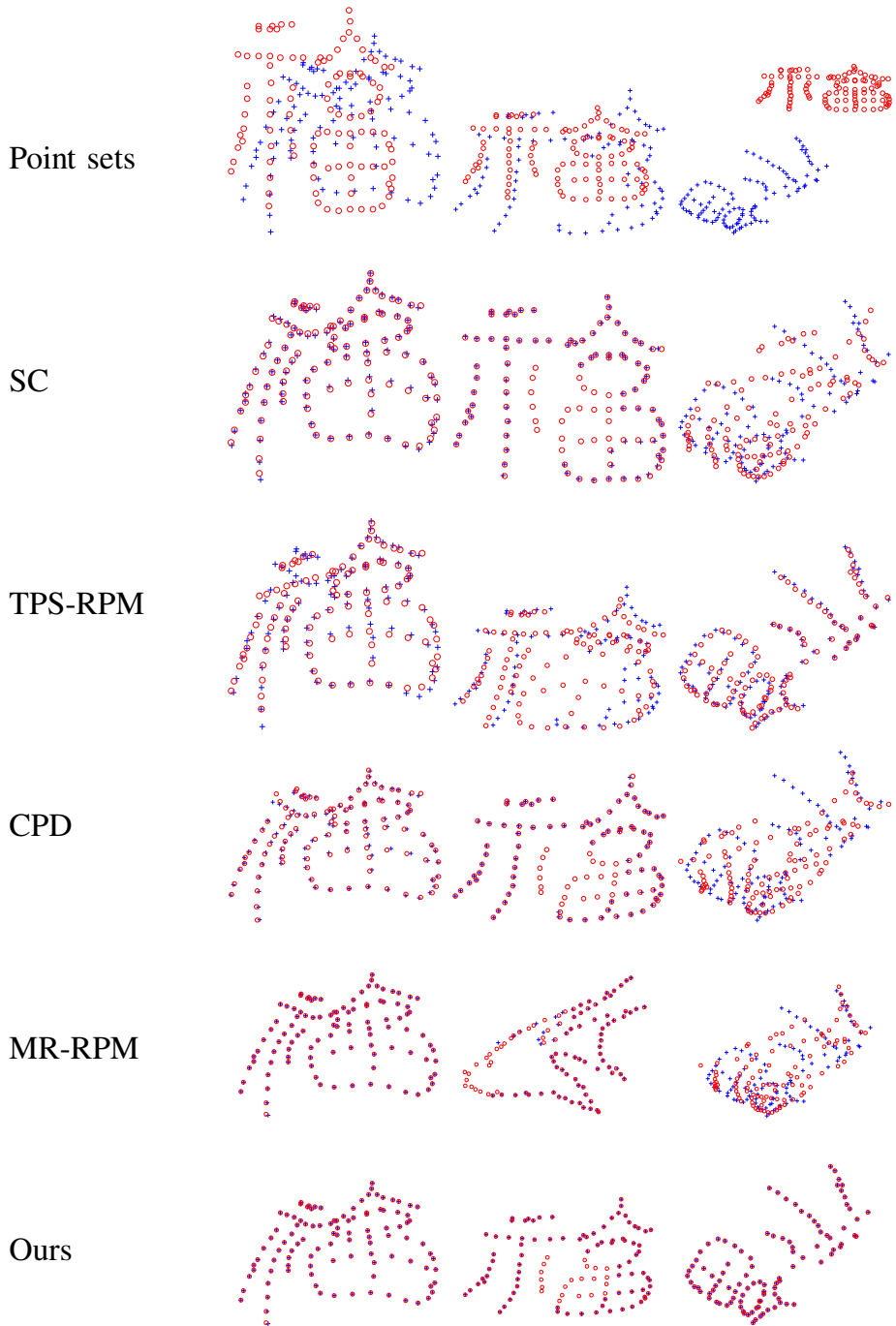


Fig. 5 Registration examples by different methods on the Chinese character shape. The goal is to align the model point set (red circles) onto the target point set (blue pluses). Top row: template and target. From the second row to the bottom row: registration results by SC, TPS-RPM, CPD, MR-RPM, and our method, respectively. From left to right column: the deformation, occlusion, and rotation tests, respectively

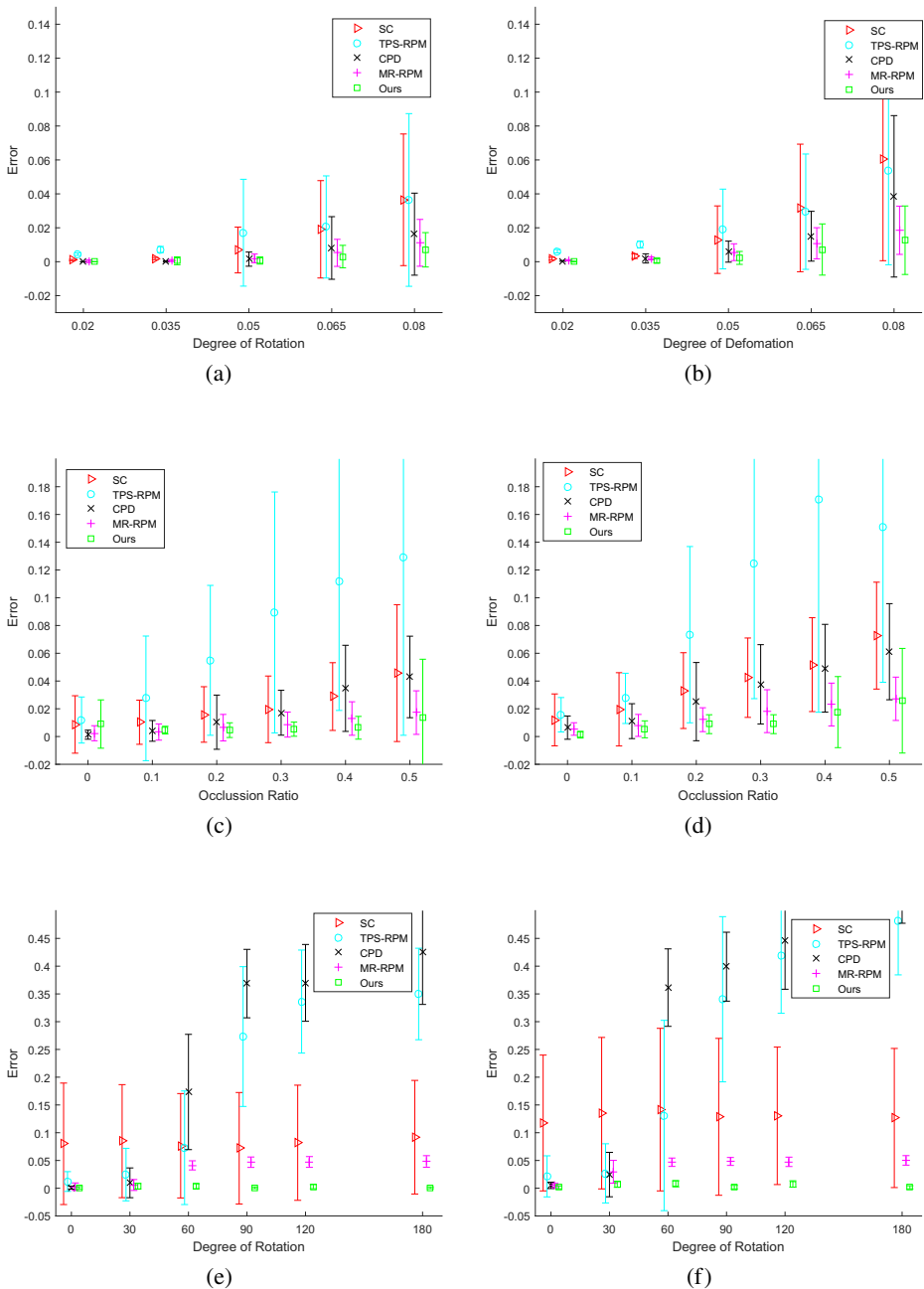


Fig. 6 Comparison of the registration performance of our method with SC, TPS-RPM, and CPD on the fish and Chinese character shapes [8]. Left column: the shape of fish. Right column: the shape of a Chinese character. The error bars indicate the registration error means and standard deviations over 100 trials



Fig. 7 The first sample in each category used as template

invariant shape context as the feature descriptor in our approach. As shown in the results, we see that SC and CPD can only generate satisfactory alignments for the shapes with deformation and occlusion, however, our method is able to produce an almost perfect alignment in all experiments. Note that the other four algorithms are not robust to large rotation, as shown in the right column of the figures. By contrast, our method is not affected by rotation since we use a rotation invariant feature descriptor.

We further provide a quantitative comparison on the two shape models, and we test on all the three sets. The registration error on a pair of shapes is quantified as the average Euclidean distance between a point in the warped model and the corresponding point in the target. Then the registration performance of each algorithm is compared by the mean and standard deviation of the registration error of all the 100 samples. The performance of the proposed method is compared with four state-of-the-art point matching methods, SC [1], TPS-RPM [8], CPD [29], and MR-RPM [26]. The statistical results for each setting are summarized in Fig. 6. The results for the shapes of fish and Chinese character with deformation are shown in Fig. 6a and b, respectively. For minor deformations, four methods behave similarly. But our method performs best, especially for large deformations. Figure 6c and d show the results for the shapes of fish and Chinese character with occlusion, respectively. The error means of SC, CPD, MR-RPM, and our approach are nearly horizontal, showing good robustness to occlusion. However, our approach performs best, and it consistently outperforms CPD. Figures 6e and f show the results for the shapes of fish and Chinese character with rotation, respectively. The best results are obtained by our method. These plots clearly show the sensitivity of the CPD, and TPS-RPM to the rotation.

3.2 Experiments with MNIST handwritten digit database

In this section, we perform experiments on the MNIST handwritten digit database [16]. The database consists 10 categories corresponding to the 10 digits (from 0 to 9), where each category has 1000 images. The size of each image is 28×28 . Figure 7 shows the ten templates. For each image, the 48 brightest points were sampled to represent a digit from the Canny edges. The first sample in each category is used as template and the rest images are used as the targets for registration test. The spatial mapping \mathbf{f} between these two sets of points is established by the proposed method. Then the estimated \mathbf{f} is used to warp the template shape.

We compare the proposed method with state-of-the-art point methods including SC [1], TPS-RPM [8], CPD [29], and MR-RPM [26]. The registration error of a pair of digits, which is quantified as the average Euclidean distance between a point in the warped template and

Table 1 Performance comparison on the MNIST database

	SC	TPS-RPM	CPD	MR-RPM	Ours
Average error	0.0595	0.0511	0.0833	0.0275	0.0172

Bold emphasis are the best results

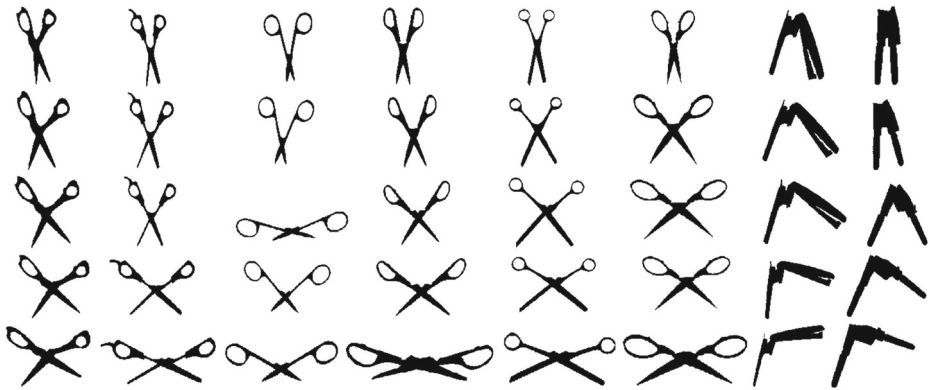


Fig. 8 Articulated shape database. The dataset contains 40 images from 8 articulated objects. Each column has 5 images from the same object

the corresponding point in the target, is used to quantify the accuracy of the registration methods. Table 1 shows the average error of the evaluated methods. As shown in Table 1, our method outperforms the other three algorithms. The average error of the proposed method is 0.0172 and is less than 0.0595, 0.0511, 0.0833, and 0.0275 respectively, provided by the SC, TPS-RPM, CPD, and MR-RPM methods.

3.3 Shape classification results

In this section we test our method on the articulated shape data set [18], which contains 40 images from 8 different objects with articulation, as shown in Fig. 8. Each object has 5 images articulated to different degrees. The dataset is extremely challenging due to the high similarity between different objects. We compare our method to SC [1], CPD [29], and MR-RPM [26]. In our evaluation, we first sampled two hundred points from the outer contours of every shape, and run each point matching method to find the correspondence between two shapes and use the correspondence to warp one of the shapes. We then use SC distance to measure the similarity between the warped template shape and the target. For both shape representations, the χ^2 distance was used to compare the SC histograms and the Hungarian method [13] is applied to compute distances between pairs of shapes.

For each image, the four most similar matches from other images in the dataset are chosen to evaluate the retrieval results. Table 2 gives the number of 1st, 2nd, 3rd and 4th most similar matches that come from the correct object. From Table 2, we can note that our method outperforms the other methods. For these shapes, the SC does not work well since a large deformation in the histogram is incurred by the articulation.

Table 2 Retrieval results on the articulated shapes dataset [18]

	Top 1	Top 2	Top 3	Top 4
SC	20/40	10/40	11/40	5/40
CPD	32/40	26/40	17/40	6/40
MR-RPM	35/40	25/40	18/40	12/40
Ours	36/40	26/40	19/40	17/40

Bold emphasis are the best results

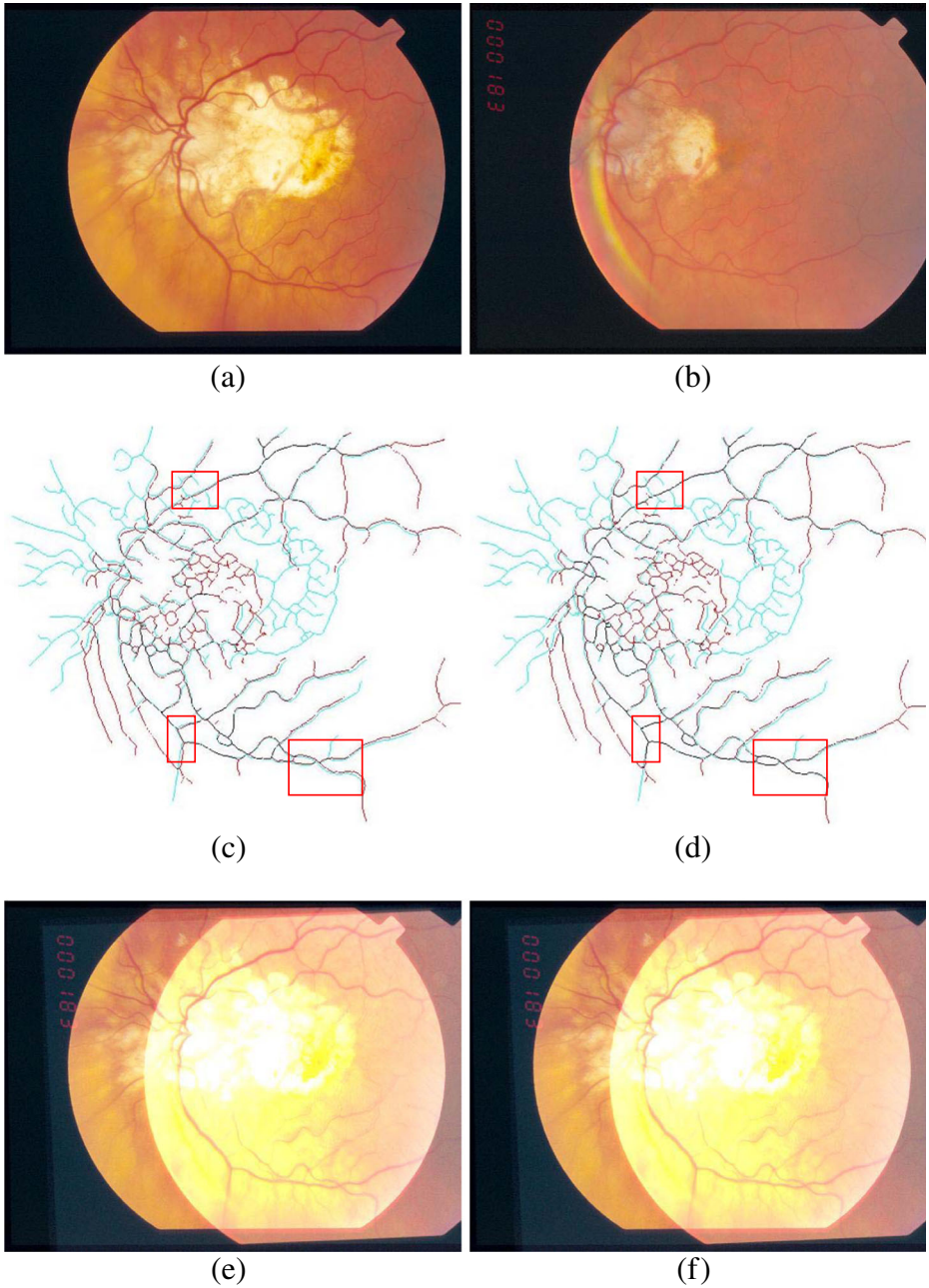


Fig. 9 Comparison of the retinal image registration performance of our method with SM. **a** and **b** a pair of retinal images taken at different times. **c** and **d** mosaic vessel image results by MS and our method, respectively. **e** and **f** mosaic retinal image results by MS and our method, respectively

Table 3 comparison of the Root Mean Squared Error(RMSE) of our method with SM on the vessel retinal image in Fig. 9c and d

	SM	Ours
Average error	3.52	1.24

Bold emphasis are the best results

3.4 Retinal image registration results

In this section, we apply the proposed methods to the tasks of retinal image registration. Given an retinal image, the ridge based vascular tree segmentation [12] is adopted to detect the vascular trees. We use the structure-matching method (SM) [7] to establish the initial correspondence. Then the proposed method is used to estimate the transformation for the inlier. Figure 9a and b are a typical pair of retinal images taken at different times. Figure 9c and d are the mosaic vessel image results by MS and our method, respectively. Figure 9e and f are the mosaic retinal image results by MS and our method, respectively. From the seams marked by the red rectangles in Fig. 9c, we can see that the vessel image pairs are not accurately aligned. By contrast, the matching accuracy of our method is much higher, as shown the seams marked by the red rectangles in Fig. 9c. A quantitative comparison of our method with SM is reported in Table 3. As shown in Table 3, our method outperform SM method. In conclusion, our method is robust for non-rigid registration.

4 Conclusion

In this paper, a new approach is proposed for non-rigid point set registration. The two steps of estimating the point correspondences and transformation between two point sets are iterated to obtained a reliable result. To this end, we preserve the local structures by expressing each point as a weighted sum of several nearest neighbors. Experiments on various synthetic and real data demonstrate that the proposed approach yields superior results to those of the state-of-the-art methods. We have also presented a method for using the proposed approach for retinal image registration. Since the matrix inversion and matrix multiplication operations in (18) and (19), the design of a more robust and efficient measure is required. It might then be combined sparse approximation based methods or parallel based methods [31–34] to reduce the time and space complexities.

Acknowledgements This work is supported in part by the National Natural Science Foundation of China under Grant 61501120 and 41501505 and in part by 2016 Outstanding Youth Research Talent Cultivation Program in Colleges and Universities in Fujian Province.

References

1. Belongie S, Malik J, Puzicha J (2002) Shape matching and object recognition using shape contexts. *IEEE Trans Pattern Anal Mach Intell* 24(24):509–522
2. Besl PJ, McKay ND (1992) A method for registration of 3-D shapes. *IEEE Trans Pattern Anal Mach Intell* 14(2):239–256
3. Bishop CM (2006) *Pattern recognition and machine learning*. Springer
4. Brown LG (1992) A survey of image registration techniques. *ACM Comput Surv* 24(4):325–376

5. Carpaneto PG (1980) Algorithm 548: solution of the assignment problem. *ACM Trans Math Softw* 6(1):104–111
6. Chen J, Ma J, Yang C, Wei J, Ma L (2015) Non-rigid point set registration via coherent spatial mapping. *Signal Process* 106:62–72
7. Chen L, Huang X, Tian J (2015) Retinal image registration using topological vascular tree segmentation and bifurcation structures. *Biomed Signal Process Control* 16:22–31
8. Chui H, Rangarajan A (2003) A new point matching algorithm for non-rigid registration. *Comput Vis Image Underst* 89:114–141
9. Dempster A, Laird N, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Statist Soc Series B* 39(1):1–38
10. Du S, Bi B, Xu G, Zhu J, Zhang X (2017) Robust non-rigid point set registration via building tree dynamically. *Multimed Tools Appl* 76(9):12065–12081
11. Fitzgibbon AW (2003) Robust registration of 2D and 3D point sets. *Image Vis Comput* 21:1145–1153
12. Freeman WT, Adelson EH (1991) The design and use of steerable filters. *IEEE Trans Pattern Anal Mach Intell* 13(9):891–906
13. Giorgio Carpaneto PT (1980) Algorithm 548: solution of the assignment problem. *ACM Trans Math Softw* 6(1):104–111
14. Hartley R, Zisserman A (2003) *Multiple view geometry in computer vision*, 2nd edn. Cambridge University Press, Cambridge
15. Jian B, Vemuri BC (2011) Robust point set registration using Gaussian mixture models. *IEEE Trans Pattern Anal Mach Intell* 33(8):1633–1645
16. LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86(11):2278–2324
17. Lian W, Zhang L (2012) Robust point matching revisited: a concave optimization approach. In: *Proceedings of European conference on computer vision*
18. Ling H, Jacobs DW (2007) Shape classification using the inner-distance. *IEEE Trans Pattern Anal Mach Intell* 29(2):840–853
19. Ma J, Zhao J, Tian J, Bai X, Tu Z (2013) Regularized vector field learning with sparse approximation for mismatch removal. *Pattern Recogn* 46(12):3519–3532
20. Ma J, Zhao J, Tian J, Tu Z, Yuille AL (2013) Robust estimation of nonrigid transformation for point set registration. In: *IEEE Conference on computer vision and pattern recognition*. IEEE, pp 2147–2154
21. Ma J, Zhao J, Tian J, Yuille AL, Tu Z (2014) Robust point matching via vector field consensus. *IEEE Trans Image Process* 23(4):1706–1721
22. Ma J, Chen J, Ming D, Tian J (2014) A mixture model for robust point matching under multi-layer motion. *PloS one* 9(3):e92282
23. Ma J, Zhou H, Zhao J, Gao Y, Jiang J, Tian J (2015) Robust feature matching for remote sensing image registration via locally linear transforming. *IEEE Trans Geosci Remote Sens* 53(12):6469–6481
24. Ma J, Qiu W, Zhao J, Ma Y, Yuille AL, Tu Z (2015) Robust L2E estimation of transformation for non-rigid registration. *IEEE Trans Signal Process* 63(5):1115–1129
25. Ma J, Zhao J, Yuille AL (2016) Non-rigid point set registration by preserving global and local structures. *IEEE Trans Image Process* 25(1):53–64
26. Ma J, Zhao J, Jiang J, Zhou H (2017) Non-rigid point set registration with robust transformation estimation under manifold regularization. In: *Inproceedings of the Thirty-First AAAI conference on artificial intelligence (AAAI)*. AAAI, pp 4218–4224
27. Ma J, Jiang J, Liu C, Li Y (2017) Feature guided Gaussian mixture model with semi-supervised em and local geometric constraint for retinal image registration. *Inf Sci* 417:128–142
28. Ma J, Jiang J, Zhou H, Zhao J, Guo X (2018) Guided locality preserving feature matching for remote sensing image registration. *IEEE Trans Geosci Remote Sensing*. <https://doi.org/10.1109/TGRS.2018.2820040>
29. Myronenko A, Song X (2010) Point set registration: coherent point drift. *IEEE Trans Pattern Anal Mach Intell* 32(12):2262–2275
30. Wahba G (1990) *Spline models for observational data*. SIAM, Philadelphia
31. Yan C, Zhang Y, Xu J, Li L, QionghaiDai F, Zhang Y, QionghaiDai A (2014) Highly parallel framework for HEVC coding unit partitioning tree decision on many-core processors. *IEEE Signal Process Lett* 21(5):573–576
32. Yan C, Xie H, Liu S, Yin J, Zhang Y, Dai Q (2018) Effective Uyghur language text detection in complex background images for traffic prompt identification. *IEEE Trans Intell Transp Syst* 19(1):220–229
33. Yan C, Xie H, Yang D, Yin J, Zhang Y (2018) Supervised hash coding with deep neural network for environment perception of intelligent vehicles. *IEEE Trans Intell Transp Syst* 19(1):284–295

34. Yan C, Zhang Y, Xu J, Dai F, Li L, Dai Q, Wu F (2014) Efficient parallel framework for HEVC motion estimation on many-core processors. *IEEE Trans Circ Syst Vid Technol* 24(12):2077–2089
35. Zheng Y, Doermann D (2006) Robust point matching for nonrigid shapes by preserving local neighborhood structures. *IEEE Trans Pattern Anal Mach Intell* 28(4):643–649



Changcai Yang received his M.S. degree in control theory and control engineering from China Three Gorges University, China, in 2008, and his Ph.D. in pattern recognition and intelligent systems at Huazhong University of Science and Technology (HUST), China, in 2012. He worked as Post Doctor at HUST, China, from 2012 to 2014. He is now an associate professor and MS supervisor of computer science and its application at Fujian Agriculture and Forestry University. His research interests include image processing, point set registration. He has published more than 30 papers.



Riqing Chen received the B.Eng. degree in communication engineering from Tongji University, China, in 2001, the M.Sc. degree in communications and signal processing from Imperial College London, U.K., in 2004, and the Ph.D. degree in engineering science from the University of Oxford, U.K., in 2010. Currently, he is a professor affiliated with Digital Fujian Institute of Big Data for Agriculture and Forestry, Fujian Agriculture and Forestry University, China. His current research interests include big data and visualization, pattern recognition, cloud computing and communications etc.