Pattern Anal Applic (2014) 17:379–388
DOI 10.1007/s10044-013-0324-z

SHORT PAPER

A robust and outlier-adaptive method for non-rigid point registration

Yuan Gao · Jiayi Ma · Ji Zhao · Jinwen Tian · Dazhi Zhang

Received: 29 November 2011 / Accepted: 30 January 2013 / Published online: 19 February 2013

Abstract In this paper, a flexible probabilistic method is introduced for non-rigid point registration, which is motivated by the pioneering research named Coherent Point Drift (CPD). Being different from CPD, our algorithm is robust and outlier-adaptive, which does not need prior information about data such as the appropriate outlier ratio when the point sets are perturbed by outliers. We consider the registration as the alignment of the data (one point set) to a set of Gaussian Mixture Model centroids (the other point set), and initially formulate it as maximizing the likelihood problem, then the problem is solved under Expectation–Maximization (EM) framework. The outlier ratio is also formulated in EM framework and will be updated during the EM iteration. Moreover, we use the volume of the point set region to determine the uniform distribution for modeling the outliers. The resulting registration algorithm exhibits inherent statistical robustness and has an explicit interpretation. The experiments demonstrate that our algorithm outperforms the state-of-the-art method.

Keywords Non-rigid point registration · Outlier-adaptive · Coherent Point Drift · Gaussian Mixture Model · Expectation–Maximization · Robust

Electronic supplementary material The online version of this article (doi:10.1007/s10044-013-0324-z) contains supplementary material, which is available to authorized users.

Y. Gao (✉) · J. Ma · J. Zhao · J. Tian · D. Zhang
State Key Laboratory for Multi-spectral Information Processing Technologies, IPRAI, Huazhong University of Science and Technology, Wuhan 430074, China
e-mail: kipykspy@gmail.com

1 Introduction

Registration of images is playing an increasingly important role in both researches and applications. It is a universal problem in medical image analysis that has benefited from over 20 years of development [1]. It is also applied in other areas including but not limited to geo-spatial imaging, satellite imaging, movie editing, archeology for shape recognition, image alignment, range datasets integration, etc. [2]. Among all the technologies, feature-based registration is a very productive field due to its suitability for situations when illumination changes are expected or multisensor analysis is demanded [2]. The features here involve surfaces, lines, curves and points. The registration of point feature, which can be regarded as the most fundamental of all features, has aroused great interesting of researchers. The objective is to find the optimal spatial transformation to map one point set to the other between two given point sets that are extracted from a pair of images such as the locations of corners, boundary points or salient regions.

However, point set registration is of great difficulty in three aspects: (1) point correspondence is unknown, (2) transformation is unknown, and (3) there is not enough information about the physical properties of objects [3]. Besides, the point set registration algorithms should possess several necessarily desirable properties to ensure its practicality: (1) capable of robustness to degradations such as noises and outliers arising from the process of image acquisition and feature extraction; (2) ability to incorporate high dimensional mappings in order to account for deformations of the point-sets; (3) the accurate transformation with tractable computational complexity for alignment of the point sets [4].

The transformation of registration can be broadly classified into rigid and non-rigid. In rigid transformations, there are only translation, rotation and scaling included. And for non-rigid case, usually more complex transformations are contained such as affine transformation which allows anisotropic scaling and skews. Compared with rigid case, non-rigid transformation occurs ubiquitously in real-world problems including deformation field recovery during surgery, deformable motion tracking, surgical procedure planning and evaluation [5]. It is such a common problem that many methods have been proposed to attack its various aspects.

Decades ago, a straightforward approach was proposed to solve the point set registration problem, which is known as iterative closest point (ICP) algorithm [6, 7]. The ICP method can well deal with rigid registration up to 3D. It assumes that every closest point pair should correspond to each other. Using this simple and intuitive criterion, ICP iteratively assigns correspondences by finding the least-squares rigid transformation between the two point sets. However, the ICP method requires a close initialization because the assumption of ICP could fail if the two point sets are not coarsely aligned. There are several variations of ICP, but some of them focus to speed it up [8], while others dedicate to extend its application to non-rigid case [3, 9]. Nevertheless, the vital problem that a proper initialization should be given still remains.

Due to the limitation of ICP, researchers turn to consider the point set registration problem by another delicate aspect. Rather than modeling the one-to-one correspondence by the closest distance criterion, more advanced fuzzy correspondence using one-to-many relaxations is implemented. The problem is modeled by Gaussian Mixture Model (GMM) in these types of methods. Generally, these methods can be classified into two broad categories.

One category of methods is that they first represent both of the two point sets as GMM centroids, then the problem is transferred to align the two GMM centroids by minimizing the discrepancy between them [10–12]. In [10], a kernel correlation (KC) based point set registration approach is proposed, and KL-divergence between the two distributions is used as the cost function, while the cost function is the L2-distance between the distributions in [11]. And in [12], the authors systematically summarize the motivation of key idea of these methods and develop a unified framework of them. These methods can only get the transformation of the two point sets.

The other category can solve for both the correspondence and the transformation simultaneously. They treat one point set as GMM centroids, and align the other point set to them by maximizing the likelihood [4, 13–17]. The TPS–RPM in [15] used the thin-plate spline (TPS) to parameterize the non-rigid spatial transformation and applied the soft-assign and deterministic annealing to estimate the correspondence. Rather than using a binary correspondence variable in ICP, the key idea of TPS–RPM is that they assume the point in one set corresponds to a weighted sum of the points in the other set. And in [14], the authors proved that, with several modifications, TPS–RPM is equivalent to EM for GMM. Myronenko et al. [4] propose a probabilistic method, called the Coherent Point Drift (CPD), which fits one point set to the other by maximizing the likelihood. This method could handle high dimensionality of the point sets and is robust to outliers. However, an appropriate outlier ratio should be given in advance, which limits its applicability. The difference between these two algorithms is that the Gaussian radial basis functions is used in [4] while the non-rigid deformation is modeled by TPS in [15]. The most recently method of this category is proposed in [17], but it can only be used in rigid point set registration.

In this paper, our method is proposed to remedy the problems stated above, which is mainly based on the state-of-art CPD method [4]. The main contributions of our paper are twofold: (1) We integrates the estimation of outlier ratio into the EM framework. Which allows us to implement this method without giving the real outlier ratio. That enlarges the practicability of CPD method since it is intuitively difficult to get a reasonable assumption of outlier ratio before the two point sets have been registered. (2) Though the outliers is also modeled by a uniform distribution as in [4], we use the volume of the set region to present the uniform interval rather than a stable item in [4]. The experiment results show that our method is more robust than the state-of-the-art CPD.

The paper is organized as follows: In Sect. 2, the stages of the CPD including methodology and the underlying problem are reviewed, followed by our method which properly overcomes the problem in Sect. 3. Section 4 presents experiments on various non-rigid data sets from 2D to 3D, and the performance of our algorithm is evaluated. Finally, we make concluding remarks in Sect. 5.

2 Review of Coherent Point Drift algorithm

The Coherent Point Drift algorithm [4] is known as the state-of-the-art method for non-rigid registration. The CPD algorithm treat one point set as GMM centroids, and then fit it to the other point set. The method obtain the optimal correspondence when the GMM posterior probability is maximized. In this section, we first briefly introduce the CPD algorithm. And then, the problems underlying the CPD method is highlighted.

2.1 Coherent Point Drift algorithm

Given two sets of point: $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_N)^{\mathrm{T}} \in \mathbb{R}^{N \times D}$ and $\mathbf{Y} = (\mathbf{y}_1, \cdots, \mathbf{y}_M)^{\mathrm{T}} \in \mathbb{R}^{M \times D}$, where the two point sets may contain unknown outliers and D denotes the dimension of the points. The purpose is to learn a transformation $\mathbf{T}: \mathbf{X} = \mathbf{T}(\mathbf{Y})$, in which the T can be defined as the initial position plus a displacement function $\mathbf{f}: \mathbf{X} = \mathbf{Y} + \mathbf{f}(\mathbf{Y})$. Then, the authors make the assumption that for inliers, the noise is Gaussian with zero mean and standard deviation σ; and for outliers, the distribution is a uniform term $\frac{1}{N}$, where N is the total number of the points in the data set $\mathbf{X}$. Then to formulate the problem, the CPD method need the input of the outlier ratio ω, which should be given in advance. And all the candidate inliers ratios are equally as $\frac{1}{M}(1-\omega)$, where M is the total number of the points in the GMM centroids set $\mathbf{Y}$. Thus the GMM takes the form:

$$\begin{aligned} p(\mathbf{x}) &= \sum_{m=1}^{M+1} P(m)p(\mathbf{x}|m) \\ &= \frac{1-\omega}{M}\sum_{m=1}^{M}\mathcal{N}(\mathbf{x}|\mathbf{y}_m + \mathbf{f}(\mathbf{y}_m), \sigma^2\mathbf{I}_{D\times D}) + \omega\mathcal{U}(N) \\ &= \frac{1-\omega}{M}\sum_{m=1}^{M}\frac{1}{(2\pi\sigma^2)^{D/2}}e^{\frac{\|\mathbf{x}-\mathbf{y}_m-\mathbf{f}(\mathbf{y}_m)\|^2}{2\sigma^2}} + \omega\frac{1}{N}, \end{aligned} \tag{1}$$

where $p(\mathbf{x}|m) = \frac{1}{(2\pi\sigma^2)^{D/2}}e^{\frac{\|\mathbf{x}-\mathbf{y}_m-\mathbf{f}(\mathbf{y}_m)\|^2}{2\sigma^2}}, P(m) = \frac{1}{M}(1-\omega)$, $m = 1, \ldots, M, P(M+1) = \omega$ and $p(\mathbf{x}|M+1) = \frac{1}{N}$.

The displacement function $\mathbf{f}$ should be smooth because the points close to one another tend to move coherently. Thus the norm of $\mathbf{f}$ should be regularized to enforce the smoothness, the details of proof can be referred to [20]. Combined with the smooth constrain, the negative log likelihood function $L(\theta)$ can be written as:

$$\begin{aligned} L(\boldsymbol{\theta}) &= -\sum_{n=1}^{N}\ln p(\mathbf{x}) + \frac{\lambda}{2}\phi(\mathbf{f}) \\ &= -\sum_{n=1}^{N}\ln\sum_{m=1}^{M+1}P(m)p(\mathbf{x}|m) + \frac{\lambda}{2}\phi(\mathbf{f}), \end{aligned} \tag{2}$$

where $\boldsymbol{\theta} = \{\mathbf{f}, \sigma^2\}$ is the set of unknown parameters, $\phi(\mathbf{f})$ is a regularization term, and $\lambda > 0$ is a trade-off parameter. Then, the EM framework is used to solve the optimal estimate of $\boldsymbol{\theta}$ by maximizing the likelihood.

2.2 The limits of the CPD algorithm

From the former subsection, both inliers and outliers has been subtly formulated in the mixture model in CPD. However, there are two problems severely influence its performance. The one is the parameter of the uniform distribution which models the outliers. CPD use a fixed item N which is the number of points in the data set, which is obscure. The intuitive definition of the uniform distribution parameter should be related to the coordinate range of the data points rather than the total number of them. For example, given two pairs of data sets with same number of points, the points distribution range of one data sets pair is large while the other is small (or one pair contains 2D points which the other is constituted by 3D points), the outliers distribution of the two pairs of data sets should not be equal.

The other problem existing in CPD is the input outlier ratio ω, which should be given in advance. That also drastically limits the application of the CPD because the one seldom know the outlier ratio of the two point sets before registration. An improper value of ω leads to unpredictability result. Assume two extreme cases, when the ω is manually set to 0, the mixture model degenerates to a pure GMM which is intuitively not rubout to outliers. While if $\omega = 1$, the mixture model becomes a pure uniform distribution which cannot deal with the point sets registration at all. In addition, the ratios of candidate inliers (GMM centroids) are equal to each other in CPD. But the GMM centroids cannot be ensured that they are all inliers and the ratio of GMM centroids should not be same. Intuitively, an updateable frame of the candidate inliers ratios should yield better performance. Moreover, the latter experiment shows that even we set the real outlier ratio to ω, the CPD algorithm still cannot guarantee a good result. Thus, the input outlier ratio ω and candidate inliers ratios play such an important role that must be optimized in the algorithm.

3 Methodology

The paper proposed a method to overcome the two problems of CPD stated above. For the outliers distribution, the volume is set to the parameter of uniform distribution so that it can be calculated initiatively in the algorithm. And for the input outlier ratio, we formulate it in the EM framework so that it can be optimized iteratively. Also, each candidate inliers ratio is added and can be updated adeptly in our algorithm. Furthermore, to avoid overlearning of each candidate inliers ratio, the learning rate is introduced in our algorithm. In this section, our method is explicitly demonstrated. Firstly, we give the derivation of our method. And then, the parameters initialization of our method is discussed.

3.1 Problem formulation and its EM solution

Suppose the candidate inliers ratio $p(m) = \pi_m, m = 1, \cdots, M$, and to distinguish with the fixed input outlier ω ratio in CPD, we suppose the input outlier ratio in our method is γ. Referring to Eq. (1), replace ω using π_m and γ, we get:

$$\begin{aligned} p(\mathbf{x}) &= \sum_{m=1}^{M+1} P(m)p(\mathbf{x}|m) \\ &= \sum_{m=1}^{M} \pi_m \mathcal{N}(\mathbf{x}|\mathbf{y}_m + \mathbf{f}(\mathbf{y}_m), \sigma^2 \mathbf{I}_{D\times D}) + \gamma \mathcal{U}(a) \\ &= \sum_{m=1}^{M} \frac{\pi_m}{(2\pi\sigma^2)^{D/2}} e^{\frac{\|\mathbf{x}-\mathbf{y}_m-\mathbf{f}(\mathbf{y}_m)\|^2}{2\sigma^2}} + \gamma \frac{1}{a}, \end{aligned} \tag{3}$$

where $p(\mathbf{x}|m) = \frac{1}{(2\pi\sigma^2)^{D/2}} e^{\frac{\|\mathbf{x}-\mathbf{y}_m-\mathbf{f}(\mathbf{y}_m)\|^2}{2\sigma^2}}$, $P(m) = \pi_m$, $m = 1$, ..., M, $P(M+1) = \gamma$ and $p(\mathbf{x}|M+1) = \frac{1}{a}$.

With the use of reproducing kernel Hilbert space, we consider the regularization term $\phi(\mathbf{f}) = \|\mathbf{f}\|^2_{\mathcal{H}}$ similar to [18]. Combined with the negative log likelihood function (2) and the standard texts [19], the complete-data log posterior can be given as follows with terms independent of θ omitted:

$$\begin{aligned} &\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) \\ &= \sum_{n=1}^{N}\sum_{m=1}^{M+1} P(m|\mathbf{x}_n, \boldsymbol{\theta}^{\text{old}}) \ln(P(m)p(\mathbf{x}_n|m, \boldsymbol{\theta})) + \frac{\lambda}{2}\phi(\mathbf{f}) \\ &= \frac{1}{2\sigma^2}\sum_{n=1}^{N}\sum_{m=1}^{M} P(m|\mathbf{x}_n, \boldsymbol{\theta}^{\text{old}}) \|\mathbf{x}_n - \mathbf{y}_m - \mathbf{f}(\mathbf{y}_m)\|^2 \\ &\quad + \frac{N_P D}{2}\ln\sigma^2 - \sum_{n=1}^{N}\sum_{m=1}^{M} P(m|\mathbf{x}_n, \boldsymbol{\theta}^{\text{old}}) \ln \pi_m \\ &\quad - \sum_{n=1}^{N} P(M+1|\mathbf{x}_n, \boldsymbol{\theta}^{\text{old}}) \ln\gamma + \frac{\lambda}{2}\|\mathbf{f}\|^2_{\mathcal{H}}, \end{aligned} \tag{4}$$

here $\boldsymbol{\theta} = \{\mathbf{f}, \sigma^2, \pi_m, \gamma\}$, which denotes the set of unknown parameters. $N_P = \sum_{n=1}^{N}\sum_{m=1}^{M+1} P(m|\mathbf{x}_n, \boldsymbol{\theta}^{\text{old}})$, and $\|\cdot\|_{\mathcal{H}}$ is the Hilbertian norm of the RKHS $\mathcal{H}$. To state our method in an explicit way, the integrated stages for solve $\boldsymbol{\theta}$ by EM algorithm are given as follows:

E-step: Denote $\mathbf{P} = (\mathbf{P}_{mn})_{M\times N}$, where the posterior probability $\mathbf{P}_{mn} = P(m|\mathbf{x}_n, \boldsymbol{\theta}^{\text{old}})$ can be computed by applying Bayes rule:

$$P(m|\mathbf{x}_n, \boldsymbol{\theta}^{\text{old}}) = \frac{\pi_m^{\text{old}} e^{-\frac{1}{2}\|\frac{\mathbf{x}_n - \mathbf{y}_m - \mathbf{f}^{\text{old}}(\mathbf{y}_m)}{\sigma^{\text{old}}}\|^2}}{\sum_{k=1}^{M} \pi_k^{\text{old}} e^{-\frac{1}{2}\|\frac{\mathbf{x}_n - \mathbf{y}_k - \mathbf{f}^{\text{old}}(\mathbf{y}_k)}{\sigma^{\text{old}}}\|^2} + \frac{\gamma^{\text{old}}(2\pi\sigma^{\text{old}}\sigma^{\text{old}})^{D/2}}{a}}, \quad \text{for} \quad m = 1, \ldots, M, \tag{5}$$

$$P(M+1|\mathbf{x}_n, \boldsymbol{\theta}^{\text{old}}) = 1 - \sum_{m=1}^{M} \mathbf{P}_{mn}. \tag{6}$$

The posterior probability P_{mn} is a soft decision, which indicates to what degree the point $\mathbf{y}_m$ corresponds to $\mathbf{x}_n$.

M-step: We determine the revised parameter $\boldsymbol{\theta}^{\text{new}}$ as $\boldsymbol{\theta}^{\text{new}} = \arg\max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$. Update σ as same as CPD:

$$\sigma^2 = \frac{\sum_{n=1}^{N}\sum_{m=1}^{M} \mathbf{P}_{mn}\|\mathbf{x}_n - \mathbf{y}_k - \mathbf{f}(\mathbf{y}_m)\|^2}{N_P D}. \tag{7}$$

Now we maximize $\mathcal{Q}(\boldsymbol{\theta})$ with respect to the candidate inlier ratios π_m and outlier ratio γ. Here we must take account of the constraint, which requires the summation of all the candidate inlier ratios and the outlier ratio to be one. This can be achieved using a Lagrange multiplier and maximizing the following quantity which has omitted terms that are independent of π_m,

$$\begin{aligned} \mathcal{Q}(\pi_m, \gamma) = &- \sum_{n=1}^{N}\sum_{m=1}^{M} P_{mn}\ln\pi_m - \sum_{n=1}^{N} P_{(M+1)n}\ln\gamma \\ &+ \mu\left(\sum_{m=1}^{M}\pi_m + \gamma - 1\right). \end{aligned} \tag{8}$$

we can get

$$\pi_m = \sum_{n=1}^{N} \mathbf{P}_{mn}/N, \quad m = 1, \ldots, M, \quad \gamma = 1 - \sum_{m=1}^{M}\pi_m. \tag{9}$$

When the outlier ratio is large, updating π_m using Eq. (9) directly may lead to the problem of over-fitting. To remedy this, we use a learning rate:

$$\pi = \pi^{\text{old}} + \frac{1}{\alpha_t}(\pi^{\text{new}} - \pi^{\text{old}}), \tag{10}$$

where π^{new} is defined by Eq. (9), π^{old} is the old π at the last iteration, and α_t is the learning rate.

The learning rate here is as the descent step size in the gradient descent algorithm, or as the annealing rate in simulated annealing algorithm. It is an empirical parameter to avoid over-learning. Without loss of generality, we suggest the learning rate α_t be set to a decreasing value as the reciprocal of the current iteration time t, so that the π can update faster at the initial iterations to get rapid optimization and update slower at the latter iterations to prevent over-learning.

Thus, combined with learning rate, we can update the outlier ratio γ and the candidate inlier ratio π_m efficiently during the EM iteration by Eqs. (9) and (10).

Considering the terms of $\mathcal{Q}(\boldsymbol{\theta})$ in Eq. (4) related to $\mathbf{f}$, and multiplying them by -1, we get an energy function:

$$\mathrm{E}(\mathbf{f}) = \frac{1}{2\sigma^2}\sum_{\mathrm{n}=1}^{\mathrm{N}}\sum_{\mathrm{m}=1}^{\mathrm{M}} \mathbf{P}_{\mathrm{mn}}\|\mathbf{x}_{\mathrm{n}} - \mathbf{y}_{\mathrm{m}} - \mathbf{f}(\mathbf{y}_{\mathrm{m}})\|^2 + \frac{\lambda}{2}\|\mathbf{f}\|^2_{\mathcal{H}}. \tag{11}$$

Then the maximization of $\mathcal{Q}$ with respect to $\mathbf{f}$ is equivalent to minimize energy function $\mathrm{E}(\mathbf{f})$. This energy function is a vector-valued extension of Tikhonov regularization, and the first term could be regarded as weighted empirical error [18].

For the registration problem, we define the hypothesis space $\mathcal{H}$ using a Gaussian matrix kernel. Using the vector-valued representer theorem [18], the optimal $\mathbf{f}$ has the following form:

$$\mathbf{f}(\mathbf{y}) = \sum_{m=1}^{M}\Gamma(\mathbf{y}, \mathbf{y}_m)c_m = \sum_{m=1}^{M}(e^{-\beta\|\mathbf{x}-\mathbf{y}\|^2} \cdot I_{D\times D})c_m. \tag{12}$$

Minimizing $E(\mathbf{f})$ with respect to c_m, we can get the optimal c_m as:

$$(\Gamma + \lambda\sigma^2(\mathrm{diag}(\mathbf{P1}))^{-1})\mathbf{C} = (\mathrm{diag}(\mathbf{P1}))^{-1}\mathbf{PX} - \mathbf{Y}, \qquad (13)$$

where $\Gamma = (\Gamma_{i,j})_{M\times M}$ is the kernel matrix, here $\Gamma_{i,j} = e^{-\beta\|\mathbf{x}_i - \mathbf{y}_j\|^2}$, $\mathbf{C} = (\mathbf{c}_1, \ldots, \mathbf{c}_M)^{\mathrm{T}} \in R^{M\times D}$, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)^{\mathrm{T}} \in R^{N\times D}$, $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_M)^{\mathrm{T}} \in R^{M\times D}$, $\mathrm{diag}(\cdot)$ denotes the diagonal of the input matrix, and $\mathbf{1}$ is a unit column.

The algorithm iterates E-step and M-step until $\mathcal{Q}$ converges. We summarize our algorithm in algorithm 1. The computational complexity of our method is $O(M^3)$, which is the same as CPD.

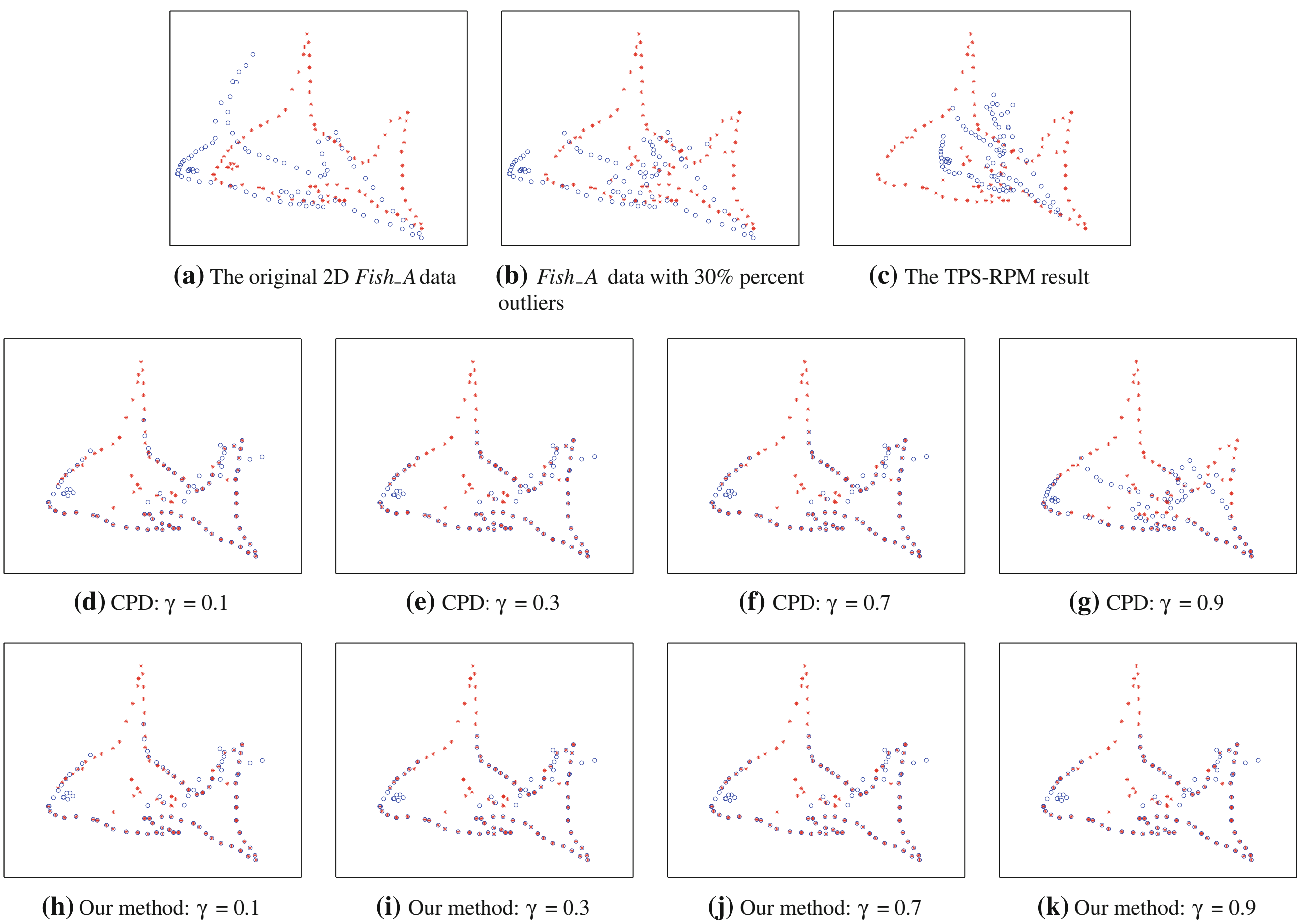
3.2 Parameters initialization

There are mainly three inputs in our algorithm: β, λ and γ. Parameters β and λ both reflect the amount of smoothness regularization. The kernel Γ is determined by the parameter β as shown in Eq. (12) which indicates how wide the range of interaction between samples. The trade-off λ presents the weight of the closeness to the data and the smoothness of the solution. A detailed description of them can be found in [20]. The β and λ are set to 0.1 and 4, respectively through this paper. And parameter γ reflects the initial assumption on the amount of outliers in correspondence sets, our algorithm is not sensitive to its value according to the formulation, and which are also proved by the experiments.

4 Experiments

To test the performance of our approach, we present experimental results on various data source including 2D shapes namely *Fish_A* and *Fish_B*, 3D surfaces namely *Face* and 3D random point sets namely *3D Random*. For each data set, quantitative comparisons with two different approaches named TPS–RPM [15] and CPD [4] are also presented. The algorithms are implemented using MATLAB R2012a, and all the experiments run on a laptop with dual-core 2.1GHz processor and 2 G RAM.

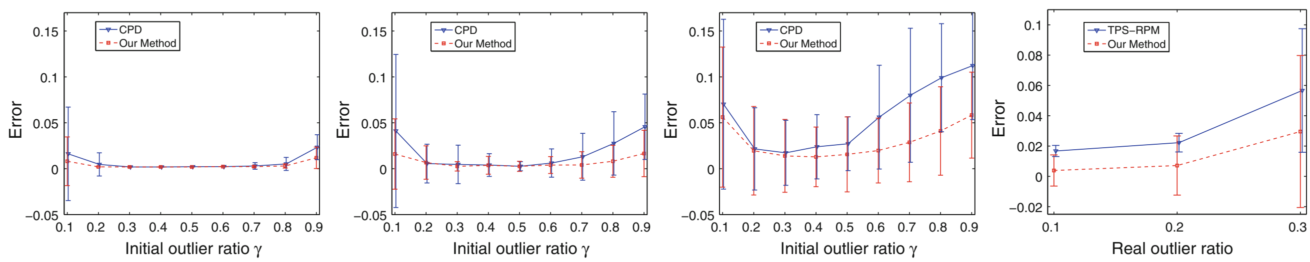
The merit of our method is illustrated by comparing the registration error with the competing registration methods. Since the ground truth are known, the registration error is



(a) The original 2D *Fish_A* data
(b) *Fish_A* data with 30% percent outliers
(c) The TPS-RPM result
(d) CPD: γ = 0.1
(e) CPD: γ = 0.3
(f) CPD: γ = 0.7
(g) CPD: γ = 0.9
(h) Our method: γ = 0.1
(i) Our method: γ = 0.3
(j) Our method: γ = 0.7
(k) Our method: γ = 0.9

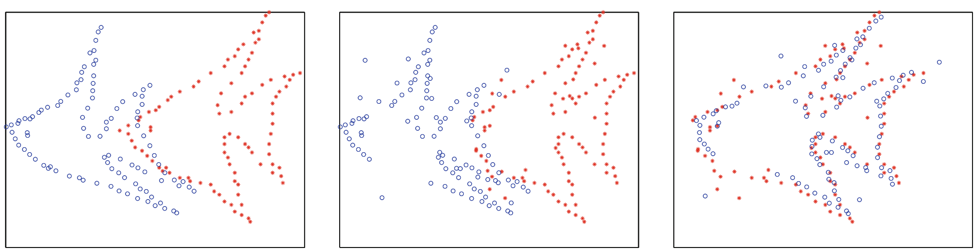
Fig. 1 Some results on 2D *Fish_A* data with 30 % outlier ratio measured as the mean squared distance between the corresponding points after the registration, which is $\frac{1}{N}\sum_{\forall(i,j)\in\mathbf{C}} D(\mathbf{x_i}, T(\mathbf{y_j}))$, where N is the total number of true correspondences, $\mathbf{C}$ is the true correspondences set, $D(\cdot,\cdot)$ is the Euclidean distance between the two input points and $\mathbf{T}$ is the registering transformation.

The algorithms are implemented on three real outlier ratio settings: 0.1, 0.2, 0.3, where the outliers are generated randomly from a normal zero-mean distribution. For comparing with CPD, we repeat the experimental process for both methods by 50 times with different randomly generated outliers at each same ratio to obtain reliable results. the results under different initial outlier ratio is reported. For testing against TPS–RPM, since the TPS–RPM method do not require the initial outlier ratio, without loss of generality, we conduct it for 450 runs, and compare

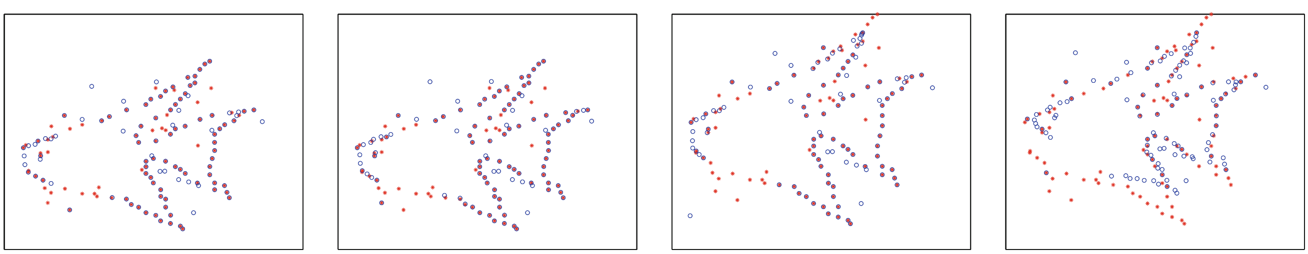


(a) **(b)** **(c)** **(d)**

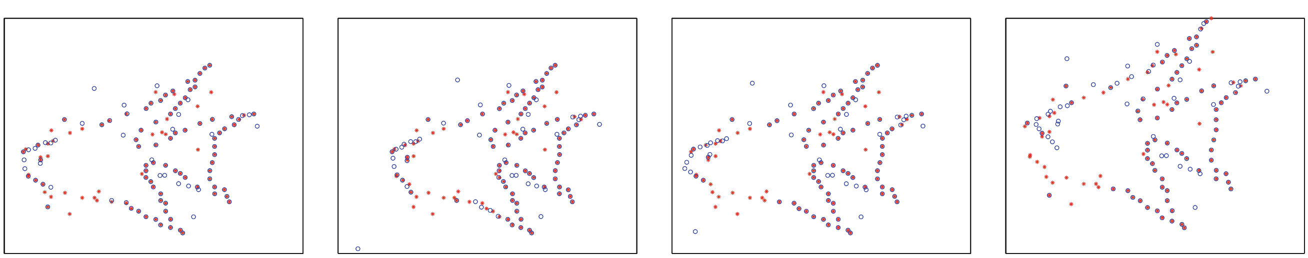
Fig. 2 A comparison of our algorithm with CPD and TPS–RPM on 2D *Fish_A* data: the comparison of our algorithm and CPD with 10 % real outlier ratio is in (**a**), 20 % real outlier ratio in (**b**), 30 % real outlier ratio in (**c**), respectively; the comparison of our algorithm and TPS–RPM is in (**d**)



(a) The original 2D *Fish_B* data **(b)** *Fish_B* data with 30% percent outliers **(c)** The TPS-RPM result



(d) CPD: $\gamma = 0.1$ **(e)** CPD: $\gamma = 0.3$ **(f)** CPD: $\gamma = 0.7$ **(g)** CPD: $\gamma = 0.9$



(h) Our method: $\gamma = 0.1$ **(i)** Our method: $\gamma = 0.3$ **(j)** Our method: $\gamma = 0.7$ **(k)** Our method: $\gamma = 0.9$

Fig. 3 Some results on 2D *Fish_B* data with 30 % outlier ratio the result with the overall mean and standard deviation of registration error in our algorithm. Where our algorithm is operated by 50 runs for each initial outlier ratio from 0.1 to 0.9.

2D Fish_A Registration: We first process our algorithm and the competing method on 2D *Fish_A* data. The original *Fish_A* data set is illustrated in Fig. 1a, which is available at http://www.csee.ogi.edu/myron. For brevity, we presented few of the experimental results with 30 % real outlier ratio in Fig. 1b–k. The statistical registration errors of our method and CPD for different outlier ratios are depicted in Fig. 2a–c, and the comparisons of our method and TPS–RPM are shown in Fig. 2d.

2D Fish_B Registration: Then, another 2D data set is applied to test the validation of our method. The original *Fish_B* in Fig. 3a is available from http://noodle.med.yale.edu/~chui/tps-rpm.html. The experimental results with 30 % real outlier ratio is partly presented in Fig. 3b–k. The statistical registration errors of our method and CPD for different outlier ratios are depicted in Fig. 4a–c, and the comparisons of our method and TPS–RPM are shown in Fig. 4d.

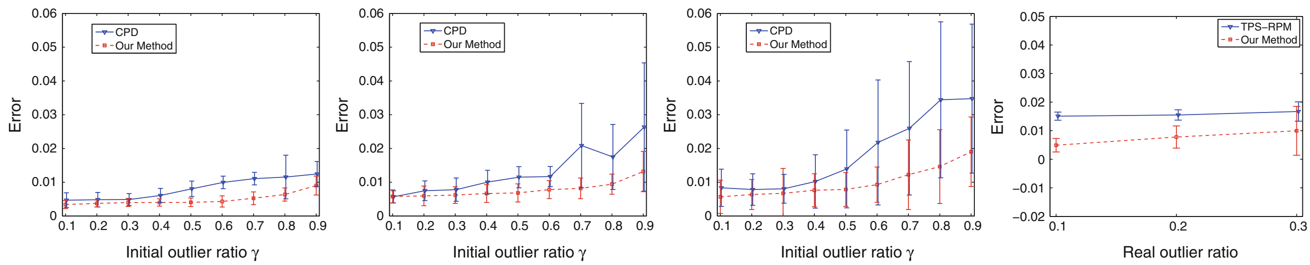
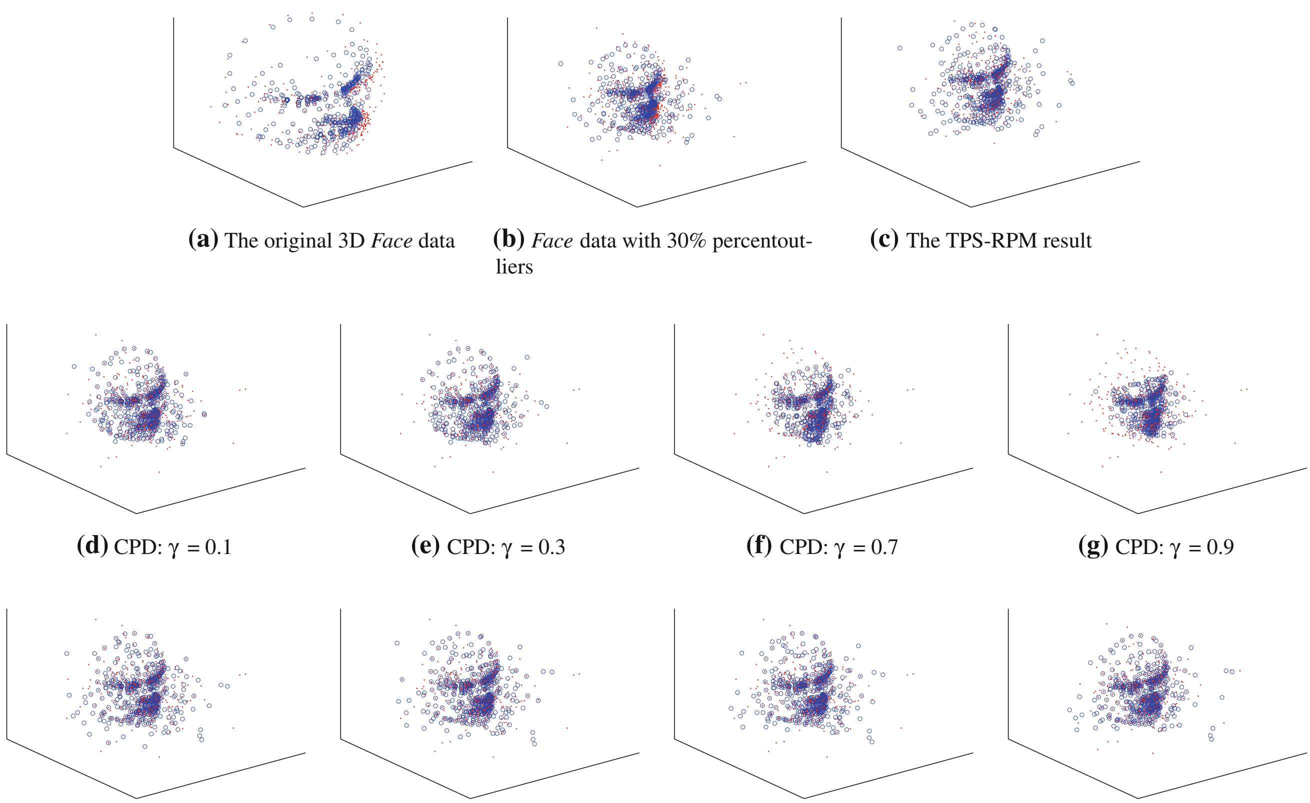


Fig. 4 A comparison of our algorithm with CPD and TPS–RPM on 2D *Fish_B* data: the comparison of our algorithm and CPD with 10 % real outlier ratio is in (**a**), 20 % real outlier ratio in (**b**), 30 % real outlier ratio in (**c**), respectively; the comparison of our algorithm and TPS–RPM is in (**d**)



(a) The original 3D *Face* data **(b)** *Face* data with 30% percentoutliers **(c)** The TPS-RPM result

(d) CPD: $\gamma = 0.1$ **(e)** CPD: $\gamma = 0.3$ **(f)** CPD: $\gamma = 0.7$ **(g)** CPD: $\gamma = 0.9$

(h) Our method: $\gamma = 0.1$ **(i)** Our method: $\gamma = 0.3$ **(j)** Our method: $\gamma = 0.7$ **(k)** Our method: $\gamma = 0.9$

Fig. 5 Some results on 3D *Face* data with 30 % outlier ratio

3D Face Registration: The meliority of our method on 2D data has been demonstrated above. Now we prepare to implement our method on 3D *Face* data along with the competing methods. This data set is obtained at http://www.csee.ogi.edu/myron and shown as Fig. 5a. Some experimental results with 30 % real outlier ratio are also in Fig. 5. The statistical registration errors of our method and CPD for different outlier ratios are depicted in Fig. 6a–c, and the comparisons of our method and TPS–RPM are shown in Fig. 6d.

3D Random Registration: Finally, a 3D *Random* data set is given to test the performance of our work. This data set containing 350 points randomly drawn from a region $[0, 1] \times [0, 1] \times [0, 1]$, then we use TPS to get the transformation map. The data set and some experimental results with 30 % real outlier ratio are presented as Fig. 7. The statistical registration errors are depicted in Fig. 8.

From the experiments, the TPS–RPM could successfully deal with both 2D and 3D data, and it performs better on 2D data than 3D. However, our method outperform TPS–RPM for all the data sets. For comparing with the CPD method, we can see that under an appropriate initial outlier ratio γ, CPD could work quite well. But beyond that, our

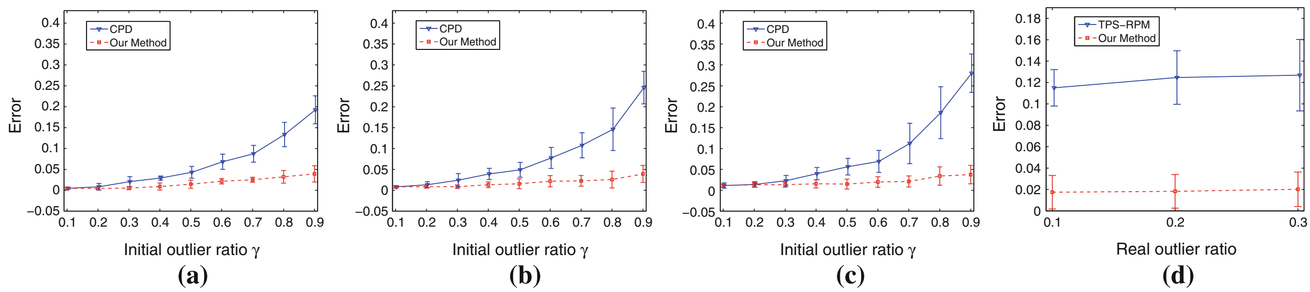
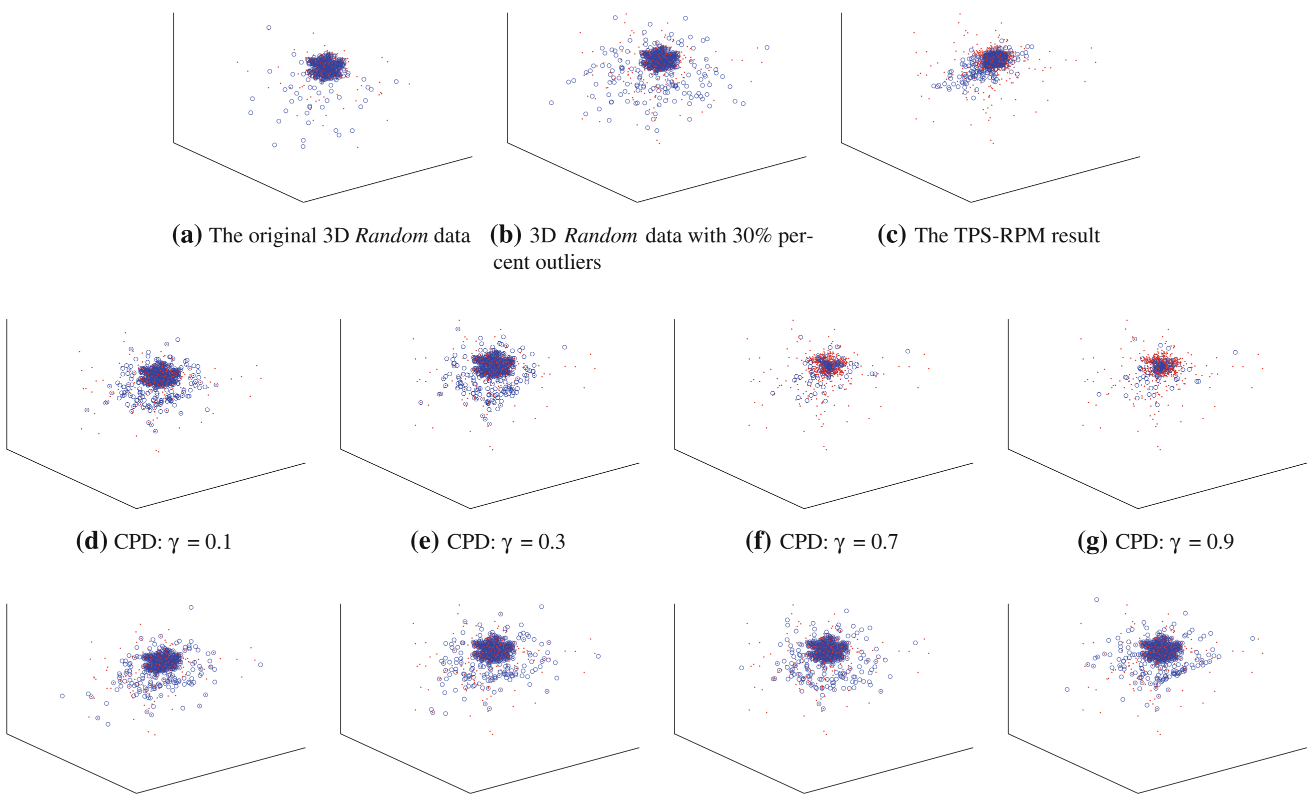


Fig. 6 A comparison of our algorithm with CPD and TPS–RPM on 3D *Face* data: the comparison of our algorithm and CPD with 10 % real outlier ratio is in (**a**), 20 % real outlier ratio in (**b**), 30 % real outlier ratio in (**c**), respectively; the comparison of our algorithm and TPS–RPM is in (**d**)



(a) The original 3D *Random* data **(b)** 3D *Random* data with 30% percent outliers **(c)** The TPS-RPM result

(d) CPD: $\gamma = 0.1$ **(e)** CPD: $\gamma = 0.3$ **(f)** CPD: $\gamma = 0.7$ **(g)** CPD: $\gamma = 0.9$

(h) Our method: $\gamma = 0.1$ **(i)** Our method: $\gamma = 0.3$ **(j)** Our method: $\gamma = 0.7$ **(k)** Our method: $\gamma = 0.9$

Fig. 7 Some results on 3D *Random* data with 30 % outlier ratio

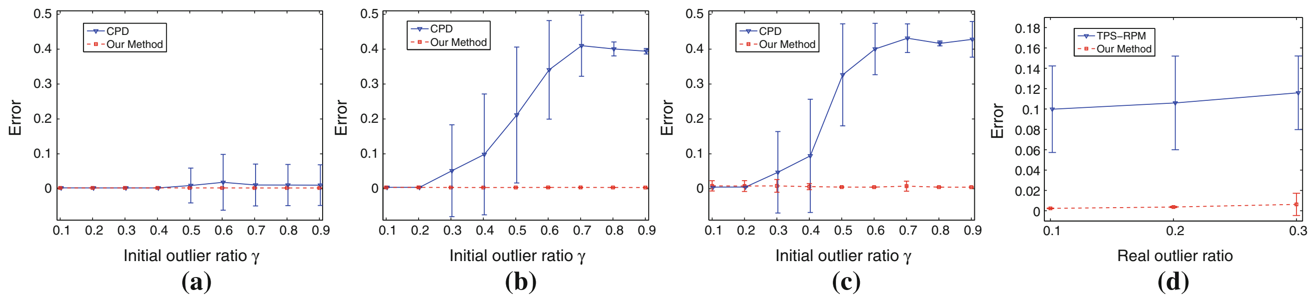


Fig. 8 A comparison of our algorithm with CPD and TPS–RPM on 3D *Random* data: the comparison of our algorithm and CPD with 10 % real outlier ratio is in (**a**), 20 % real outlier ratio in (**b**), 30 % real outlier ratio in (**c**), respectively; the comparison of our algorithm and TPS–RPM is in (**d**)

algorithm could outperforms CPD with almost any initial outlier ratio γ on both 2D and 3D point sets. Actually, our method is not sensitive to γ. We can practically fix the γ to 0.5 in practice to get robust results. Besides, to get acceptable performance, CPD has to be tested on several values of γ even the real outlier ratio is acquired. In other words, the γ which leads to the best performance does not reflect the real outlier ratio in CPD as the experimental results illustrated. While such drawback has less affection on our method. We give an explanation as follows: considering the problem formulation Eqs. (1) and (3), we model the outlier as a uniform distribution $\frac{1}{a}$, while CPD uses a stable item $\frac{1}{N}$, where N is the point number of one set, this stable item is not a probability distribution, and then it will lead to a scaling problem on the outlier ratio γ.

5 Conclusion

In this letter, we present a robust non-rigid registration algorithm which exhibits better tolerance to outliers than competing methods. Based on CPD, the main virtue of our method is that the outlier ratio can be updated during the EM iteration, which provides robust results without any assumption of prior information for outlier-perturbed data. Also, the uniform distribution which models the outlier is determined by the volume of the point set region in our method.

We compare the performance of the our method to CPD and TPS–RPM on both 2D and 3D data, and the experimental results demonstrate that our method is more robust in the presence of outliers with unknown ratios. In our future work, we plan to focus on the applicability of other distributions for modeling, such as to model the outliers other than the uniform distribution.

Acknowledgments We would like to thank the anonymous reviewers for their constructive comments. This work was funded by National Natural Science Foundation of China (Grant No.61004111) and partly supported by Key Laboratory of Geo-informatics of State Bureau of Surveying and Mapping.

Appendix: The summary of our algorithm

Algorithm 1: The summary of our algorithm

Input: Two point sets $\{\mathbf{x}_1, ..., \mathbf{x}_N\}, \{\mathbf{y}_1, ..., \mathbf{y}_N\}$, kernel $\boldsymbol{\Gamma}$, trade-off λ.
Output: Transformation $\mathbf{T}$, correspondence confidence $\mathbf{P}$.

Initialize $\mathbf{f}(\mathbf{y}_m) = \mathbf{0}$, $\sigma^2 = \frac{1}{DMN}\sum_{m,n=1}^{M,N} \|\mathbf{x}_n - \mathbf{y}_m\|^2$, kernel matrix $\boldsymbol{\Gamma}$;
Initialize $\pi_m = (1-\gamma)/M,\ m = 1, ..., M,$, where $0 < \gamma < 1$ is the initial outlier ratio and could be any real value;
Calculate a as the volume of the output space;
Set the learning rate $\alpha = t$, where t is the current iteration time;
repeat
 E-step:
 Update $\mathbf{P}$ by equations (5) and (6);
 M-step:
 Update $\mathbf{C}$ by solving linear system (13);
 Update $\mathbf{f}, \sigma^2, \pi_m, \gamma$ by equations (12), (7), (9), (10) respectively;
until Q *converges*;
Transformation $\mathbf{T}(\mathbf{y}) = \mathbf{y} + \mathbf{f}(\mathbf{y})$, correspondence confidence is given by $\mathbf{P}$

References

1. Maintz JBA, Viergever MA (1998) A survey of medical image registration. Med Image Anal 2:1–36
2. Zitova B, Flusser J (2003) Image registration methods: a survey. Image Vis Comput 21:977–1000
3. Du S, Zheng N, Ying S, Liu J (2010) Affine iterative closest point algorithm for point set registration. Pattern Recogn Lett 31: 791–799
4. Myronenko A, Song X (2010) Point set registration: coherent point drift. IEEE Trans Pattern Anal Mach Intell 32(12):2262–2275
5. Rueckert D (2001) Non-rigid registration: techniques and applications. In: Medical Image Registration. CRC Press, London.
6. Besl PJ, McKay ND (1992) A method for registration of 3-D shapes. IEEE Trans Pattern Anal Mach Intell 14(2):239–256
7. Zhang Z (1994) Iterative point matching for registration of free-form curves and surfaces. Int J Comput Vis 13(2):119C152
8. Fitzgibbon AW (2003) Robust registration of 2D and 3D point sets. Image Vis Comput 21(13–14):1145–1153
9. Amberg B, Romdhani S, Vetter T (2007) Optimal step nonrigid ICP algorithms for surface registration. In: Proceedings of International Conference on Computer Vision and Pattern Recognition, pp 1–8
10. Tsin Y, Kanade T (2004) A correlation-based approach to robust point set registration. In: Proceedings of European Conference on Computer Vision, pp 558–569
11. Jian B, Vemuri BC (2005) A robust algorithm for point set registration using mixture of Gaussians. In: Proceedings of IEEE International Conference on Computer Vision, pp 1246–1251
12. Jian B, Vemuri BC (2011) Robust point set registration using Gaussian mixture models. IEEE Trans Pattern Anal Mach Intell 33:1633–1645
13. Myronenko A, Song X, Miguel A (2006) Non-rigid point set registration: coherent point drift. In: Proceedings of Neural Information Process Systems
14. Chui H, Rangarajan A (2000) A feature registration framework using mixture models. In: Proceedings of IEEE workshop on Mathematical Methods in Biomedical Image Analysis, pp 190–197
15. Chui H, Rangarajan A (2003) A new point matching algorithm for non-rigid registration. Comput Vis Image Underst 89(2–3): 114–141
16. Chui H, Rangarajan A, Zhang J, Leonard CM (2004) Unsupervised learning of an atlas from unlabeled point-sets. IEEE Trans Pattern Anal Mach Intell 26(2):160–172
17. Horaud R, Forbes F, Yguel M, Dewaele G, Zhang J (2011) Rigid and articulated point registration with expectation conditional maximization. IEEE Trans Pattern Anal Mach Intell 33:587–602
18. Zhao J, Ma J, Tian J, Ma J, Zhang D (2011) A robust method for vector field learning with application to mismatch removing. In: Proceedings of International Conference on Computer Vision and Pattern Recognition, pp 2977–2984
19. Bishop CM (2006) Pattern recognition and machine learning. Springer, Heidelberg
20. Yuille AL, Grzywacz NM (1989) A mathematical analysis of the motion coherence theory. Int J Comput Vis 3:155–175