Signal Processing 106 (2015) 62–72

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Non-rigid point set registration via coherent spatial mapping

Jun Chen [a], Jiayi Ma [b], Changcai Yang [c,*], Li Ma [d], Sheng Zheng [e]

[a] Experimental Teaching Center of Information Technology, China University of Geosciences, Wuhan, 430074, China
[b] Electronic Information School, Wuhan University, Wuhan 430072, China
[c] College of Computer and Information Science, Fujian Agriculture and Forestry University, Fuzhou 350002, China
[d] Faculty of Mechanical and Electronic Information Engineering, China University of Geosciences, Wuhan 430074, China
[e] Institute of Intelligent Vision and Image Information, College of Science, China Three Gorges University, Yichang 443002, China

ARTICLE INFO

Article history:
Received 26 January 2014
Received in revised form
9 June 2014
Accepted 7 July 2014
Available online 15 July 2014

Keywords:
Point set registration
Coherent spatial mapping (CSM)
Maximum likelihood problem
Thin-plate spline (TPS)
Expectation Maximization (EM) algorithm

ABSTRACT

Point set registration is essential for many computer vision technologies. In this paper, we proposed a new point set registration method based on *coherent spatial mapping* (CSM), whose solution is formulated as a maximum likelihood problem. Our algorithm iteratively recovers the point correspondence and estimates the transformation between two point sets. In the first step of the iteration, feature descriptors such as shape context are used to establish rough correspondence. In the second step, each training sample is associated with a latent variable which indicates whether it is an inlier, and then the thin-plate spline (TPS) is chosen to parameterize the CSM and Expectation Maximization (EM) algorithm is used to solve it. The registration experiments are undertaken on various synthetic and real data. The results demonstrate that the proposed approach is robust and accurate for scenes under deformation, noise, outlier, occlusion, and rotation and is superior to the state-of-the-art methods.

1. Introduction

Point set registration is a fundamental problem which is widely used in computer vision, image analysis, and pattern recognition [1–5]. Many tasks in these fields – such as stereo matching, shape matching, image registration and content-based image retrieval – can be formulated as a point matching problem because point representations are general and easy to extract [1]. The points in these tasks are typically the locations of interest points extracted from an image, or the edge points sampled from a shape contour. The registration problem then reduces to determining the correct correspondence and to find the underlying spatial transformation between two point sets extracted from the input data.

The registration problem can be categorized into rigid or non-rigid registration depending on the application and the form of the data. Rigid registration, which only involves a small number of transformation parameters, is relatively easy and has been widely studied [1,2,6]. By contrast, non-rigid registration is more difficult since the underlying non-rigid transformations are often unknown, complex, and hard to model [3]. However, non-rigid registration is very important because it is required for many real world tasks including hand-written character recognition, shape recognition, deformable motion tracking and medical image registration.

In this paper, we focus on the non-rigid case and present a robust algorithm for non-rigid point set registration. There are two unknown variables we have to solve for in this problem: the correspondence and the transformation. Although solving for either variable without

* Corresponding author.
E-mail addresses: chenjun71983@163.com (J. Chen), jyma2010@gmail.com (J. Ma), changcaiyang@gmail.com (C. Yang), maryparisster@gmail.com (L. Ma), zsh@ctgu.edu.cn (S. Zheng).

http://dx.doi.org/10.1016/j.sigpro.2014.07.004
0165-1684/

information regarding the other is difficult, an iterated estimation framework can be used [2,3,7]. In this iterative process, the estimate of the correspondence is used to refine the estimate of the transformation, and vice versa. But a problem arises if there are errors in the correspondence which occurs in many applications particularly if the transformation is large and/or there are outliers in the data (e.g., data points that are not undergoing the non-rigid transformation). In this situation, the estimate of the transformation will degrade badly unless it is performed robustly.

To tackle the problem, various approaches have been proposed. The iterated closest point (ICP) algorithm [2] is one of the best known point registration approaches. It uses nearest-neighbor relationships to assign a binary correspondence, and then uses estimated correspondence to refine the transformation. Belongie et al. [7] introduced a method for registration based on the shape context descriptor, which incorporates the neighborhood structure of the point set and thus helps establish correspondence between the point sets. But these methods ignore robustness when they recover the transformation from the correspondence. To address the robust issue, Ma et al. [8] proposed to use a robust estimator to seek the transformation modeled by Gaussian radial basis functions. In related work, Chui and Rangarajan [3] established a general framework for estimating correspondences and transformations for non-rigid point matching. They modeled the transformation as a thin-plate spline and did robust point matching by an algorithm (TRS-RPM) which involved deterministic annealing and soft-assignment. Lian and Zhang [9] reduced the energy function of RPM to a concave function with very few non-rigid terms after eliminating the transformation variables and minimized the resulting energy function by a concave optimization technique. Alternatively, the coherence point drift (CPD) algorithm [10] uses Gaussian radial basis functions instead of thin-plate splines, and it was shown to be robust in the presence of outliers and noises. Other related work includes the kernel correlation (KC) based method [11], multi-layer motion based method [12], neighborhood structure preserving method [13], Gaussian mixture model based method [14], as well as vector field interpolation based method [15,16].

The main contribution of our approach is to robustly estimate the transformations from the contaminated correspondences. More precisely, our approach iteratively recovers the point correspondences and estimates the transformation between two point sets. In the first step of the iteration, feature descriptors such as shape context are used to establish correspondence. In the second step, we estimate a coherent spatial mapping fitting for the inlier as the transformation. To this end, we model the outlier process using extra (hidden) variables which enable us to identify and reject outliers. Moreover, the non-rigid coherent spatial mapping is parameterized by a general purpose spline tool – the thin-plate spline (TPS) [17].

The rest of the paper is organized as follows. Section 2 describes our robust algorithm for estimating coherent spatial mappings from point correspondences with some amount of unknown outliers. In Section 3, we lay out our method for non-rigid point set registration based on the estimation of coherent spatial mapping. Section 4 demonstrates our experimental results on various data. Finally, we conclude this paper in Section 5.

2. Estimating coherent spatial mapping from correspondence

In this section, we consider the global spatial relationship 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. The main idea of our approach is illustrated in Fig. 1. In Fig. 1a, we are given two sets of points, where the goal is to align the model points (circles) to the target points (pluses). To this end, we use the shape context feature [7] combining with Hungarian method [18] to establish a set of initial point correspondences (which will be described in detail later); the result is shown in Fig. 1b, where blue lines indicate correct correspondences and red lines indicate false correspondence. Fig. 1c is the corresponding motion field, in which each arrow denotes a correspondence: the head and tail correspond to the positions of points in two point sets. We see that the blue arrows are regular and the global spatial motion is slow-and-smooth [19], while the distribution of the red arrows seems random. Therefore, it is possible to learn a spatial mapping function for the underlying correct correspondence by using a smooth constraint. We refer to the spatial mapping as *coherent spatial mapping* since it fits only those correspondences (i.e., inliers) have the slow-and-smooth spatial relationship.

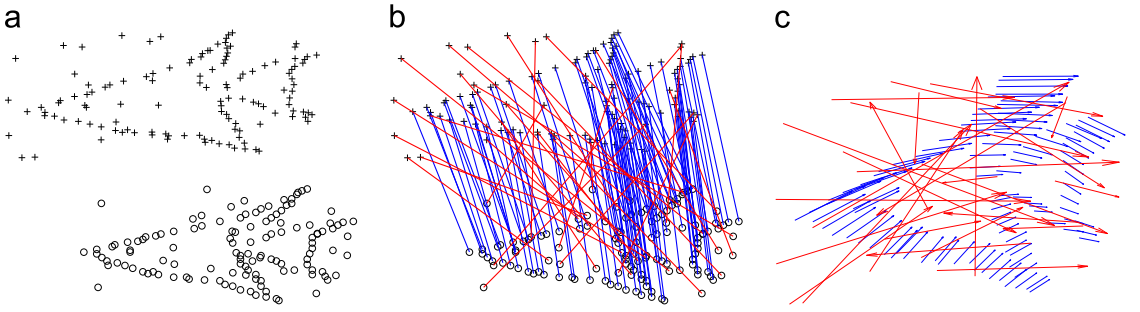

Fig. 1. Point matching and motion field between two point sets. (a) Point sets of the model (circles) and target (pluses) shapes. (b) Point matching result. (c) Motion field between two point sets. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

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, the goal is to remove outliers contained in the point correspondences, and simultaneously estimate a coherent spatial mapping $\mathbf{f}:\mathbf{y}_n=\mathbf{f}(\mathbf{x}_n)$ to fit the underlying inliers.

Due to the existence of outliers, it is desirable to have a robust estimation of $\mathbf{f}$. To this end, we make the assumption that, for the inliers, the noise is Gaussian on each component with zero mean and uniform standard deviation σ; for the outliers, the output space is a bounded region of $\mathbb{R}^2$, and the distribution is assumed to be uniform $1/a$, where a is a constant. We then associate the n-th sample with a latent variable $z_n\in\{0,1\}$, where $z_n=1$ indicates a Gaussian distribution and $z_n=0$ points to a uniform distribution. Let $\mathbf{X}$ and $\mathbf{Y}$ be the set of observed input and output data, in which the n-th rows represent $\mathbf{x}_n^{\mathrm{T}}$ and $\mathbf{y}_n^{\mathrm{T}}$. We use homogeneous coordinates for the points, e.g., $\mathbf{x}=(\mathbf{x}_x,\mathbf{x}_y,1)$. Thus, the likelihood is a mixture model given by

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

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., $\forall z_n$, $p(z_n=1)=\gamma$. Note that the uniform distribution function is nonzero only in a bounded region (here we omit the indicator function for clarity).

Generally speaking, the true parameter set $\boldsymbol{\theta}$ maximizes likelihood (1). Now we give a maximum likelihood estimation of $\boldsymbol{\theta}$, i.e. $\boldsymbol{\theta}^*=\arg\max_{\boldsymbol{\theta}}p(\mathbf{Y}|\mathbf{X},\boldsymbol{\theta})$. This is equivalent to seeking the minimal energy

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

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

2.2. The EM algorithm

The well known EM algorithm [20] provides a natural framework for solving the problem. It alternates with two steps: an expectation step (E-step) and a maximization step (M-step). The E-step basically estimates the responsibility indicating to what degree a correspondence belongs to inlier under the given coherent spatial mapping $\mathbf{f}$, while the M-step updates $\mathbf{f}$ based on the current estimate of the responsibility. We follow standard notations [21] and omit some terms that are independent of $\boldsymbol{\theta}$. Considering the negative log likelihood function, i.e. Eq. (2), the complete-data log likelihood is then given by

$$\begin{aligned}\mathcal{Q}(\boldsymbol{\theta},\boldsymbol{\theta}^{\mathrm{old}})=&-\frac{1}{2\sigma^2}\sum_{n=1}^{N}P(z_n=1|\mathbf{x}_n,\mathbf{y}_n,\boldsymbol{\theta}^{\mathrm{old}})\|\mathbf{y}_n-\mathbf{f}(\mathbf{x}_n)\|^2\\&-\ln\sigma^2\sum_{n=1}^{N}P(z_n=1|\mathbf{x}_n,\mathbf{y}_n,\boldsymbol{\theta}^{\mathrm{old}})\\&+\ln\gamma\sum_{n=1}^{N}P(z_n=1|\mathbf{x}_n,\mathbf{y}_n,\boldsymbol{\theta}^{\mathrm{old}})\\&+\ln(1-\gamma)\sum_{n=1}^{N}P(z_n=0|\mathbf{x}_n,\mathbf{y}_n,\boldsymbol{\theta}^{\mathrm{old}}).\end{aligned}\quad(3)$$

This may be maximized by treating the z_n as the missing data from the mixture model in Eq. (1).

E-step: Denote $\mathbf{P}=\mathrm{diag}(p_1,\ldots,p_N)$ a diagonal matrix, where $p_n=P(z_n=1|\mathbf{x}_n,\mathbf{y}_n,\boldsymbol{\theta}^{\mathrm{old}})$ can be computed by applying Bayes rule:

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

The posterior probability p_n is a soft decision, which indicates to what degree the sample n agrees with the current estimated coherent spatial mapping $\mathbf{f}$.

M-step: We determine the revised parameter estimate $\boldsymbol{\theta}^{\mathrm{new}}$ as follows: $\boldsymbol{\theta}^{\mathrm{new}}=\arg\max_{\boldsymbol{\theta}}\mathcal{Q}(\boldsymbol{\theta},\boldsymbol{\theta}^{\mathrm{old}})$. Taking derivative of $\mathcal{Q}(\boldsymbol{\theta})$ with respect to the variance σ^2 and the mixing coefficient γ and setting them to zero, we obtain

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

$$\gamma=\mathrm{tr}(\mathbf{P})/N,\quad(6)$$

where $\mathbf{V}=(\mathbf{f}(\mathbf{x}_1),\ldots,\mathbf{f}(\mathbf{x}_N))$, and $\mathrm{tr}(\cdot)$ is the trace. Notice that $\mathbf{f}(\mathbf{x}_n)$ is a homogeneous coordinate, we should normalize it with scale 1 before computing p_n and σ^2.

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

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

This is in general ill-posed since $\mathbf{f}$ is not unique. To generate a smooth mapping fitting for the image point correspondences, we choose the TPS for parametrization. The TPS is a general purpose spline tool which produces a smooth functional mapping for supervised learning [17]. It has no free parameters that need manual tuning, and also has a close-form solution which can be decomposed into a global linear affine motion and a local non-affine warping component controlled by coefficients $\mathbf{A}$ and $\mathbf{W}$ respectively,

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

where $\tilde{K}(\mathbf{x})$ 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 $K_{ij}=K(|\mathbf{x}_i-\mathbf{x}_j|)$. If we substitute (8) into the expression (7), the error function can be written in the form

$$\mathcal{Q}(\mathbf{f})=\frac{1}{2\sigma^2}\sum_{n=1}^{N}p_n\|\mathbf{y}_n-\mathbf{x}_n\mathbf{A}-\tilde{K}(\mathbf{x}_n)\mathbf{W}\|^2.\quad(9)$$

With a regularization parameter λ, the non-rigid coherent spatial mapping $\mathbf{f}$ can be then estimated by minimizing

a TPS energy function as

$$\mathcal{E}(\mathbf{A},\mathbf{W}) = \frac{1}{2\sigma^2}\|\mathbf{P}^{1/2}(\mathbf{Y}-\mathbf{X}\mathbf{A}-\mathbf{K}\mathbf{W})\|^2 + \frac{\lambda}{2}\operatorname{tr}(\mathbf{W}^{\mathrm{T}}\mathbf{K}\mathbf{W}) = \frac{1}{2\sigma^2}\|\tilde{\mathbf{Y}}-\tilde{\mathbf{X}}\mathbf{A}-\mathbf{P}^{1/2}\mathbf{K}\mathbf{W}\|^2 + \frac{\lambda}{2}\operatorname{tr}(\mathbf{W}^{\mathrm{T}}\mathbf{K}\mathbf{W}), \tag{10}$$

where $\tilde{\mathbf{X}} = \mathbf{P}^{1/2}\mathbf{X}$, and $\tilde{\mathbf{Y}} = \mathbf{P}^{1/2}\mathbf{Y}$. The second smoothness term is the standard TPS regularization term, and it is the bending energy which has a physical explanation and is independent on the linear component of the coherent spatial mapping.

To solve the TPS parameter pair $\mathbf{A}$ and $\mathbf{W}$, we use a QR decomposition [17,3]

$$\tilde{\mathbf{X}} = [\mathbf{Q}_1\mathbf{Q}_2]\begin{bmatrix}\mathbf{R}\\ \mathbf{0}\end{bmatrix}, \tag{11}$$

where $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are orthonormal matrices of sizes $N\times 3$ and $N\times(N-3)$, respectively, and $\mathbf{R}$ is an upper triangular matrix of size 3×3. With the QR decomposition in place, Eq. (10) becomes

$$\mathcal{E}(\mathbf{A},\boldsymbol{\Gamma}) = \frac{1}{2\sigma^2}(\|\mathbf{Q}_2^{\mathrm{T}}\tilde{\mathbf{Y}}-\mathbf{Q}_2^{\mathrm{T}}\mathbf{P}^{1/2}\mathbf{K}\mathbf{Q}_2\boldsymbol{\Gamma}\|^2 + \|\mathbf{Q}_1^{\mathrm{T}}\tilde{\mathbf{Y}}-\mathbf{R}\mathbf{A}-\mathbf{Q}_1^{\mathrm{T}}\mathbf{P}^{1/2}\mathbf{K}\mathbf{Q}_2\boldsymbol{\Gamma}\|^2) + \frac{\lambda}{2}\operatorname{tr}(\boldsymbol{\Gamma}^{\mathrm{T}}\mathbf{Q}_2^{\mathrm{T}}\mathbf{K}\mathbf{Q}_2\boldsymbol{\Gamma}), \tag{12}$$

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

Minimizing the energy function (12) with respect to $\boldsymbol{\Gamma}$ and $\mathbf{A}$, we obtain

$$\mathbf{W} = \mathbf{Q}_2\boldsymbol{\Gamma} = \mathbf{Q}_2(\mathbf{S}^{\mathrm{T}}\mathbf{S}+\lambda\sigma^2\mathbf{T}+\tilde{\epsilon}\mathbf{I})^{-1}\mathbf{S}^{\mathrm{T}}\mathbf{Q}_2^{\mathrm{T}}\tilde{\mathbf{Y}}, \tag{13}$$

$$\mathbf{A} = \mathbf{R}^{-1}\mathbf{Q}_1^{\mathrm{T}}(\tilde{\mathbf{Y}}-\mathbf{P}^{1/2}\mathbf{K}\mathbf{W}), \tag{14}$$

where $\mathbf{S} = \mathbf{Q}_2^{\mathrm{T}}\mathbf{P}^{1/2}\mathbf{K}\mathbf{Q}_2$, $\mathbf{T} = \mathbf{Q}_2^{\mathrm{T}}\mathbf{K}\mathbf{Q}_2$, and $\tilde{\epsilon}\mathbf{I}$ is used for numerical stability. Thus we obtain the coherent spatial mapping $\mathbf{f}$ in Eq. (8). The time complexity of our algorithm is $O(N^3)$ due to the matrix inversion and matrix multiplication operations in Eqs. (13) and (14).

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 $T_{\mathbf{x}}$ and $T_{\mathbf{y}}$ for the point sets $\{\mathbf{x}_n\}$ and $\{\mathbf{y}_n\}$, respectively, i.e. $\hat{\mathbf{x}}_n = T_{\mathbf{x}}\mathbf{x}_n$, which ensure that point sets have zero means and average distance $\sqrt{2}$ to the origin [22]. The regularization parameter λ controls the trade-off between the closeness to the data and the smoothness of the mapping. We set it 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. Estimation of coherent spatial transformation from correspondences.

Input: Putative correspondence set $S = \{\mathbf{x}_n, \mathbf{y}_n\}_{n=1}^{N}$, kernel K, and parameters λ, a
Output: Coherent spatial mapping $\mathbf{f}$

1 Normalization: $\hat{\mathbf{x}}_n = T_{\mathbf{x}}\mathbf{x}_n$, $\hat{\mathbf{y}}_n = T_{\mathbf{y}}\mathbf{y}_n$;
2 Construct kernel matrix $\mathbf{K}$ using the definition of K;
3 Initialization: $\gamma = 0.9$, $\mathbf{W} = \mathbf{0}$, $\mathbf{A} = \mathbf{I}_{3\times 3}$, $\mathbf{P} = \mathbf{I}_{N\times N}$;
4 Initialize σ^2 by Eq. (5);
5 **repeat**
6 *E-step*:
7 Update the responsibility p_n by Eq. (4);
8 *M-step*:
9 Update mapping $\mathbf{f}$ by using Eqs. (13) and (14);
10 Update σ^2 and γ by Eqs. (5) and (6);
11 **until** *some stopping criterion is satisfied*;
12 The spatial mapping $\mathbf{f}$ is determined by Eq. (8).

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. We have shown above that once we have established the correspondence between the two point sets even with noise and outliers, we are able to estimate the underlying coherent spatial mapping between them. Next, we discuss how to establish rough correspondence between two point sets.

3.1. Establishment of point correspondence

Recall that our method described above does not jointly solve the transformation and point correspondence. In order to use Algorithm 1 to solve the transformation between two point sets, we need initial correspondences.

In general, if the two point sets have similar shapes, the corresponding points have similar neighborhood structures which could be incorporated into a feature descriptor. Thus finding correspondences between two point sets is equivalent to 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, the initial correspondences need not be very accurate, since our method is robust to noise and outliers. Inspired by these facts, we use shape context (SC) [7] as the feature descriptor. Consider two points $\mathbf{x}_m$ and $\mathbf{y}_l$, their SCs are histograms $\{p_m(k)\}_{k=1}^{K}$ and $\{q_l(k)\}_{k=1}^{K}$, respectively. The χ^2 distance was used to measure their difference

$$C_{ml} = C(\mathbf{x}_m,\mathbf{y}_l) = \frac{1}{2}\sum_{k=1}^{K}\frac{[p_m(k)-q_l(k)]^2}{p_m(k)+q_l(k)}. \tag{15}$$

After computing all of $\{C_{ml}, m=1,\ldots,M, l=1,\ldots,L\}$, the Hungarian method [18] is applied to find the putative correspondences between $\{\mathbf{x}_m\}_{m=1}^{M}$ and $\{\mathbf{y}_l\}_{l=1}^{L}$.

In the 3D case, the spin image [23] can be used as a feature descriptor, where the local similarity is measured by an improved correlation coefficient. Then the matching is performed by a method which encourages geometrically consistent groups.

The two steps of estimating correspondences and transformations are iterated to obtain a reliable result. In this paper, we use a fixed number of iterations, typically 10 but more when the noise is big or when there are a large percentage of outliers contained in the original point sets. We summarize our point set registration method in Algorithm 2.

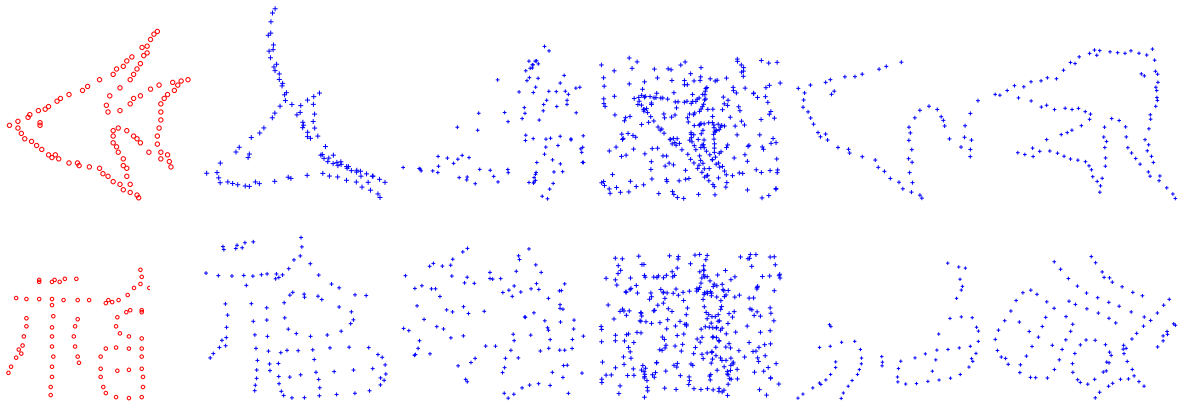

Fig. 2. Examples of synthetic data used in the experiments. Top row: the shapes of the fish. Bottom row: the shapes of the Chinese character. The left column: the template point sets. From the second to the right column: examples of the target point sets for the deformation, noise, outlier, occlusion,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

Algorithm 2. Non-rigid point set registration.

Input: Two point sets $\{\mathbf{x}_m\}_{m=1}^{M}$, $\{\mathbf{y}_l\}_{l=1}^{L}$

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 initial correspondences based on the feature descriptors of two point sets;
5 Solve the transformation $\mathbf{f}$ warping the model point set to the target point set using Algorithm 1;
6 Update model point set $\{\mathbf{x}_m\}_{m=1}^{M} \leftarrow T_{\mathbf{y}}^{-1}\{\mathbf{f}(\mathbf{x}_m)\}_{m=1}^{M}$;
7 **until** *reach the maximum iteration number*;
8 The aligned model point set $\{\hat{\mathbf{x}}_m\}_{m=1}^{M}$ is given by $T_{\mathbf{y}}^{-1}\{\mathbf{f}(\mathbf{x}_m)\}_{m=1}^{M}$ in the last iteration.

4. Experimental results

In order to evaluate the performance of our proposed approach, we compare our algorithm to other state-of-the-art approaches using the synthetic and real data.

4.1. Experiments with synthesized data

We test our method on the same synthesized data as in [3] and [13]. The data consists of two different shape models: a *fish* and a *Chinese character*. For each model, there are five sets of data designed to measure the robustness of registration algorithms under deformation, occlusion, rotation, noise and outliers. 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. Fig. 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, rotation invariance is also required. We use the rotation invariant shape context as in [13] if necessary.

Figs. 3 and 4 show the registration results of our method with comparison to four state-of-the-art methods: SC [7], TPS-RPM [3], CPD [10], and COA-RPM [9], which are implemented using publicly available codes. In the tests of rotation, we use the rotation invariant shape context as the feature descriptor in our approach.[1] 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 five 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 statistical results for each setting are summarized in Figs. 5 and 6. Fig. 5a and b shows the results for the shapes of fish and Chinese character with deformation, respectively. For the shape of fish, our method and CPD perform best, followed by SC and TPS-RPM. For the shape of Chinese character, our approach performs a litter better than CPD, and they both achieve quite satisfactory performance. Fig. 5c and d shows the results for the shapes of fish and Chinese character undergoing the Gaussian noise, respectively. CPD performs best, followed by TPS-RPM and our method. Fig. 5e and f shows the results for the shapes of fish and Chinese character with outlier, respectively. The best results are obtained by TPS-RPM for both shapes. The error means of our approach are greater than SC, CPD , and COA-RPM. In fact, the registration errors of CPD, COA-RPM, and

[1] We do not use the rotation invariant shape context in the other cases, since the rotation invariant may reduce the discriminative ability of a feature descriptor when rotation is indeed not involved (which is also more realistic), for example, recognizing the numbers '6' and '9'.

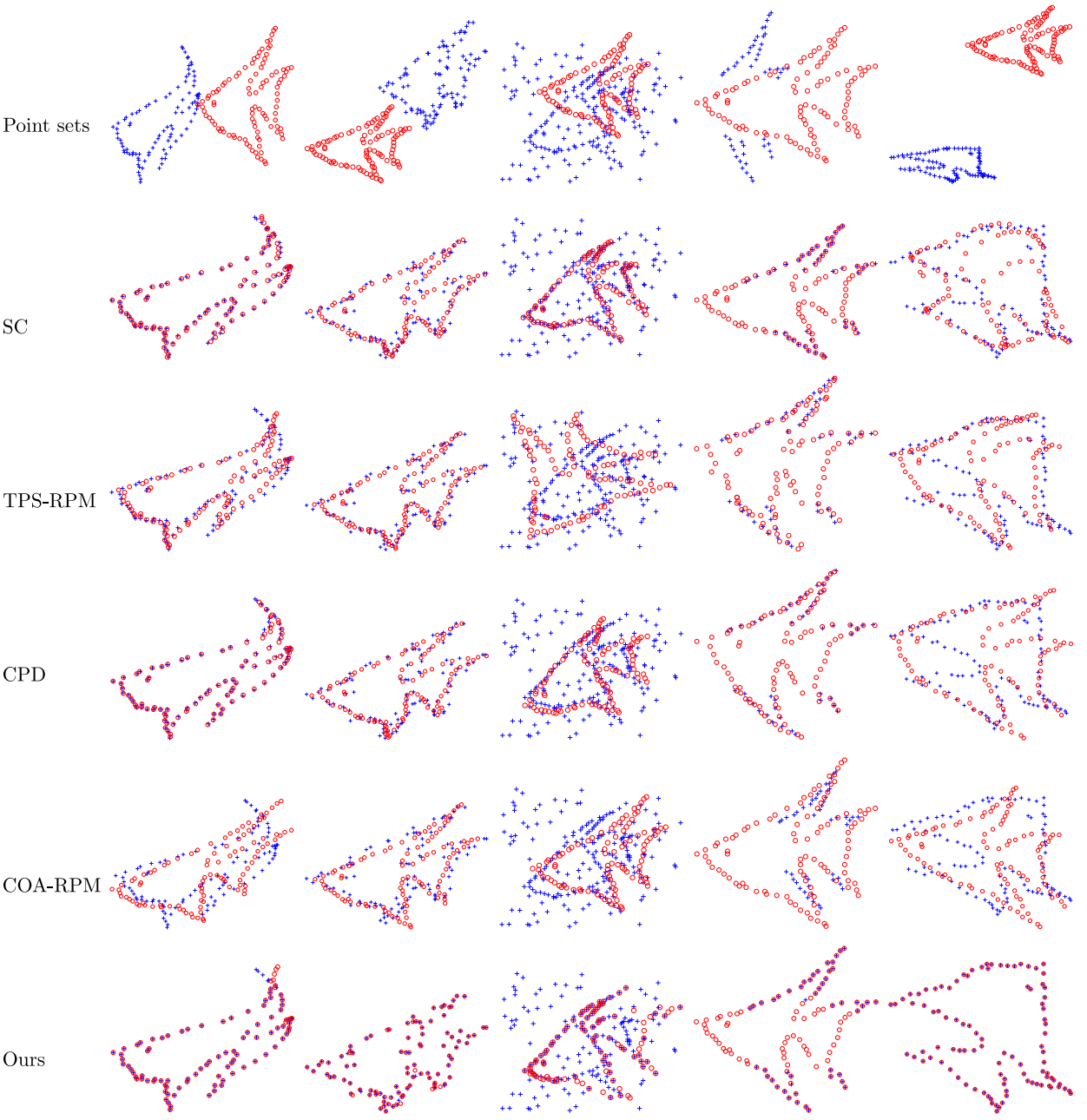

Fig. 3. Registration examples by different methods on the fish shape. The goal is to align the model point set (red circles) onto the target point set (blue pluses). Top row: template and target. From the second row to the bottom row: registration results by SC, TPS-RPM, CPD, COA-RPM, and our method, respectively. From left to right column: the deformation, noise, outlier, occlusion, and rotation tests, respectively. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

our method are large. The main reason that our method is not robust to high proportion of outliers (as shown in Fig. 7) is that the shape context in this situation could not work well and we cannot obtain good initial correspondences, and hence the estimation of transformation will fail. In addition, we also test our method in the case of outliers not satisfying the uniform distribution assumption, as shown in the middle and right columns of Fig. 7. From the results, we see that the uniform distribution assumption also works well when the outliers concentrate in a specific part of the space.

Fig. 6a and b shows the results for the shapes of fish and Chinese character with occlusion, respectively. Our approach performs best, and it consistently outperforms CPD. The error means of our approach are nearly horizontal, showing good robustness to occlusion. Fig. 6c and d

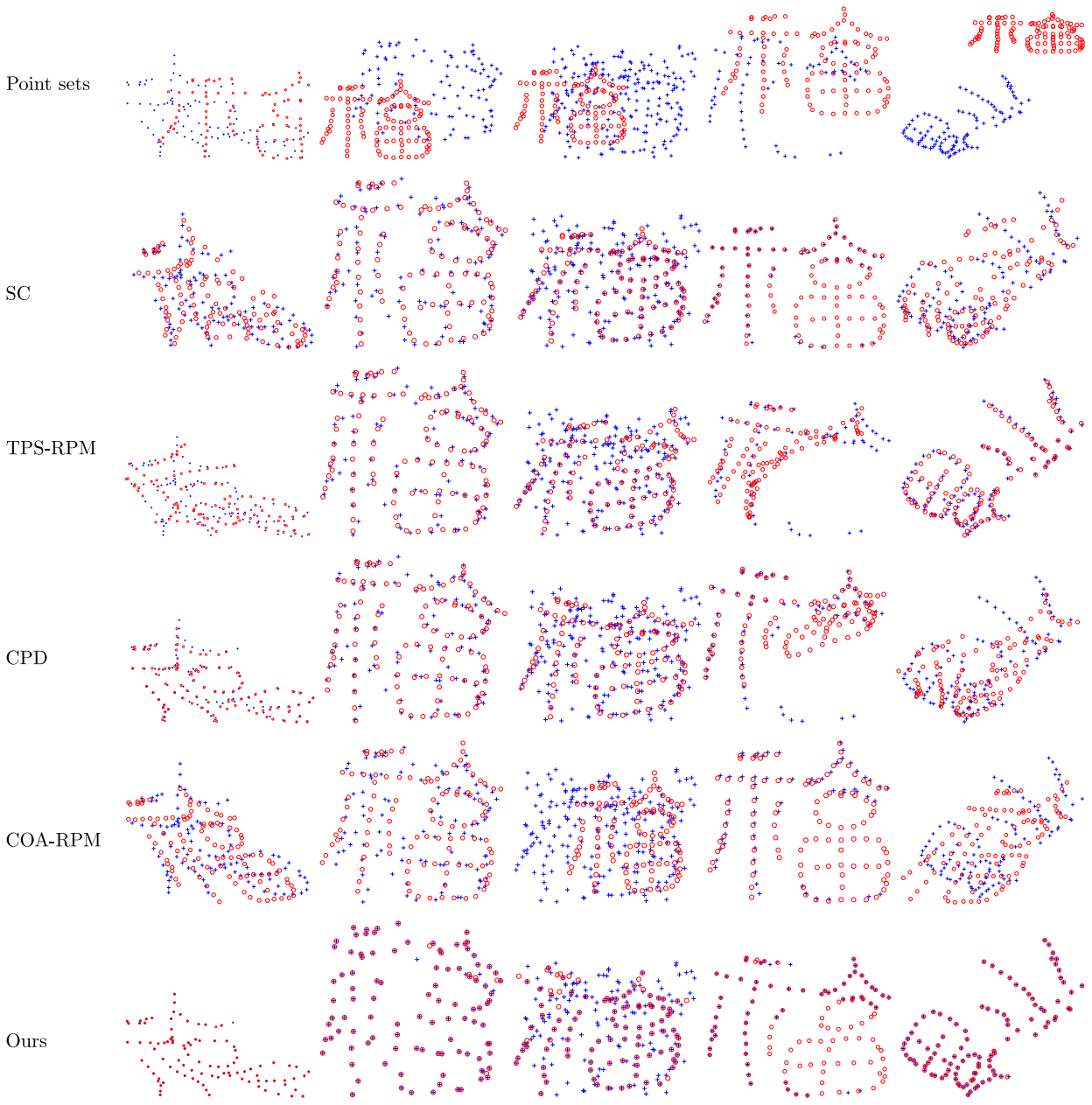

Fig. 4. Registration examples by different methods on the Chinese character shape. The goal is to align the model point set (red circles) onto the target point set (blue pluses). Top row: template and target. From the second row to the bottom row: registration results by SC, TPS-RPM, CPD, COA-RPM, and our method, respectively. From left to right column: the deformation, noise, outlier, occlusion, and rotation tests, respectively. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

shows the results for the shapes of fish and Chinese character with rotation, respectively. The best results are obtained by our method. TPS-RPM, CPD, and COA-RPM, however, are not robust to large rotation.

In summary, our method performs best on the deformation, occlusion, and rotation tests and achieve comparable results for the noise tests except the shapes with outlier. SC gives average error in all experiments. TPS-RPM obtains the best performances on the outlier tests. CPD performs well on the different types of transformations, except for the rotation tests in Fig. 6c and d, where CPD performs significantly worse than our method when the disturbance becomes higher.

4.2. Experiments with MNIST handwritten digit database

In this section, we perform experiments on the MNIST handwritten digit database [24], which include 10 categories corresponding to the 10 digits (from 0 to 9). Each category has 1000 images. The size of each image is

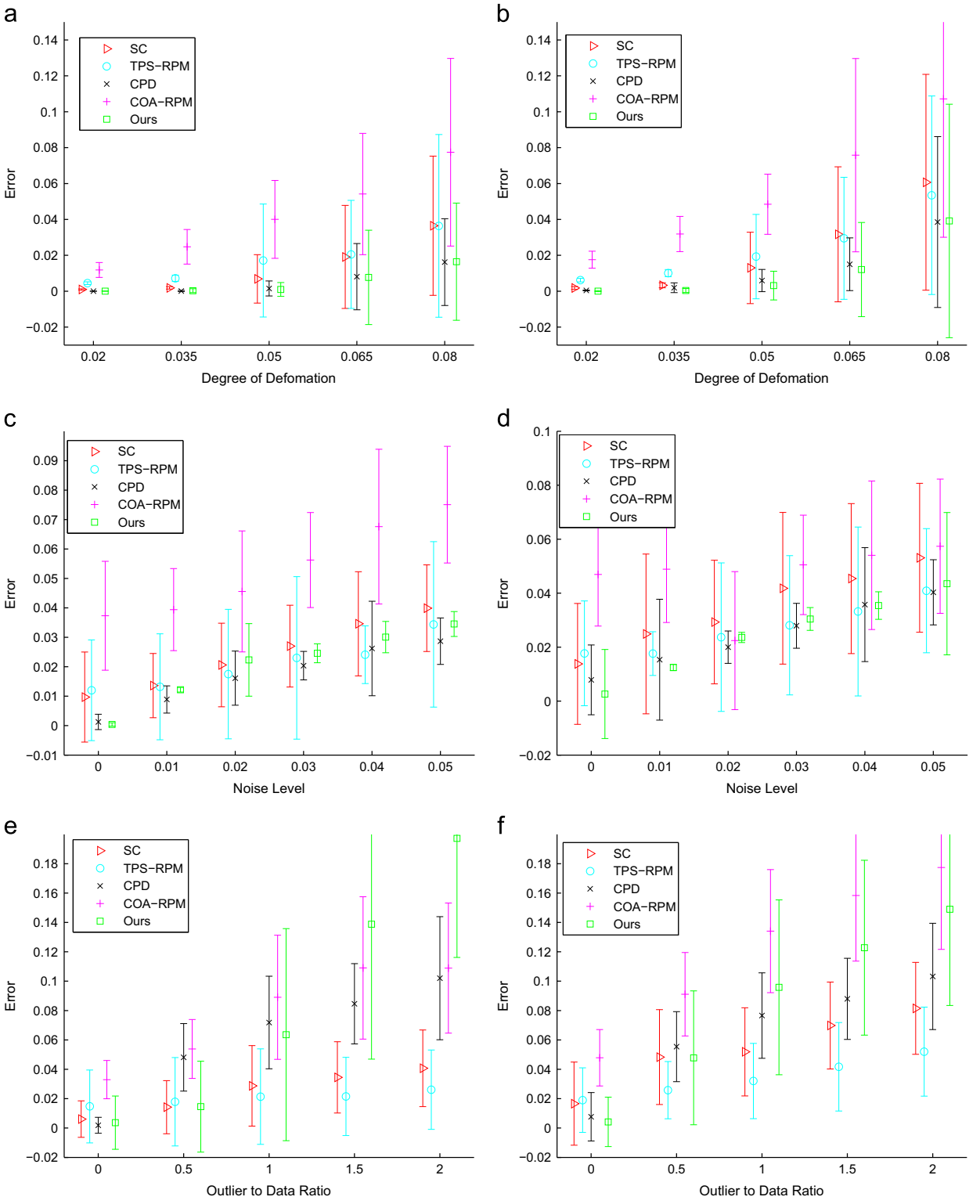

Fig. 5. Comparison of the registration performance of our method with SC, TPS-RPM, CPD, and COA-RPM on the fish and Chinese character shapes [3]. Left column: the shape of fish. Right column: the shape of a Chinese character. The error bars indicate the registration error means and standard deviations over 100 trials.

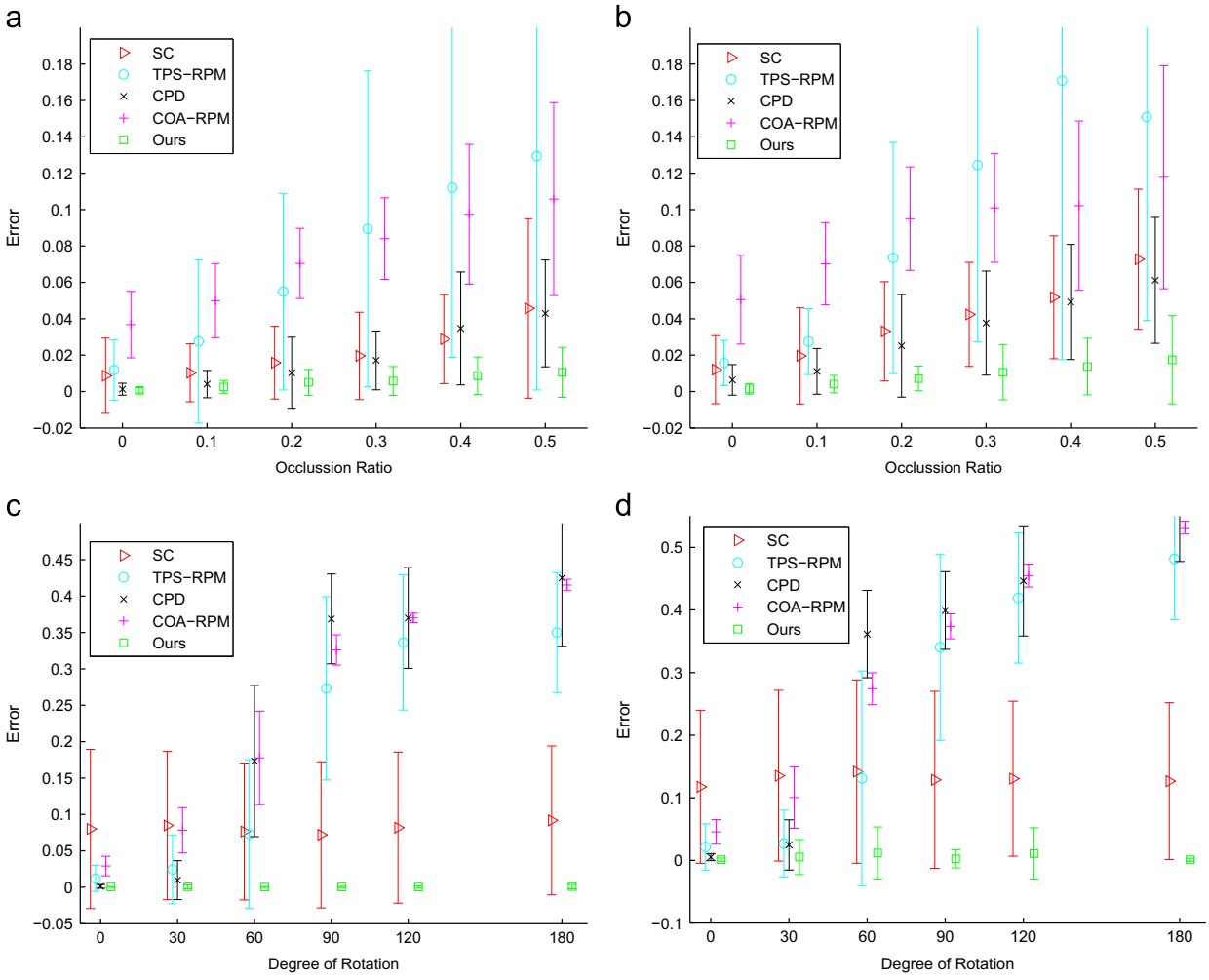

Fig. 6. Comparison of the registration performance of our method with SC, TPS-RPM, CPD, and COA-RPM on the fish and Chinese character shapes [13]. Left column: the shape of fish. Right column: the shape of a Chinese character. The error bars indicate the registration error means and standard deviations over 100 trials.

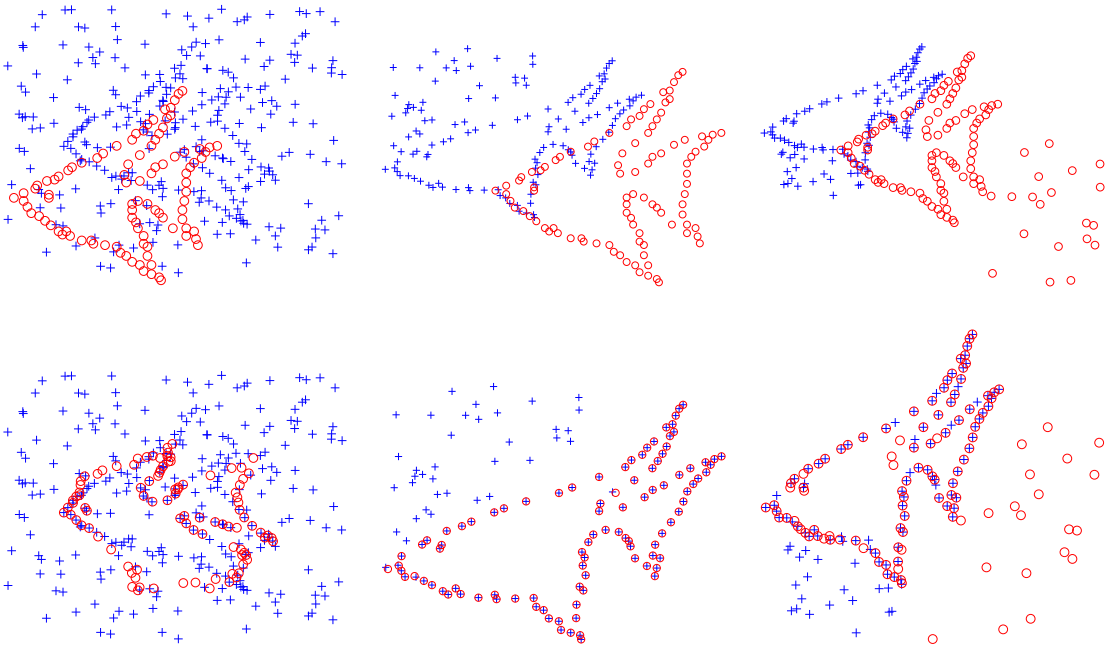

Fig. 7. Registration examples for the outlier test by the proposed method. Top row: template and target. Bottom row: registration results by our method.

28×28. The first sample in each category is used as template and the rest images are used as the targets for matching test. Fig. 8 shows the ten templates. In the experiments, the 48 brightest points were sampled from the Canny edges. The correspondence between these two sets of points is established by the proposed method. Then the estimated correspondence is used to warp the template shape. The average Euclidean distance between a point in the warped template and the corresponding point in the target is used to quantify the accuracy of the registration methods. The performance of the proposed method is compared with three state-of-the-art point matching methods, SC [7], TPS-RPM [3], and CPD [10]. The average error for each registration method is shown in Table 1. It can be seen that our method performs best. The average error of the proposed method is 0.0326 and is less than 0.0595, 0.0511, and 0.0833, respectively, provided by the SC, TPS-RPM, and CPD methods.

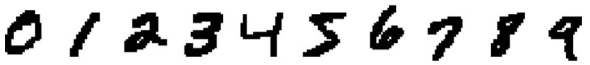

Fig. 8. The first sample in each category used as template.

Table 1
Performance comparison on the MNIST database.

	SC	TPS-RPM	CPD	Ours
Average error	0.0595	0.0511	0.0833	**0.0326**

Table 2
Retrieval results on the articulated shapes dataset [25].

Method	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
ours	**34/40**	**32/40**	**27/40**	**20/40**

4.3. Shape classification results

In this section we test our method on the articulated shape data set [25], which contains 40 images from 8 different objects with articulation. Each object has 5 images articulated to different degrees, as shown in Fig. 9. The dataset is extremely challenging due to the high similarity between different objects. We compare our method to SC [7] and CPD [10]. 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 [18] 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.

4.4. Non-rigid 3D registration

We test our method against CPD [10] on the 3D wolf point sets taken from the non-rigid objects database [26]. Fig. 10 shows the templates and targets point sets and the registration results by the two methods. The average matching error of our method is 0.2691 and is lower than 0.4714, provided by the CPD, and hence the result of our method is much more accurate and outperforms the CPD.

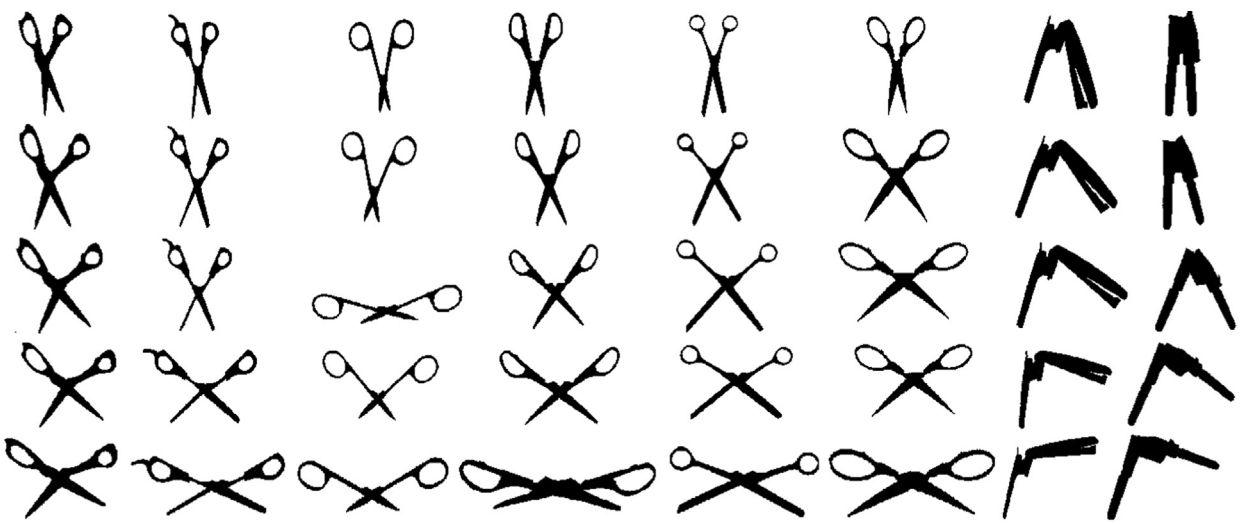

Fig. 9. Articulated shape database. The dataset contains 40 images from 8 articulated objects. Each column has 5 images from the same object.

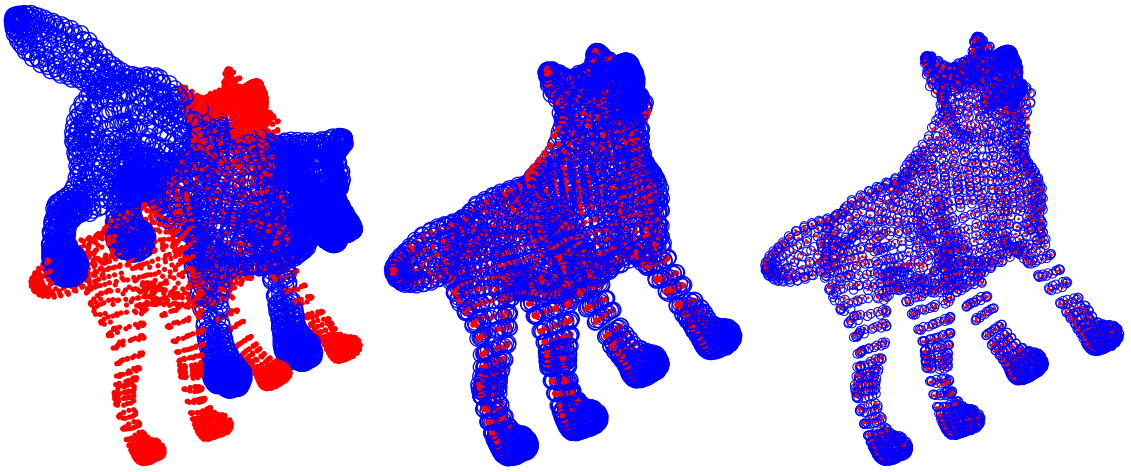

Fig. 10. Registration of 3D wolf point sets. Left column: template and target. Second column: registration result by CPD. Right column: registration result by our method. The average matching error of the CPD and our method are 0.4714 and 0.2691, respectively. Our method outperforms the CPD.

5. Conclusion

In this paper, we have presented a new approach for non-rigid point set registration. A key characteristic of our approach is the estimation of coherent spatial mapping from correspondences based on a robust maximum likelihood estimation which is solved by an iterative EM algorithm. We have to test the proposed approach on various synthetic and real examples and compare it to SC, TPS-RPM, CPD, and COA-RPM. The experimental results show that the proposed approach outperforms the state-of-the-art methods on most scenarios.

Acknowledgments

This work was supported in part by the National Natural Science Fund Committee of the Chinese Academy of Sciences astronomical union funds under Grants U1331113, by the National Natural Science Foundation of China under Grants 61102104, by the Science Foundation of China University of Geosciences under Grant CUGL110243, and by the China Postdoctoral Science Foundation under Grants 2013M540582 and 2014T70702.

References

[1] L.G. Brown, A survey of image registration techniques, ACM Comput. Surv. 24 (4) (1992) 325–376.
[2] P.J. Besl, N.D. McKay, A method for registration of 3-D shapes, IEEE Trans. Pattern Anal. Mach. Intell. 14 (2) (1992) 239–256.
[3] H. Chui, A. Rangarajan, A new point matching algorithm for non-rigid registration, Comput. Vis. Image Understand. 89 (2003) 114–141.
[4] J. Kybic, I. Vnucko, Approximate all nearest neighbor search for high dimensional entropy estimation for image registration, Signal Process. 92 (5) (2012) 1302–1316.
[5] A. Wong, P. Fieguth, Fast phase-based registration of multimodal image data, Signal Process. 89 (5) (2009) 724–737.
[6] A.W. Fitzgibbon, Robust registration of 2D and 3D point sets, Image Vis. Comput. 21 (2003) 1145–1153.
[7] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, IEEE Trans. Pattern Anal. Mach. Intell. 24 (24) (2002) 509–522.
[8] J. Ma, J. Zhao, J. Tian, Z. Tu, A. Yuille, Robust estimation of nonrigid transformation for point set registration, in: IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 2147–2154.
[9] W. Lian, L. Zhang, Robust point matching revisited: a concave optimization approach, in: Proceedings of the European Conference on Computer Vision, 2012.
[10] A. Myronenko, X. Song, Point set registration: coherent point drift, IEEE Trans. Pattern Anal. Mach. Intell. 32 (12) (2010) 2262–2275.
[11] Y. Tsin, T. Kanade, A correlation-based approach to robust point set registration, in: Proceedings of the European Conference on Computer Vision, 2004, pp. 558–569.
[12] J. Ma, J. Chen, D. Ming, J. Tian, A mixture model for robust point matching under multi-layer motion, PloS ONE 9 (3) (2014) e92282.
[13] Y. Zheng, D. Doermann, Robust point matching for nonrigid shapes by preserving local neighborhood structures, IEEE Trans. Pattern Anal. Mach. Intell. 28 (4) (2006) 643–649.
[14] B. Jian, B.C. Vemuri, Robust point set registration using gaussian mixture models, IEEE Trans. Pattern Anal. Mach. Intell. 33 (8) (2011) 1633–1645.
[15] J. Ma, J. Zhao, J. Tian, A.L. Yuille, Z. Tu, Robust point matching via vector field consensus, IEEE Trans. Image Process. 23 (4) (2014) 1706–1721.
[16] J. Ma, J. Zhao, J. Tian, X. Bai, Z. Tu, Regularized vector field learning with sparse approximation for mismatch removal, Pattern Recognit. 46 (12) (2013) 3519–3532.
[17] G. Wahba, Spline Models for Observational Data, SIAM, Philadelphia, PA, 1990.
[18] P.T. Giorgio Carpaneto, Algorithm 548: solution of the assignment problem, ACM Trans. Math. Softw. 6 (1) (1980) 104–111.
[19] J. Zhao, J. Ma, J. Tian, J. Ma, D. Zhang, A robust method for vector field learning with application to mismatch removing, in: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2011, pp. 2977–2984.
[20] A. Dempster, N. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. R. Stat. Soc. Ser. B 39 (1) (1977) 1–38.
[21] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, New York, 2006.
[22] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, 2nd edition, Cambridge University Press, Cambridge, 2003.
[23] A.E. Johnson, M. Hebert, Using spin-images for efficient object recognition in cluttered 3-D scenes, IEEE Trans. Pattern Anal. Mach. Intell. 21 (5) (1999) 433–449.
[24] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.
[25] H. Ling, D.W. Jacobs, Shape classification using the inner-distance, IEEE Trans. Pattern Anal. Mach. Intell. 29 (2) (2007) 840–853.
[26] A. Bronstein, M. Bronstein, R. Kimmel, Numerical Geometry of Non-rigid Shapes, Springer, New York, 2008.