

Non-Rigid Point Set Registration by Preserving Global and Local Structures

Jiayi Ma, Ji Zhao, and Alan L. Yuille, *Fellow, IEEE*

Abstract—In previous work on point registration, the input point sets are often represented using Gaussian mixture models and the registration is then addressed through a probabilistic approach, which aims to exploit global relationships on the point sets. For non-rigid shapes, however, the local structures among neighboring points are also strong and stable and thus helpful in recovering the point correspondence. In this paper, we formulate point registration as the estimation of a mixture of densities, where local features, such as shape context, are used to assign the membership probabilities of the mixture model. This enables us to preserve both global and local structures during matching. The transformation between the two point sets is specified in a reproducing kernel Hilbert space and a sparse approximation is adopted to achieve a fast implementation. Extensive experiments on both synthesized and real data show the robustness of our approach under various types of distortions such as deformation, noise, outliers, rotation and occlusion. It greatly outperforms state-of-the-art methods, especially when the data is badly degraded.

Index Terms—Registration, shape matching, non-rigid, Gaussian mixture model, global/local.

I. INTRODUCTION

Finding an optimal alignment between two sets of points is a fundamental problem in computer vision, image analysis and pattern recognition [6], [3], [8]. Many tasks in these fields – such as stereo matching, image registration and shape recognition – can be formulated as a point set registration problem, which requires addressing two issues: find point-to-point correspondences and estimate the underlying spatial transformation which best aligns the two sets.

The registration problem can be roughly categorized into rigid or non-rigid registration, depending on the application and the form of the data. With a small number of transformation parameters, rigid registration is relatively easy and has been widely studied [6], [3], [27], [11], [21]. By contrast, non-rigid registration is more difficult because the true underlying non-rigid transformations are often unknown and modeling them is a challenging task [8]. But non-rigid registration is very important because it is required for many real world tasks, including shape recognition, deformable motion tracking and medical image registration.

Existing non-rigid registration methods typically consider the alignment of two point sets as a probability density estimation problem and more specifically, they use Gaussian Mix-

ture Models (GMMs) [24], [12]. This makes the reasonable assumption that points from one set are normally distributed around points belonging to the other set. Hence, the point-to-point assignment problem can be recast into that of estimating the parameters of a mixture distribution. In general, these probabilistic methods are able to exploit global relationships in the point sets, since the rough structure of a point set is typically preserved; otherwise, even people cannot find correspondences reliably under arbitrarily large deformations. However, even for non-rigid shapes, the local structures among neighboring points are strong and stable. For example, most neighboring points on a non-rigid shape (e.g., a human face) cannot move independently under deformation due to physical constraints [36]. Furthermore, preserving local neighborhood structures is also very important for people to detect and recognize shapes efficiently and reliably [10]. As a major contribution of this paper, we propose a uniform framework, which exploits both the global and the local structures in the point sets.

More precisely, we formulate point registration as the estimation of a mixture of densities: A GMM is fitted to one point set, such that the centers of the Gaussian densities are constrained to coincide with the other point set. Meanwhile, we use local features such as shape context to assign the membership probabilities of the mixture model, so that both global and local structures can be preserved during the matching process. The non-rigid transformation is modeled in a functional space, called the reproducing kernel Hilbert space (RKHS) [1], in which the transformation function has an explicit kernel representation.

II. RELATED WORK

The iterated closest point (ICP) algorithm [3] is one of the best known point registration approaches. It uses nearest-neighbor relationships to assign a binary correspondence and then uses the estimated correspondences to refine the transformation. Efficient versions of ICP use sampling processes, either deterministic or based on heuristics [27]. The nearest point strategy of ICP can be replaced by soft assignments within a continuous optimization framework, e.g., the TPS-RPM [25], [8]. In the recent past, the point registration is typically solved by probabilistic methods [31], [13], [24], [12]. The Kernel Correlation Based Method [31] models each one of the two point sets as two probability distributions and measures the dissimilarity between the two distributions. It was later improved in [13], which represents the point sets using GMMs. In [24] and [12], the GMM is used to recast the point-to-point

J. Ma is with the Electronic Information School, Wuhan University, Wuhan, 430072, China (e-mail: jyama2010@gmail.com).

J. Zhao is with the Samsung Advanced Institute of Technology, Beijing, 100027, China (e-mail: zhaoji84@gmail.com).

A. L. Yuille is with the Department of Statistics, UCLA, Los Angeles, CA, 90095, USA (e-mail: yuille@stat.ucla.edu).

assignment problem into that of estimating the parameters of a mixture. This is done within the framework of maximum likelihood and the Expectation-Maximization (EM) algorithm [9]. In addition, point registration for articulated shapes has also been investigated using similar methods and shows promising results [22], [12], [29]. The methods mentioned above mainly focus on exploiting the global relationships in the point sets.

To preserve local neighborhood structures, Belongie *et al.* [2] introduced a registration method based on the shape context descriptor, which incorporates the neighborhood structure of the point set, and helps to establish better point correspondences. Using shape context as a lever, Zheng and Doermann [36] proposed a matching method, called the Robust Point Matching-preserving Local Neighborhood Structures (RPM-LNS) algorithm, which was later improved in [15]. Tu and Yuille [32] presented a shape matching method, based on a generative model and informative features. Wasserman *et al.* used local structure information to achieve point cloud registration for registering medical images [33], [34]. Recently, Ma *et al.* introduced Robust Point Matching methods based on local features and a robust L_2E estimator (RPM- L_2E) [18] or Vector Field Consensus [19], which can exploit both global and local structures. However, they find the correspondences and the transformation separately. In this paper, we propose a method which estimates these two variables jointly under the GMM formulation.

III. METHOD

We start by introducing a general methodology for non-rigid point set registration using Gaussian mixture models. We next present the layout of our model. We then explain the way we search for an optimal solution, followed by a description of our fast implementation. We also state the differences between our algorithm and several related state-of-the-art algorithms in detail. Finally, we provide the implementation details.

A. Point Registration and Gaussian Mixtures

We denote by $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_N) \in \mathbb{R}^{D \times N}$ the coordinates of a set of model points and by $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_M) \in \mathbb{R}^{D \times M}$ the coordinates of a set of observed data points, where each point is represented as a $D \times 1$ dimensional column vector. The model points undergo a non-rigid transformation $\mathcal{T} : \mathbb{R}^D \rightarrow \mathbb{R}^D$, and the goal is to estimate \mathcal{T} , which warps the model points to the observed data points, so that the two point sets become aligned. In this paper, we will formulate the point registration as the estimation of a mixture of densities, where a Gaussian mixture model (GMM) is fitted to the observed data points \mathbf{Y} , such that the GMM centroids of the Gaussian densities are constrained to coincide with the transformed model points $\mathcal{T}(\mathbf{X})$ [24], [13], [12].

Let us introduce a set of latent variables $\mathcal{Z} = \{z_m \in \mathbb{N}_{N+1} : m \in \mathbb{N}_M\}$, where each variable z_m assigns an observed data point \mathbf{y}_m to a GMM centroid $\mathcal{T}(\mathbf{x}_n)$, if $z_m = n$, $1 \leq n \leq N$, or to an additional outlier class, if $z_m = N + 1$. The GMM probability density function then can be defined as

$$P(\mathbf{y}_m) = \sum_{n=1}^{N+1} P(z_m = n)P(\mathbf{y}_m | z_m = n). \quad (1)$$

Typically, for point registration, we can use equal isotropic covariances $\sigma^2 \mathbf{I}$ for all GMM components and the outlier distribution is supposed to be uniform $\frac{1}{a}$. Note that a more general assumption of anisotropic Gaussian noise in the data, which uses the full Gaussian model, has also been investigated in [12]. We denote by $\boldsymbol{\theta} = \{\mathcal{T}, \sigma^2, \gamma\}$ the set of unknown parameters, where $\gamma \in [0, 1]$ is the percentage of outliers. Let π_{mn} be the membership probability of the GMM, such that $\sum_{n=1}^N \pi_{mn} = 1$. The mixture model then takes the form

$$P(\mathbf{y}_m | \boldsymbol{\theta}) = \gamma \frac{1}{a} + (1 - \gamma) \sum_{n=1}^N \frac{\pi_{mn}}{(2\pi\sigma^2)^{D/2}} e^{-\frac{\|\mathbf{y}_m - \mathcal{T}(\mathbf{x}_n)\|^2}{2\sigma^2}}. \quad (2)$$

We want to recover the non-rigid transformation \mathcal{T} . The slow-and-smooth model [35], which has been shown to account for a range of motion phenomena, suggests that the prior of \mathcal{T} has the form: $P(\mathcal{T}) \propto e^{-\frac{\lambda}{2}\phi(\mathcal{T})}$, where $\phi(\mathcal{T})$ is a smoothness functional and λ is a positive real number (we will discuss the details of $\phi(\mathcal{T})$ later). Using Bayes rule, we estimate a MAP solution of $\boldsymbol{\theta}$, i.e., $\boldsymbol{\theta}^* = \arg \max_{\boldsymbol{\theta}} P(\boldsymbol{\theta} | \mathbf{Y}) = \arg \max_{\boldsymbol{\theta}} P(\mathbf{Y} | \boldsymbol{\theta})P(\mathcal{T})$. This is equivalent to minimizing the negative log posterior

$$\mathcal{L}(\boldsymbol{\theta} | \mathbf{Y}) = - \sum_{m=1}^M \ln P(\mathbf{y}_m | \boldsymbol{\theta}) - \ln P(\mathcal{T}), \quad (3)$$

where we make the i.i.d. data assumption. The transformation \mathcal{T} will be obtained from the optimal solution $\boldsymbol{\theta}^*$.

B. Registration Using Global and Local Structures

The above method treats the point set as an instance of a GMM distribution, which aims to exploit the global structures in the point sets. For non-rigid point set registration, the local structures among neighboring points are also very strong and stable and can be used to improve the point correspondences. Next, we show how to use the local structures.

Consider the mixture model in Eq. (2), where we need to assign the membership probability π_{mn} that the observed data point \mathbf{y}_m belongs to the GMM centroid $\mathcal{T}(\mathbf{x}_n)$. It is a prior and is typically assumed to be equal for all GMM components [24], [12], i.e., $\pi_{mn} = \frac{1}{N}$, $\forall m \in \mathbb{N}_M, n \in \mathbb{N}_N$. In this paper, we initialize π_{mn} by incorporating the local neighborhood structure information. More specifically, we compute a local feature descriptor for each point [2], [17], and then initialize π_{mn} based on the matching of the feature descriptors of the two sets.

We use shape context [2] as the feature descriptor in the 2D case, using the Hungarian method for matching, with the χ^2 test statistic as the cost measure. In the 3D case, the Fast Point Feature Histograms (FPFH) [28] can be used as a feature descriptor and the matching is performed by a Sample Consensus Initial Alignment method. After we obtain the correspondences between the two point sets by matching their feature descriptors, we use them to initialize π_{mn} according to the following two rules:

- (i) For a data point \mathbf{y}_m , we denote by \mathcal{I} the set of model

points, which it corresponds to¹. We let

$$\pi_{mn} = \begin{cases} \frac{\tau}{|\mathcal{I}|}, & \forall \mathbf{x}_n \in \mathcal{I}, \\ \frac{1-\tau}{N-|\mathcal{I}|}, & \forall \mathbf{x}_n \notin \mathcal{I}, \end{cases} \quad (4)$$

where parameter τ , $0 \leq \tau \leq 1$, could be considered as the confidence of a feature correspondence, and $|\cdot|$ denotes the cardinality of a set.

- (ii) If a data point \mathbf{y}_m does not have a corresponding model point, we use equal membership probabilities for all GMM components, i.e. $\pi_{mn} = \frac{1}{N}$, $\forall n \in \mathbb{N}_N$.

Note that in our formulation the membership probability π_{mn} is not assigned as a prior, since it depends on the data.

C. The EM Algorithm

There are several ways to estimate the parameters of the mixture model, such as the EM algorithm, gradient descent and variational inference. The EM algorithm [9] is a technique for learning and inference in the context of latent variables. It alternates between two steps: an expectation step (E-step) and a maximization step (M-step).

We follow standard notation [4] and omit some terms that are independent of θ . Considering the negative log posterior function, i.e., Eq. (3), the complete-data log posterior (also referred to as the posterior expectation of the complete-data log likelihood) is then given by

$$\begin{aligned} \mathcal{Q}(\theta, \theta^{\text{old}}) &= \frac{1}{2\sigma^2} \sum_{m=1}^M \sum_{n=1}^N P(z_m = n | \mathbf{y}_m, \theta^{\text{old}}) \|\mathbf{y}_m - \mathcal{T}(\mathbf{x}_n)\|^2 \\ &+ \frac{M_{\mathbf{P}} D}{2} \ln \sigma^2 - M_{\mathbf{P}} \ln(1 - \gamma) \\ &- (M - M_{\mathbf{P}}) \ln \gamma + \frac{\lambda}{2} \phi(\mathcal{T}), \end{aligned} \quad (5)$$

where $M_{\mathbf{P}} = \sum_{m=1}^M \sum_{n=1}^N P(z_m = n | \mathbf{y}_m, \theta^{\text{old}}) \leq M$.

E-step: We use the current parameter values θ^{old} to find the posterior distribution of the latent variables. To this end, we first renew the membership probability π_{mn} based on the feature correspondence between the data points \mathbf{Y} and the transformed model points $\mathcal{T}(\mathbf{X})$, as described in Section III-B. We denote by \mathbf{P} the posterior probability matrix of size $M \times N$, where the (m, n) -th element $p_{mn} = P(z_m = n | \mathbf{y}_m, \theta^{\text{old}})$ can be computed by applying Bayes rule:

$$\begin{aligned} p_{mn} &= \frac{P(\mathbf{y}_m | z_m = n, \theta^{\text{old}}) P(z_m = n | \theta^{\text{old}})}{P(\mathbf{y}_m | \theta^{\text{old}})} \\ &= \frac{\pi_{mn} e^{-\frac{\|\mathbf{y}_m - \mathcal{T}(\mathbf{x}_n)\|^2}{2\sigma^2}}}{\sum_{k=1}^N \pi_{mk} e^{-\frac{\|\mathbf{y}_m - \mathcal{T}(\mathbf{x}_k)\|^2}{2\sigma^2}} + \frac{\gamma(2\pi\sigma^2)^{D/2}}{(1-\gamma)^a}}. \end{aligned} \quad (6)$$

Note that the local structure is used throughout the EM process. The posterior probability p_{mn} is a soft assignment, which indicates to what degree the observed data point \mathbf{y}_m coincides with the model point \mathbf{x}_n under the current estimated transformation \mathcal{T} .

¹Typically, \mathcal{I} contains only one point, but multiple correspondences could exist in some cases, for instance, if the matching strategy is not bipartite matching.

M-step: We compute the revised parameters as²: $\theta^{\text{new}} = \arg \max_{\theta} \mathcal{Q}(\theta, \theta^{\text{old}})$. Taking derivatives of $\mathcal{Q}(\theta)$ with respect to γ and σ^2 , and setting them to zero, we obtain

$$\gamma = 1 - M_{\mathbf{P}}/M, \quad (7)$$

$$\sigma^2 = \frac{\sum_{m=1}^M \sum_{n=1}^N p_{mn} \|\mathbf{y}_m - \mathcal{T}(\mathbf{x}_n)\|^2}{M_{\mathbf{P}} D}. \quad (8)$$

The estimate of the variance σ^2 with a large initial value is conceptually similar to deterministic annealing [8], which uses the solution of an easy (e.g., smoothed) problem to recursively provide the initial conditions to increasingly harder problems, but differs in several aspects (e.g., by not requiring any annealing schedule).

Next we consider the terms of $\mathcal{Q}(\theta)$ that are related to \mathcal{T} . We define the transformation as the initial position plus a displacement function \mathbf{v} : $\mathcal{T}(\mathbf{X}) = \mathbf{X} + \mathbf{v}(\mathbf{X})$, and then obtain a regularized risk functional as:

$$\mathcal{E}(\mathbf{v}) = \frac{1}{2\sigma^2} \sum_{m=1}^M \sum_{n=1}^N p_{mn} \|\mathbf{y}_m - \mathbf{x}_n - \mathbf{v}(\mathbf{x}_n)\|^2 + \frac{\lambda}{2} \phi(\mathbf{v}). \quad (9)$$

This is a particular form of Tikhonov regularization [30], and the first term can be seen as a weighted empirical error. Thus the maximization of \mathcal{Q} with respect to \mathcal{T} is equivalent to minimizing the regularized risk functional (9). To proceed, we model \mathbf{v} by requiring it to lie within a vector-valued reproducing kernel Hilbert space (RKHS) \mathcal{H} defined by a diagonal Gaussian matrix kernel $\Gamma : \mathbb{R}^D \times \mathbb{R}^D \rightarrow \mathbb{R}^{D \times D}$ [20], where $\Gamma(\mathbf{x}_i, \mathbf{x}_j) = \exp\{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\beta^2}\} \mathbf{I}$. A brief introduction of the vector-valued RKHS is given in Appendix A. By using the square norm for the smoothness functional $\phi(\mathbf{v})$, i.e., $\phi(\mathbf{v}) = \|\mathbf{v}\|_{\mathcal{H}}^2$, we then have the following representer theorem [23], and the proof is given in Appendix B.

Theorem 1. *The optimal solution of the regularized risk functional (9) has the form*

$$\mathbf{v}(\mathbf{x}) = \sum_{n=1}^N \Gamma(\mathbf{x}, \mathbf{x}_n) \mathbf{c}_n. \quad (10)$$

with the coefficient set $\{\mathbf{c}_n : n \in \mathbb{N}_N\}$ determined by the following linear system:

$$\mathbf{C}(\Gamma d(\mathbf{1}^T \mathbf{P}) + \lambda \sigma^2 \mathbf{I}) = \mathbf{Y} \mathbf{P} - \mathbf{X} d(\mathbf{1}^T \mathbf{P}), \quad (11)$$

where $\mathbf{C} = (\mathbf{c}_1, \dots, \mathbf{c}_N) \in \mathbb{R}^{D \times N}$ is the coefficient set, $\Gamma \in \mathbb{R}^{N \times N}$ is the so-called Gram matrix with the (i, j) -th element $\Gamma_{ij} = \exp\{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\beta^2}\}$, $\mathbf{1}$ is a column vector of all ones, and $d(\cdot)$ is the diagonal matrix.

Note that the M-step is different from the standard EM algorithm, where the free parameters are the means and the covariances of the Gaussian mixture, and the estimation of these parameters is quite straightforward. Our problem requires multiple conditional maximization steps and thus the

²Note that the parameters contained in θ cannot be updated simultaneously because of their dependencies on σ^2 and \mathcal{T} , as can be seen from Eqs. (8) and (11). However, here we use the symbol θ instead of $\{\mathcal{T}, \sigma^2, \gamma\}$ for the sake of simplicity.

Algorithm 1: The PR-GLS algorithm

Input: Two point sets \mathbf{X} and \mathbf{Y} , parameters γ, β, λ
Output: Optimal transformation \mathcal{T}

- 1 Construct the Gram matrix $\mathbf{\Gamma}$;
- 2 Initialization: $\mathbf{C} = \mathbf{0}$, $\sigma^2 = \frac{1}{DMN} \sum_{m,n=1}^{M,N} \|\mathbf{y}_m - \mathbf{x}_n\|^2$;
- 3 Compute feature descriptors for point set \mathbf{Y} ;
- 4 **repeat**
- 5 *E-step:*
- 6 Compute feature descriptors for point set $\mathcal{T}(\mathbf{X})$;
- 7 Establish correspondence between $\mathcal{T}(\mathbf{X})$ and \mathbf{Y} ;
- 8 Initialize π_{mn} based on the feature correspondence;
- 9 Update posterior probability \mathbf{P} by Eq. (6);
- 10 *M-step:*
- 11 Update \mathbf{C} by solving the linear system (11);
- 12 Update $\mathcal{T}(\mathbf{X}) = \mathbf{X} + \mathbf{C}\mathbf{\Gamma}$;
- 13 Update γ and σ^2 by Eqs. (7) and (8);
- 14 **until** \mathcal{Q} converges;
- 15 The transformation \mathcal{T} is determined as $\mathcal{T}(\mathbf{X}) = \mathbf{X} + \mathbf{C}\mathbf{\Gamma}$.

algorithm is referred as the Expectation Conditional Maximization (ECM), which was first introduced into the point registration problem by Horaud *et al.* [12]. Since our non-rigid Point set Registration algorithm exploits both Global and Local Structures, we call it PR-GLS and summarize it in Algorithm 1.

D. Fast Implementation

Solving for the transformation \mathcal{T} requires solving the linear system (11). But for large values of N , this may pose a serious problem because this has $O(N^3)$ computational complexity. Even if this is tractable, we may prefer a suboptimal, but faster method. In this section, we provide a fast implementation based on an idea related to the Subset of Regressors method [26], [20].

Instead of searching for the optimal solution (10), we use a sparse approximation and randomly pick a subset containing K ($K \ll N$) model points $\{\tilde{\mathbf{x}}_k : k \in \mathbb{N}_K\}$. Only points in this subset are allowed to have nonzero coefficients in the expansion of the solution, as in [26], [20]. Our experiments showed that simply selecting an arbitrary subset of the training inputs at random does not perform any worse than more sophisticated methods. Therefore, we seek a solution of the form:

$$\mathbf{v}(\mathbf{x}) = \sum_{k=1}^K \Gamma(\mathbf{x}, \tilde{\mathbf{x}}_k) \mathbf{c}_k, \quad (12)$$

with the coefficients $\tilde{\mathbf{C}} = (\mathbf{c}_1, \dots, \mathbf{c}_K) \in \mathbb{R}^{D \times K}$ determined by a linear system

$$\tilde{\mathbf{C}}(\mathbf{U}^T d(\mathbf{1}^T \mathbf{P}) \mathbf{U} + \lambda \sigma^2 \tilde{\mathbf{\Gamma}}) = \mathbf{Y} \mathbf{P} \mathbf{U} - \mathbf{X} d(\mathbf{1}^T \mathbf{P}) \mathbf{U}, \quad (13)$$

where $\tilde{\mathbf{\Gamma}}$ is the Gram matrix $\tilde{\mathbf{\Gamma}} \in \mathbb{R}^{K \times K}$, where $\tilde{\Gamma}_{ij} = \exp\{-\frac{\|\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j\|^2}{2\beta^2}\}$, and $\mathbf{U} \in \mathbb{R}^{N \times K}$ with the (i, j) -th element $U_{ij} = \exp\{-\frac{\|\mathbf{x}_i - \tilde{\mathbf{x}}_j\|^2}{2\beta^2}\}$. By using this sparse approximation, the time and space complexities for solving the linear system (11) are reduced from $O(N^3)$ and $O(N^2)$ to $O(K^2 N)$ and $O(KN)$, respectively.

E. Related Non-Rigid Registration Methods

The most relevant non-rigid point sets registration algorithm to ours is coherent point drift (CPD) [24], as both algorithms use the GMM formulation and Gaussian radial basis functions to parameterize the transformations. However, our PR-GLS has three major advantages compared to CPD. Firstly, CPD ignores the local structure information in the point sets; it simply uses equal membership probabilities in the mixture model, while our PR-GLS tries to preserve local structures by encouraging the matching of points with similar neighborhood structures, e.g., shape features. Secondly, the outlier percentage γ is fixed and set manually in CPD. This may be problematic in many real-world problems, where the amount of outliers is not known in advance. By contrast, our PR-GLS treats γ as an unknown parameter and estimates it for each EM iteration. Thirdly, CPD can not handle large rotations, e.g., rotations with angles larger than 60° . We solve this problem by choosing rotation invariant features in our formulation.

The two other most relevant non-rigid point sets registration algorithms to ours are RPM-LNS [36] and RPM- L_2E [18]. More specifically, they also use local structure information. RPM-LNS formulates point matching as a graph matching problem. RPM- L_2E establishes point correspondences by first constructing a large putative set of correspondence and then uses a robust L_2E estimator to get rid of the outliers. Both of the two algorithms use shape context to initialize the correspondences. One major drawback of RPM-LNS and RPM- L_2E is that they find the correspondence and the transformation separately, while our method estimates these two variables jointly using the GMM formulation. Moreover, RPM-LNS is not applicable to the 3D case.

In summary, our main contribution is that we propose an efficient framework for preserving both global and local structures during matching. Compared to other related methods, the efficient use of global and local information in our PR-GLS can lead to big improvements under various degenerations (see Fig. 3).

F. Implementation Details

The performance of point matching algorithms typically depends on the coordinate system in which the points are expressed. We use data normalization to control this. More specifically, we perform a linear scaling of the coordinates, so that the points in both of the two sets both have zero mean and unit variance.

To assign the membership probability π_{mn} , we need to establish initial correspondences which is usually time consuming. Fortunately, the deformation of the model shape is gradually reduced during the EM iterations and as such, the correspondences between the two shapes will not change too fast. Therefore, to speed-up our algorithm, we update π_{mn} in every 10 iterations rather than in each iteration.

Parameter settings. There are three main parameters in our PR-GLS algorithm: γ, β and λ . Parameter γ reflects our assumption on the amount of outliers in the point sets. Parameters β and λ reflect the amount of the smoothness constraint, where β determines the width of the range of

the interaction between samples and λ controls the trade-off between the closeness to the data and the smoothness of the solution. In general, we found that our method was robust to parameter changes. Parameter τ is used to initialize the membership probabilities, which we will discuss in the experiments. Moreover, the uniform distribution parameter a is set to be the volume of the bounding box of the data.

Analysis of convergence. The objective function (3) is not convex, so it is unlikely that any algorithm can find its global minimum. However, a stable local minimum is often enough for many practical applications. To this end, our strategy is to initialize the variance σ^2 with a large initial value and then use the EM algorithm. If the value of σ^2 is large enough, the objective function becomes convex in a large region which can filter out a lot of unstable shallow local minima. Hence we are likely to find a good minimum for large variance. As σ^2 decreases, the objective function tends to change smoothly, which makes it likely that using the old minimum as the initial value could be helpful to converge to a new good minimum. Therefore, as the iterations continue, we have a good chance of reaching a stable local minimum. This is conceptually similar to deterministic annealing [8], which uses the solution of an easy problem to recursively give initial conditions to increasingly harder problems.

IV. EXPERIMENTAL RESULTS

In order to evaluate the performance of our algorithm, we conducted three types of experiments: i) non-rigid point set registration for 2D shapes; ii) non-rigid point set registration for 3D shapes; iii) sparse feature correspondence on real images. The experiments were performed on a laptop with 2.5 GHz Intel[®] Core[™]2 Duo CPU, 8 GB memory and Matlab Code.

A. Results on 2D Non-Rigid Shape Matching

We tested our PR-GLS algorithm on the same synthesized data as in [8] and [36]. The data consists of two models with different shapes, where the first model consists of 96 points representing a *fish* and the second model consists of 108 points, representing a *Chinese character*. For each model, there are five sets of data designed to measure the robustness of registration algorithms with respect to different degrees of deformation, noise, outliers, rotation and occlusion. In the deformation test, the deformation is generated by using Gaussian radial basis functions with coefficients sampled from a Gaussian distribution with zero mean and a standard deviation ranging from 0.02 to 0.08. In the noise test, Gaussian noise is added with standard deviation ranging from 0 to 0.05. In the outlier test, random outliers are added with the outlier to original data ratio ranging from 0 to 2. In the rotation test, six rotation angles are used: 0, 30, 60, 90, 120, and 180. In the occlusion test, some parts of the data are removed with the removed data to original data ratio ranging from 0 to 0.5. In each test, one of the above distortions is applied to the model set to create an observed data set, and 100 samples are generated for each degradation level. We use shape contexts as the feature descriptor to establish the

initial correspondences. It is easy to make the shape context feature descriptors translation and scale invariant and in some applications, rotation invariance is also required. We use the rotation invariant shape context as in [36].

Generally, the parameter τ in Eq. (4) reflects the accuracy of the initial correspondences, which to some extent depends on the degree of the distortion. In these experiments we determine τ adaptively, based on the data. To this end, we first define a registration error ϵ^* , which is independent of the ground truth:

$$\epsilon^* = \left(\frac{1}{|\mathcal{S}|} \sum_{(n,m) \in \mathcal{S}} \|\mathcal{T}(\mathbf{x}_n) - \mathbf{y}_m\|^2 \right)^{1/2}, \quad (14)$$

where \mathcal{S} consists of the $\min(N, M)$ matches with the largest posterior matching probabilities, i.e., p_{mn} . For each set of data, we randomly choose 5 of the 100 samples and compute their average registration error with $\tau = 0.1, 0.3, 0.5, 0.7$ and 0.9 . Afterwards, we set τ to be the value which achieves the smallest registration error ϵ^* . Typically, $\tau \approx 0.9$ works well for small distortions and 0.3 for large distortions. For the parameters γ , β and λ , we set them to 0.1 , 2 and 3 , respectively, throughout the experiments.

The registration progress on the fish shape is illustrated schematically in Fig. 1. The columns show the iterative progress, and each row provides a different type of distortion. The goal is to align the model point set (blue pluses) onto the observed data point set (red circles). From the results, we see that our RPM-GLS registration is robust and accurate, and it typically converges in about 30 iterations. More qualitative results of our method on the two shape models are presented in Fig. 2. We organize the results in every two rows: the upper row presents the data, while the lower row shows the registration result. For the *fish* model, the shape is relatively simple and discriminative and as such, both the global and the local structures are preserved well, even under large degradations. Our method in this case produces almost perfect alignments, as shown in the left of Fig. 2. For the *Chinese character* model, the points are spread out on the shape and the local structure is not preserved as well under large degradations. The matching performance then degrades gradually, but it remains acceptable, even for large degradations. Let us consider the rotation test in (a4, b3) of Fig. 2. It is interesting that our method can handle the fold-back contained in the data shape, which is a desirable property in real-world applications. Moreover, in some extreme cases, such as the deformation test in (a1, b3), as well as the outlier test in (a3, b1), it is hard to establish visual correspondences even by human beings, however, our method can still achieve accurate alignments. The average runtime of our method on 2D shape matching with about 100 points is about 0.4 seconds.

To provide a quantitative comparison, we report the results of six state-of-the-art algorithms, such as shape context (SC) [2], TPS-RPM [8], RPM-LNS [36], GMMREG [13], CPD [24], and RPM- L_2E [18] which were implemented using publicly available codes. The registration error of a pair of shapes is quantified as the average Euclidean distance between a point in the warped model shape and the ground truth corresponding point in the observed data shape (note that it

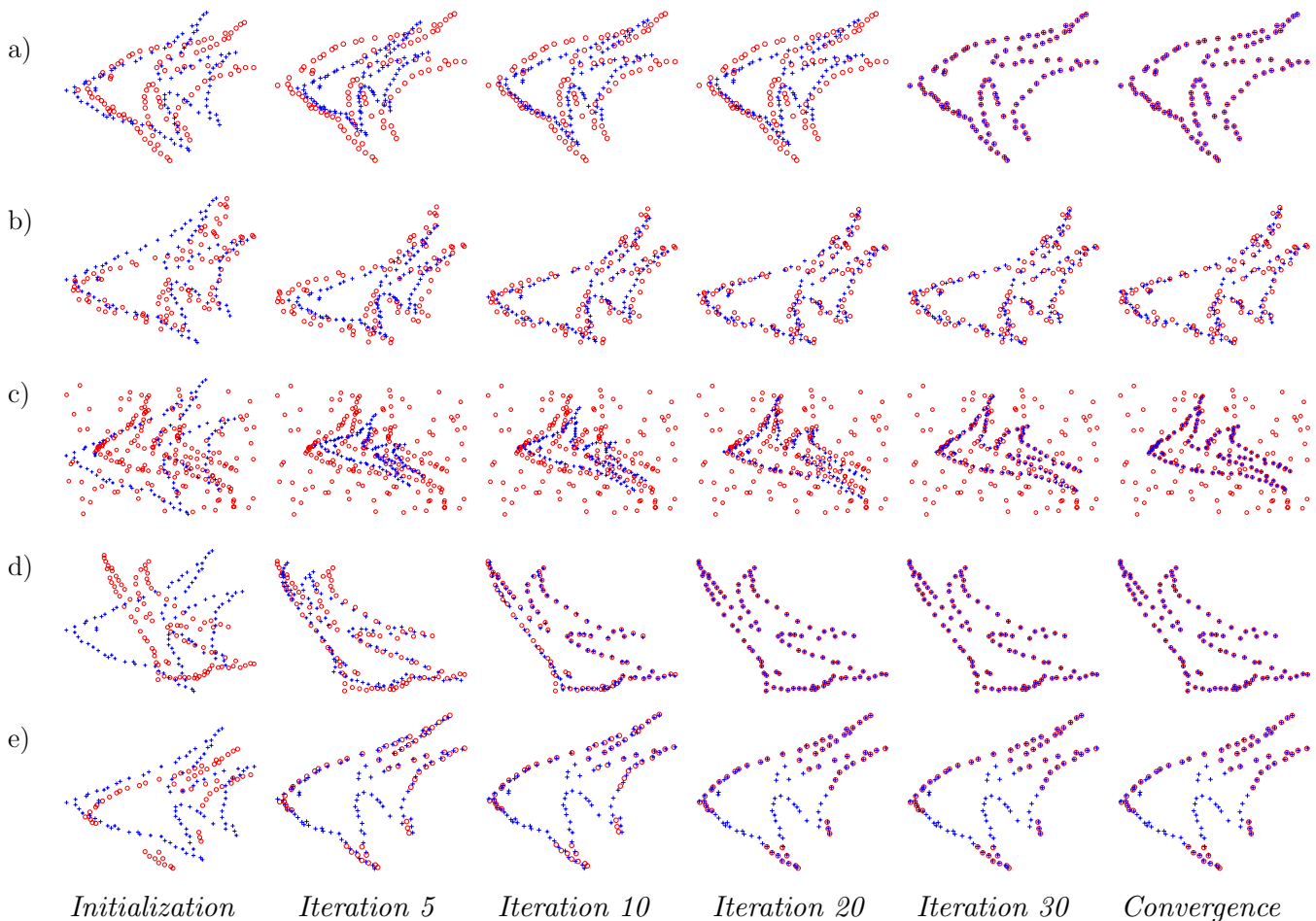


Fig. 1. Schematic illustration of our method for non-rigid point set registration. The goal is to align the model point set (blue pluses) onto the observed data point set (red circles). The columns show the iterative alignment progress. From top to bottom: results on data degraded by deformation, noise, outliers, rotation and occlusion, respectively. Our RPM-GLS registration is robust and accurate in all experiments.

is different from ϵ^* defined in Eq. (14)). Then the registration performance of each algorithm is compared by the mean and standard deviation of the registration error of all the 100 samples for each distortion level. The statistical results, error means, and standard deviations for each setting are summarized in Fig. 3.

In the deformation test results, seven algorithms achieve similar performance in both the *fish* and the *Chinese character* datasets at low deformation levels. But our PR-GLS outperforms them, especially for large degrees of deformation. In the noise test results, we observe that all algorithms are affected by this type of distortion. Still, our PR-GLS achieves the best registration performance. In the outlier test results, GMMREG is not robust to outliers. RPM- L_2E is easily confused and starts to fail once the outlier level becomes relatively high. This is due to that it solves the correspondence and the transformation problems separately. When the outlier level becomes too high, it cannot establish good correspondences in the first step and hence, the final solution will be badly degraded. In contrast, our PR-GLS shows much more robustness, regardless of the outlier level. Note that the original CPD is not robust to changes in the outlier ratios, as it involves a fixed outlier ratio parameter, i.e., parameter γ in Eq. (2). To make a fair

comparison, we set it adaptively similar to the setting of parameter τ in our PR-GLS. In the rotation test, TPS-RPM, GMMREG and CPD degenerate rapidly when the rotation angle is larger than 60° , while RPM- L_2E and our PR-GLS are not affected, because they are using the rotation invariant shape context. In the occlusion test, the registration error of TPS-RPM is the largest, and our PR-GLS again has the best results in most cases once again. We have also tested the fast version of PR-GLS on this dataset and achieved similar performance, as shown in the last column of each test in Fig. 3.

In our PR-GLS, we assign the membership probability π_{mn} based on shape context feature, so that the local structure information can be used to achieve better performance. To demonstrate the benefits for doing this, we perform experiments on the fish and Chinese character datasets with different degrees of deformations. The results are given in Tables I and II, where CPD assumes equal GMM components ($\pi_{mn} = 1/N$), PR-GLS assigns π_{mn} based on local shape features, and GT assigns π_{mn} by using the ground truth correspondences. Clearly, a good initialization of π_{mn} can improve registration performance, especially when the data is badly degraded.

In conclusion, our PR-GLS is effective for various non-rigid point set registration problems, regardless of large deforma-

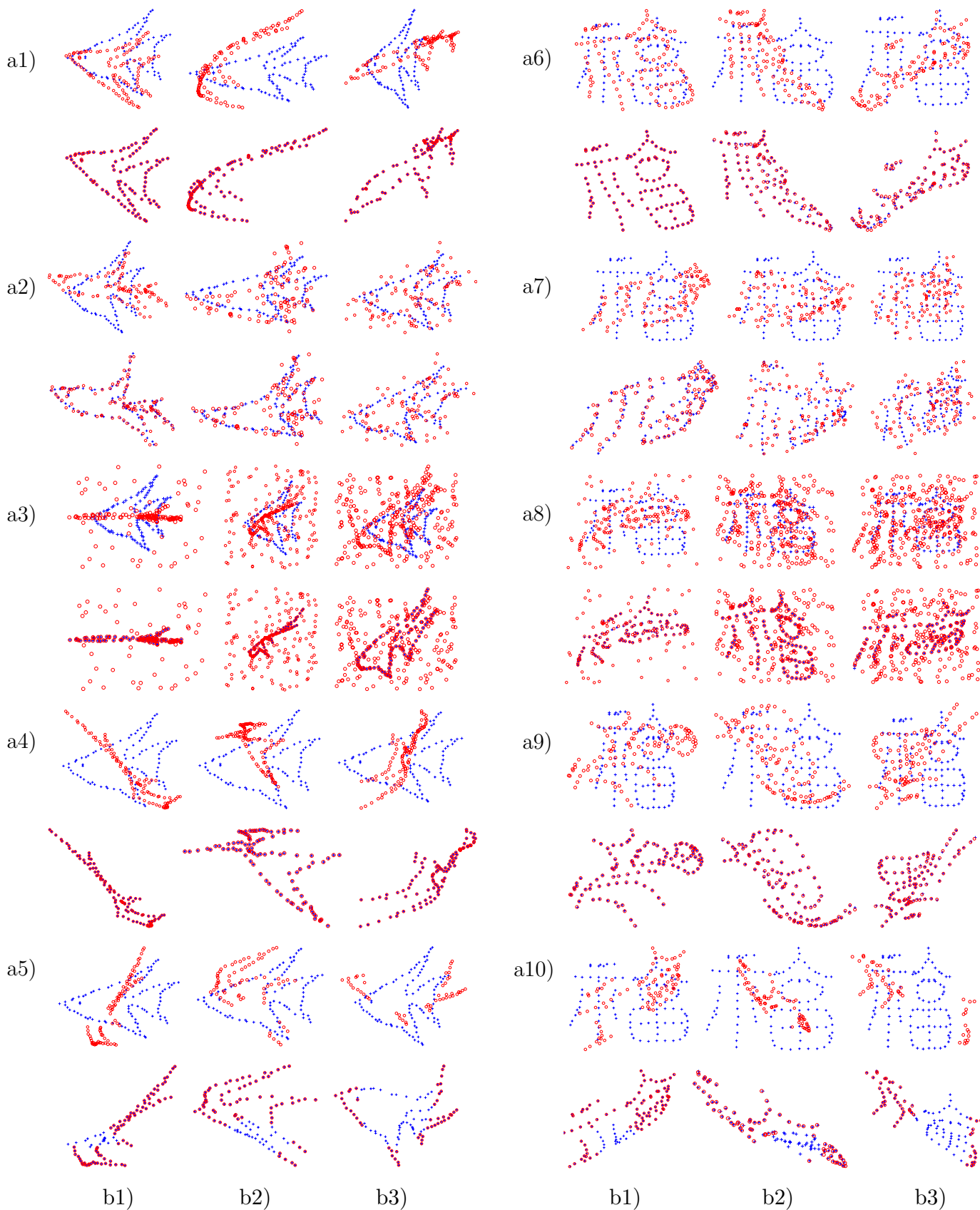


Fig. 2. Point set registration results of our method on the *fish* (left) and *Chinese character* (right) shapes [8], [36], with deformation, noise, outliers, rotation and occlusion presented in every two rows. The goal is to align the model point set (blue pluses) onto the observed data point set (red circles). For each group, the upper figure is the model and data point sets, and the lower figure is the registration result. From left to right, increasing degree of degradation.

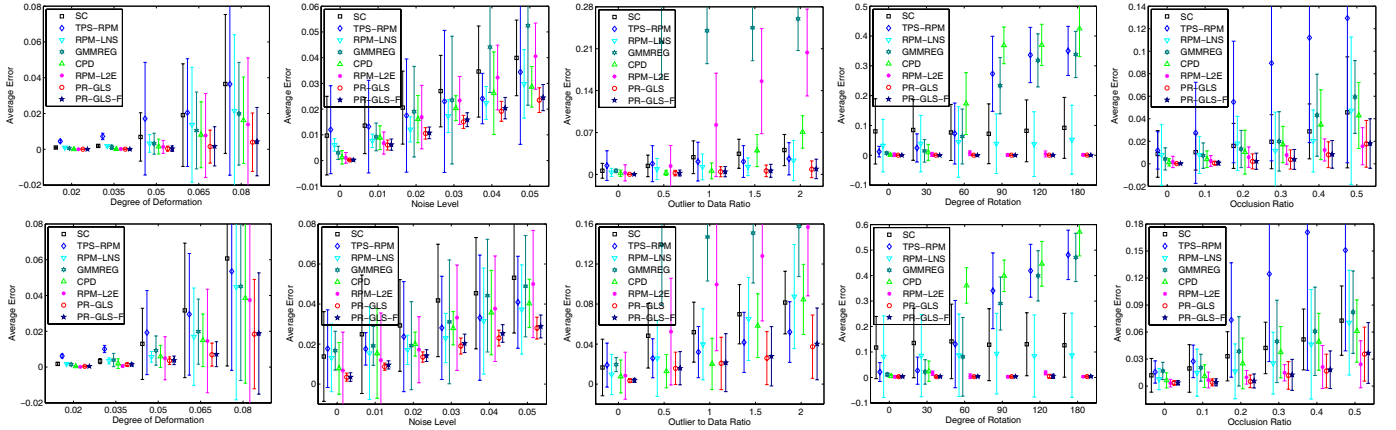


Fig. 3. Comparison of PR-GLS and its fast version (PR-GLS-F) with SC [2], TPS-RPM [8], RPM-LNS [36], GMMREG [13], CPD [24] and RPM- L_2E [18] on the *fish* (top) and *Chinese character* (bottom) [8], [36]. The error bars indicate the registration error means and standard deviations over 100 trials. Note that the variance bar can reach the negative y-axis, where large variance indicates that a method sometimes works well and sometimes totally fails.

TABLE I

AVERAGE REGISTRATION ERROR ON THE FISH DATASETS WITH DIFFERENT DEGREES OF DEFORMATIONS (DoD) OVER 100 TRIALS. THE LAST ROW (GT) USES THE GROUND TRUTH CORRESPONDENCES TO INITIALIZE π_{ij} .

DoD	0.02	0.035	0.05	0.065	0.08
CPD	2.6×10^{-5}	1.3×10^{-4}	1.5×10^{-3}	8.1×10^{-3}	1.6×10^{-2}
PR-GLS	2.5×10^{-5}	7.3×10^{-5}	3.6×10^{-4}	1.5×10^{-3}	4.0×10^{-3}
GT	2.5×10^{-5}	7.3×10^{-5}	2.1×10^{-4}	5.1×10^{-4}	1.0×10^{-3}

TABLE II

AVERAGE REGISTRATION ERROR ON THE CHINESE CHARACTER DATASETS WITH DIFFERENT DEGREES OF DEFORMATIONS (DoD) OVER 100 TRIALS. THE LAST ROW (GT) USES THE GROUND TRUTH TO INITIALIZE π_{ij} .

DoD	0.02	0.035	0.05	0.065	0.08
RPM- L_2E	4.3×10^{-4}	1.9×10^{-3}	5.9×10^{-3}	1.5×10^{-2}	3.8×10^{-2}
PR-GLS	4.3×10^{-4}	1.4×10^{-3}	3.6×10^{-3}	7.0×10^{-3}	1.8×10^{-2}
GT	4.3×10^{-4}	1.4×10^{-3}	3.5×10^{-3}	5.9×10^{-3}	9.1×10^{-3}

tion, noise, outliers, rotation or occlusion.

B. Results on 3D Non-Rigid Shape Matching

We next test our PR-GLS algorithm for 3D shape matching. We adopt the FPFH [28] to determine the initial correspondence. For the dataset, we use two pairs of point sets, representing a wolf in different poses from a surface correspondence benchmark³ [14], as shown in the left two columns of Fig. 4. The ground truth correspondences are supplied by the dataset. Some other challenging benchmark datasets with known ground truth correspondences have also been developed [5], [16]. To assign the membership probability π_{mn} , parameter τ can be determined adaptively according to Eq. (14); here, we fix it to 0.9 for efficiency. Moreover, 3D datasets typically contain thousands of points (e.g., about 5,000 in the *wolf* shape), which may pose a serious problem

³Available at http://www.cs.princeton.edu/~vk/projects/CorrsBlended/doc_data.php.

due to heavy computational or memory requirements; here, we use the fast version of our method with $K = 50$ in Eq. (12).

Fig. 4 presents the registration results of PR-GLS. In the left two columns, the data contains large rotations and slight pose changes, and our method gives almost perfect alignments. In the middle two and right two columns, we also test our method on occlusion and outliers. In each group, the first set contains a degradation on only one set of points, while the second set contains degradations on both the two point sets. The last column demonstrates a challenging example, where we biased the outliers to the different sides of point sets, and the outliers were generated randomly from a normal distribution rather than a uniform distribution. The results demonstrate that our method is quite robust and is not affected by all these degradations. The average runtime of our method with the fast implementation on 3D shape matching with about 5,000 points is about 1 minute. Note that in this experiment the PPFH is quite efficient and requires only 1.5 seconds approximately to establish the initial correspondences.

A quantitative comparison of our method with CPD and RPM- L_2E is reported in Table III. In addition, we also compare with a recent method using local structure information based on Tensor Fields (TF) [34]. In the case of CPD, we rotate the data, so that the two point sets have roughly similar orientations, since CPD is not robust to large rotations. As shown in Table III, our PR-GLS consistently outperforms the other three algorithms in every case. In conclusion, PR-GLS is effective for non-rigid registration, in both the 2D and the 3D cases.

C. Results on Image Feature Correspondence

In this section, we test our PR-GLS on real image data, focusing on establishing visual correspondences between two images. Here the images contain some deformable objects and consequently, the underlying relationships between the images are non-rigid. We use SIFT [17] for feature detection and description. The goal is to find correspondences/matches between two sets of sparse feature points $\{\mathbf{x}_n : n \in \mathbb{I}N_N\}$ and $\{\mathbf{y}_m : m \in \mathbb{I}N_M\}$, with corresponding feature descriptors.

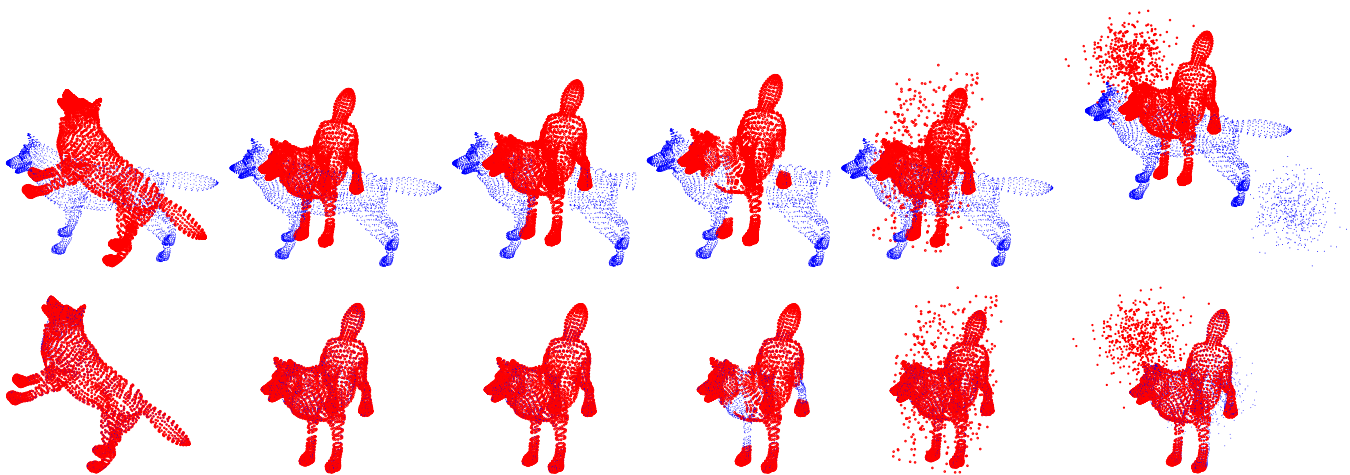


Fig. 4. Point set registration results of our method on the 3D *wolf* shape [14], with deformation, occlusion and outliers presented in every two columns. The goal is to align the model point set (blue dots) onto the observed data point set (red circles). For each group of experiments, the upper figure is the model and data point sets, and the lower figure is the registration result.

TABLE III
COMPARISON OF THE REGISTRATION ERRORS OF PR-GLS WITH RESPECT TO CPD [24], RPM- L_2E [18] AND TF [34] ON THE 3D *wolf* SHAPE IN FIG. 4. FROM LEFT TO RIGHT: DEFORMATION (DEF1, DEF2), OCCLUSION (Occ1, Occ2), AND OUTLIER (OUT1, OUT2).

	Def1	Def2	Occ1	Occ2	Out1	Out2
CPD	0.80	0.82	1.05	0.72	3.58	3.37
RPM- L_2E	1.03	1.18	1.13	0.96	1.22	1.35
TF	1.12	1.36	2.53	3.13	3.87	5.85
PR-GLS	0.50	0.71	0.58	0.45	0.74	0.92

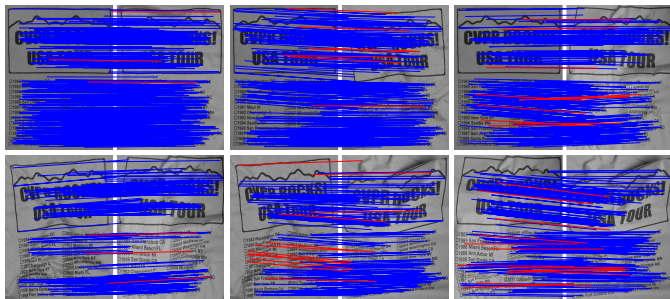


Fig. 5. Results of image feature correspondence on image pairs of deformable objects. From left to right, increasing degree of deformation. The lines indicate matching results (blue = correct matches, red = false matches).

To assign the membership probability π_{mn} , we determine the initial correspondences by searching the nearest neighbors and fix parameter τ to 0.9. Note that since the feature descriptor here is computed based on an image patch rather than the neighboring points, the value of π_{mn} does not need to be recalculated in every iteration.

Fig. 5 contains a T-shirt with different amount of spatial warps, where we aim to establish SIFT matches in each image pair. We consider (x_n, y_m) to be a correspondence/match if its posterior probability p_{mn} is larger than 0.5. Performance is characterized by the number of correct matches and the matching score (i.e., the ratio between the number of correct

matches and the number of total matches). Since the image relations are non-rigid and not known in advance, it is impossible to establish accurate ground truth. We evaluate the matching results by manually checking. Though the judgment of correct match or false match seems arbitrary, we make the benchmark to ensure objectivity.

In feature correspondence, the image typically contains hundreds or thousands of feature points, thus, for efficiency, we use the fast version of our method with $K = 15$ in Eq. (12). The results of our method are presented in Fig. 5. In the left columns, where the deformations are relatively small, our method obtains a lot of correspondences, whereas in the right columns, the deformations are large and the number of correct matches decreases. Still, even in these cases, the matching scores remain quite high, e.g., around 95% or higher. For comparison, we also report three other matching methods as shown in Fig. 6. The SIFT method can be considered as a baseline. To eliminate as many outliers as possible, while preserving the inliers, SIFT compares the distance ratio between the nearest and the second nearest neighbors against a predefined threshold to decide whether they are considered a match or not. From these results, we see that the matching score of SIFT is not good, especially in the case of large deformations. This suggests that using local image features alone is insufficient. CPD is not efficient for such problems, since it ignores the local image features' information. However, RPM- L_2E and PR-GLS consider both global structures and local features and hence, they are able to find more correct matches and achieve high matching scores. The average runtime of our method with the fast implementation on the six image pairs is about 0.5 seconds per pair.

V. CONCLUSION

In this paper, we presented a new approach for non-rigid registration. A key characteristic of our approach is the preservation of both global and local structures during matching. We also provide an efficient implementation of our

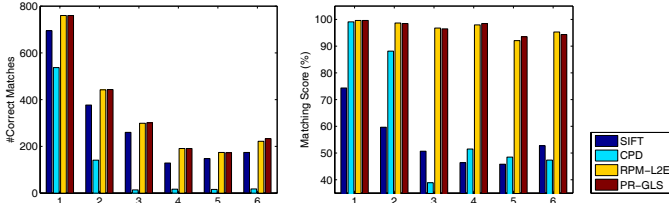


Fig. 6. Comparison of SIFT [17], CPD [24], RPM-L2E [18] and PR-GLS for feature correspondence. Left: number of correct matches; right: matching score. The horizontal axis corresponds to the six image pair in Fig. 5 (from left to right, top to bottom).

method for handling large datasets, which significantly reduces the computational complexity without reducing the quality of the matching significantly. Experiments on public datasets for 2D and 3D non-rigid point registration and on real images for sparse image feature correspondence demonstrate that our approach yields superior results to those of the state-of-the-art methods, when there are significant deformations, noise, outliers, rotations and/or occlusions in the data.

APPENDIX A. VECTOR-VALUED REPRODUCING KERNEL HILBERT SPACE

We review the basic theory of vector-valued reproducing kernel Hilbert space, and for further details and references we refer to [23], [7].

Let \mathcal{X} be a set, for example, $\mathcal{X} \subseteq \mathbb{R}^P$, \mathcal{Y} a real Hilbert space with inner product (norm) $\langle \cdot, \cdot \rangle$, $(\| \cdot \|)$, for example, $\mathcal{Y} \subseteq \mathbb{R}^D$, and \mathcal{H} a Hilbert space with inner product (norm) $\langle \cdot, \cdot \rangle_{\mathcal{H}}$, $(\| \cdot \|_{\mathcal{H}})$, where $P = D = 2$ or 3 for point matching problem. Note that a norm can be induced by an inner product, for example, $\forall \mathbf{f} \in \mathcal{H}$, $\|\mathbf{f}\|_{\mathcal{H}} = \sqrt{\langle \mathbf{f}, \mathbf{f} \rangle_{\mathcal{H}}}$. And a Hilbert space is a real or complex inner product space that is also a complete metric space with respect to the distance function induced by the inner product. Thus a vector-valued RKHS can be defined as follows.

Definition 1. A Hilbert space \mathcal{H} is an RKHS if the evaluation maps $ev_{\mathbf{x}} : \mathcal{H} \rightarrow \mathcal{Y}$ (i.e., $ev_{\mathbf{x}}(\mathbf{f}) = \mathbf{f}(\mathbf{x})$) are bounded, i.e., if $\forall \mathbf{x} \in \mathcal{X}$ there exists a positive constant $C_{\mathbf{x}}$ such that

$$\|ev_{\mathbf{x}}(\mathbf{f})\| = \|\mathbf{f}(\mathbf{x})\| \leq C_{\mathbf{x}} \|\mathbf{f}\|_{\mathcal{H}}, \quad \forall \mathbf{f} \in \mathcal{H}. \quad (15)$$

A reproducing kernel $\Gamma : \mathcal{X} \times \mathcal{X} \rightarrow \mathcal{B}(\mathcal{Y})$ is then defined as: $\Gamma(\mathbf{x}, \mathbf{x}') := ev_{\mathbf{x}} ev_{\mathbf{x}'}^*$, where $\mathcal{B}(\mathcal{Y})$ is the Banach space of bounded linear operators (i.e., $\Gamma(\mathbf{x}, \mathbf{x}')$, $\forall \mathbf{x}, \mathbf{x}' \in \mathcal{X}$) on \mathcal{Y} , for example, $\mathcal{B}(\mathcal{Y}) \subseteq \mathbb{R}^{D \times D}$, and $ev_{\mathbf{x}}^*$ is the adjoint of $ev_{\mathbf{x}}$. We have the following two properties about the RKHS and kernel.

Remark 1. The kernel Γ reproduces the value of a function $\mathbf{f} \in \mathcal{H}$ at a point $\mathbf{x} \in \mathcal{X}$. Indeed, $\forall \mathbf{x} \in \mathcal{X}$ and $\mathbf{y} \in \mathcal{Y}$ we have $ev_{\mathbf{x}}^* \mathbf{y} = \Gamma(\cdot, \mathbf{x}) \mathbf{y}$, so that $\langle \mathbf{f}(\mathbf{x}), \mathbf{y} \rangle = \langle \mathbf{f}, \Gamma(\cdot, \mathbf{x}) \mathbf{y} \rangle_{\mathcal{H}}$.

Remark 2. An RKHS defines a corresponding reproducing kernel. Conversely, a reproducing kernel defines a unique RKHS.

More specifically, for any $N \in \mathbb{N}$, $\{\mathbf{x}_n\}_{n=1}^N \subseteq \mathcal{X}$, and a reproducing kernel Γ , a unique RKHS can be defined by

considering the completion of the space

$$\mathcal{H}_N = \left\{ \sum_{n=1}^N \Gamma(\cdot, \mathbf{x}_n) \mathbf{c}_n : \mathbf{c}_n \in \mathcal{Y} \right\}, \quad (16)$$

with respect to the norm induced by the inner product

$$\langle \mathbf{f}, \mathbf{g} \rangle_{\mathcal{H}} = \sum_{i,j=1}^N \langle \Gamma(\mathbf{x}_j, \mathbf{x}_i) \mathbf{c}_i, \mathbf{d}_j \rangle \quad \forall \mathbf{f}, \mathbf{g} \in \mathcal{H}_N, \quad (17)$$

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

APPENDIX B. PROOF OF THEOREM 1

For any given reproducing kernel Γ , we can define a unique RKHS \mathcal{H}_N as in Eq. (16) in Appendix A. Let \mathcal{H}_N^{\perp} be a subspace of \mathcal{H} ,

$$\mathcal{H}_N^{\perp} = \{ \mathbf{v} \in \mathcal{H} : \mathbf{v}(\mathbf{x}_n) = 0, n \in \mathbb{N}_N \}. \quad (18)$$

From the reproducing property, i.e. Remark 1, $\forall \mathbf{v} \in \mathcal{H}_N^{\perp}$

$$\langle \mathbf{v}, \sum_{n=1}^N \Gamma(\cdot, \mathbf{x}_n) \mathbf{c}_n \rangle_{\mathcal{H}} = \sum_{n=1}^N \langle \mathbf{v}(\mathbf{x}_n), \mathbf{c}_n \rangle = 0. \quad (19)$$

Thus \mathcal{H}_N^{\perp} is the orthogonal complement of \mathcal{H}_N ; then every $\mathbf{v} \in \mathcal{H}$ can be uniquely decomposed into components along and perpendicular to \mathcal{H}_N : $\mathbf{v} = \mathbf{v}_N + \mathbf{v}_N^{\perp}$, where $\mathbf{v}_N \in \mathcal{H}_N$ and $\mathbf{v}_N^{\perp} \in \mathcal{H}_N^{\perp}$. Since by orthogonality $\|\mathbf{v}_N + \mathbf{v}_N^{\perp}\|_{\mathcal{H}}^2 = \|\mathbf{v}_N\|_{\mathcal{H}}^2 + \|\mathbf{v}_N^{\perp}\|_{\mathcal{H}}^2$ and by the reproducing property $\mathbf{v}(\mathbf{x}_n) = \mathbf{v}_N(\mathbf{x}_n)$, the regularized risk functional then satisfies

$$\begin{aligned} \mathcal{E}(\mathbf{v}) &= \frac{1}{2\sigma^2} \sum_{m=1}^M \sum_{n=1}^N p_{mn} \|\mathbf{y}_m - \mathbf{x}_n - \mathbf{v}(\mathbf{x}_n)\|^2 + \frac{\lambda}{2} \|\mathbf{v}_N + \mathbf{v}_N^{\perp}\|_{\mathcal{H}}^2 \\ &\geq \frac{1}{2\sigma^2} \sum_{m=1}^M \sum_{n=1}^N p_{mn} \|\mathbf{y}_m - \mathbf{x}_n - \mathbf{v}_N(\mathbf{x}_n)\|^2 + \frac{\lambda}{2} \|\mathbf{v}_N^{\perp}\|_{\mathcal{H}}^2. \end{aligned} \quad (20)$$

Therefore, the optimal solution of the regularized risk functional (9) comes from the space \mathcal{H}_N , and hence has the form (10). To solve the coefficients, we consider the definition of the smoothness functional $\phi(\mathbf{v})$ and the inner product (17)

$$\begin{aligned} \|\mathbf{v}\|_{\mathcal{H}}^2 &= \langle \mathbf{v}, \mathbf{v} \rangle_{\mathcal{H}} = \left\langle \sum_{i=1}^N \Gamma(\cdot, \mathbf{x}_i) \mathbf{c}_i, \sum_{j=1}^N \Gamma(\cdot, \mathbf{x}_j) \mathbf{c}_j \right\rangle_{\mathcal{H}} \\ &= \sum_{i,j=1}^N \langle \Gamma(\mathbf{x}_j, \mathbf{x}_i) \mathbf{c}_i, \mathbf{c}_j \rangle = \mathbf{C} \mathbf{C}^T, \end{aligned} \quad (21)$$

where $\mathbf{C} = (\mathbf{c}_1, \dots, \mathbf{c}_N) \in \mathbb{R}^{D \times N}$ is the coefficient set. The regularized risk functional then can be conveniently expressed in the following matrix form:

$$\mathcal{E}(\mathbf{C}) = \frac{1}{2\sigma^2} \sum_{m=1}^M \|(\mathbf{Y}_m^N - \mathbf{X} - \mathbf{C} \mathbf{\Gamma}) d(\mathbf{P}_m)^{1/2}\|_F^2 + \frac{\lambda}{2} \mathbf{C} \mathbf{C}^T, \quad (22)$$

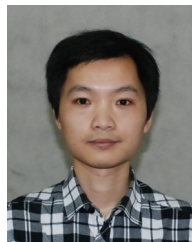
where $\mathbf{\Gamma} \in \mathbb{R}^{N \times N}$ is the so-called Gram matrix with the (i, j) -th element $\Gamma_{ij} = \exp\{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\beta^2}\}$, $\mathbf{Y}_m^N = (\mathbf{y}_m, \dots, \mathbf{y}_m) \in \mathbb{R}^{D \times N}$, \mathbf{P}_m is the m -th row of \mathbf{P} , $\|\cdot\|_F$ denotes the Frobenius norm, and $d(\cdot)$ is the diagonal matrix. Taking the derivative of the last equation with respect to \mathbf{C} and setting it to zero, we obtain the linear system in Eq. (11). Thus the coefficient set $\{\mathbf{c}_n : n \in \mathbb{N}_N\}$ of the optimal solution \mathbf{v} is determined by the linear system (11).

REFERENCES

- [1] N. Aronszajn, "Theory of reproducing kernels," *Transactions of the American Mathematical Society*, vol. 68, no. 3, pp. 337–404, 1950.
- [2] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 24, pp. 509–522, 2002.
- [3] P. J. Besl and N. D. McKay, "A method for registration of 3-d shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.
- [4] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [5] A. Bronstein, M. Bronstein, U. Castellani, B. Falcidieno, A. Fusiello, A. Godil, L. Guibas, I. Kokkinos, Z. Lian, M. Ovsjanikov *et al.*, "SHREC 2010: robust large-scale shape retrieval benchmark," *Proc. 3DOR*, vol. 5, 2010.
- [6] L. G. Brown, "A survey of image registration techniques," *ACM Computing Surveys*, vol. 24, no. 4, pp. 325–376, 1992.
- [7] C. Carmeli, E. De Vito, and A. Toigo, "Vector valued reproducing kernel hilbert spaces of integrable functions and mercer theorem," *Analysis and Applications*, vol. 4, pp. 377–408, 2006.
- [8] H. Chui and A. Rangarajan, "A new point matching algorithm for non-rigid registration," *Computer Vision and Image Understanding*, vol. 89, pp. 114–141, 2003.
- [9] A. Dempster, N. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society. Series B*, vol. 39, no. 1, pp. 1–38, 1977.
- [10] D. J. Field, A. Hayes, and R. F. Hess, "Contour integration by the human visual system: Evidence for a local "association field"," *Vision research*, vol. 33, no. 2, pp. 173–193, 1993.
- [11] A. W. Fitzgibbon, "Robust registration of 2d and 3d point sets," *Image and Vision Computing*, vol. 21, pp. 1145–1153, 2003.
- [12] R. Horaud, F. Forbes, M. Yguel, G. Dewaele, and J. Zhang, "Rigid and articulated point registration with expectation conditional maximization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 587–602, 2011.
- [13] B. Jian and B. C. Vemuri, "Robust point set registration using gaussian mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1633–1645, 2011.
- [14] V. G. Kim, Y. Lipman, and T. Funkhouser, "Blended intrinsic maps," *ACM Transactions on Graphics*, vol. 30, no. 4, p. 79, 2011.
- [15] J.-H. Lee and C.-H. Won, "Topology preserving relaxation labeling for nonrigid point matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 427–432, 2011.
- [16] Z. Lian, A. Godil, B. Bustos, M. Daoudi, J. Hermans, S. Kawamura, Y. Kurita, G. Lavoué, H. Van Nguyen, R. Ohbuchi *et al.*, "SHREC'11 track: Shape retrieval on non-rigid 3d watertight meshes," *3DOR*, vol. 11, pp. 79–88, 2011.
- [17] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [18] J. Ma, J. Zhao, J. Tian, Z. Tu, and A. Yuille, "Robust estimation of nonrigid transformation for point set registration," in *IEEE conference on Computer Vision and Pattern Recognition*, 2013, pp. 2147–2154.
- [19] J. Ma, J. Zhao, J. Tian, A. L. Yuille, and Z. Tu, "Robust point matching via vector field consensus," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1706–1721, 2014.
- [20] J. Ma, J. Zhao, J. Tian, X. Bai, and Z. Tu, "Regularized vector field learning with sparse approximation for mismatch removal," *Pattern Recognition*, vol. 46, no. 12, pp. 3519–3532, 2013.
- [21] A. Makadia, A. Patterson, and K. Daniilidis, "Fully automatic registration of 3d point clouds," in *IEEE conference on Computer Vision and Pattern Recognition*, vol. 1. IEEE, 2006, pp. 1297–1304.
- [22] D. Mateus, R. Horaud, D. Knossow, F. Cuzzolin, and E. Boyer, "Articulated shape matching using laplacian eigenfunctions and unsupervised point registration," in *IEEE conference on Computer Vision and Pattern Recognition*. IEEE, 2008, pp. 1–8.
- [23] C. A. Micchelli and M. Pontil, "On learning vector-valued functions," *Neural Computation*, vol. 17, no. 1, pp. 177–204, 2005.
- [24] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 12, pp. 2262–2275, 2010.
- [25] A. Rangarajan, H. Chui, and F. Bookstein, "The softassign procrustes matching algorithm," in *Information Processing in Medical Imaging*, 1997, pp. 29–42.
- [26] R. Rifkin, G. Yeo, and T. Poggio, "Regularized least-squares classification," in *Advances in Learning Theory: Methods, Model and Applications*, 2003.
- [27] S. Rusinkiewicz and M. Levoy, "Efficient variants of the ICP algorithm," in *3-D Digital Imaging and Modeling*, 2001, pp. 145–152.
- [28] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3d registration," in *IEEE International Conference on Robotics and Automation*, 2009, pp. 3212–3217.
- [29] A. Sharma, R. Horaud, and D. Mateus, "3D shape registration using spectral graph embedding and probabilistic matching," *Image processing and analysing with graphs: theory and practice*, pp. 441–474, 2012.
- [30] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-posed Problems*. Washington: Winston, 1977.
- [31] Y. Tsai and T. Kanade, "A correlation-based approach to robust point set registration," in *European Conference on Computer Vision*, 2004, pp. 558–569.
- [32] Z. Tu and A. L. Yuille, "Shape matching and recognition—using generative models and informative features," in *European Conference on Computer Vision*. Springer, 2004, pp. 195–209.
- [33] D. Wassermann, J. Ross, G. Washko, C.-F. Westin, and R. San Jose Estepar, "Diffeomorphic point set registration using non-stationary mixture models," in *IEEE International Symposium on Biomedical Imaging*. IEEE, 2013, pp. 1042–1045.
- [34] D. Wassermann, J. Ross, G. Washko, W. M. Wells, and R. S. Jose-Estepar, "Deformable registration of feature-endowed point sets based on tensor fields," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2014, pp. 2729–2735.
- [35] A. L. Yuille and N. M. Grzywacz, "A computational theory for the perception of coherent visual motion," *Nature*, vol. 333, pp. 71–74, 1988.
- [36] Y. Zheng and D. Doermann, "Robust point matching for nonrigid shapes by preserving local neighborhood structures," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 643–649, 2006.



Jiayi Ma received the B.S. degree from the Department of Mathematics, and the Ph.D. Degree from the School of Automation, Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2014, respectively. From 2012 to 2013, he was with the Department of Statistics, University of California at Los Angeles. He is now an assistant professor with the Electronic Information School, Wuhan University. His current research interests include in the areas of computer vision, machine learning, and pattern recognition.



Ji Zhao received the B.S. degree in automation from Nanjing University of Posts and Telecommunication in 2005. He received the Ph.D. degree in control science and engineering from HUST in 2012. Since 2012, he is a postdoctoral research associate at the Robotics Institute, Carnegie Mellon University. His research interests include image classification, image segmentation and kernel-based learning.



Alan L. Yuille received his B.A. in mathematics from the University of Cambridge in 1976, and completed his Ph.D. in theoretical physics at Cambridge in 1980 studying under Stephen Hawking. Following this, he held a postdoc position with the Physics Department, University of Texas at Austin, and the Institute for Theoretical Physics, Santa Barbara. He then joined the Artificial Intelligence Laboratory at MIT (1982–1986), and followed this with a faculty position in the Division of Applied Sciences at Harvard (1986–1995), rising to the position of associate professor. From 1995–2002 Alan worked as a senior scientist at the Smith-Kettlewell Eye Research Institute in San Francisco. In 2002 he accepted a position as full professor in the Department of Statistics at the University of California, Los Angeles. He has over two hundred peer-reviewed publications in vision, neural networks, and physics, and has co-authored two books: *Data Fusion for Sensory Information Processing Systems* (with J. J. Clark) and *Two- and Three- Dimensional Patterns of the Face* (with P. W. Hallinan, G. G. Gordon, P. J. Giblin and D. B. Mumford); he also co-edited the book *Active Vision* (with A. Blake). He has won several academic prizes and is a Fellow of IEEE.