

Shape Registration by Simultaneously Optimizing Representation and Transformation

Yifeng Jiang¹, Jun Xie², Deqing Sun¹, and Hungtat Tsui¹

¹ Department of Electronic Engineering, The Chinese University of Hong Kong
{yfjiang,dqsun,httsui}@ee.cuhk.edu.hk

² School of Computer Science, University of Central Florida, USA
jxie@cs.ucf.edu

Abstract. This paper proposes a novel approach that achieves shape registration by optimizing shape representation and transformation simultaneously, which are modeled by a constrained *Gaussian Mixture Model* (GMM) and a regularized *thin plate spline* respectively. The problem is formulated within a Bayesian framework and solved by an expectation-maximum (EM) algorithm. Compared with the popular methods based on *landmarks-sliding*, its advantages include: (1) It can naturally deal with shapes of complex topologies and 3D dimension; (2) It is more robust against data noise; (3) The registration performance is better in terms of the generalization error of the resultant statistical shape model. These are demonstrated on both synthetic and biomedical shapes.

1 Introduction

Shape registration has a long history being a key problem for *shape analysis* in various disciplines [1] and receives increasing interests because of the success of the data-driven *deformable models* [2] in *computer vision* and *medical image analysis*, where shape registration is a bottleneck in the model training stage. To construct such models is the application background of this paper.

There has been a lot of previous research on shape registration, and two basic elements are often involved in various methods:

1. **Shape representation.** The representation of shape data can directly adopt *point sets*, or adopt some parametric models: *parametric curves or surfaces* [3], *Fourier descriptors* or *spherical harmonic functions* [4], and *wavelet transform*. Sometimes non-parametric models are also used: *medial axes*, *M-reps*, *implicit distance functions* [5], and *curvature scale spaces*.
2. **Transformation model.** The mapping between shapes is usually parameterized by transformation models. Global transformation models apply to the entire shapes, *e.g. rigid, similarity (Euclidean), affine, and perspective*. Local transformation models can represent non-rigid deformations, *e.g. optical flow, Thin Plate Splines (TPS)* [6, 7], *Radial Basis Functions* and *Free Form Deformations (FFD)* [5]. In many works, *partial differential equations (PDE)* are also utilized to implicitly represent the transformation model.

To our best knowledge, most existing approaches treat representation first and separately. For example, in [4], the shape is first mapped onto a topologically equivalent parameter space before seeking for the parameters of a contiguous mapping. In [5], the shapes are first represented in the space of distance functions, then the global and local registration are sought by maximizing different shape similarity measures. This leaves the seeking for correspondence to optimizing transformation only. In other words, the shape representations are chosen to fit individual shapes best, without considering the complexity of transformation. From the view of the construction of shape model, which couples both representation and transformation, this may lead to a sub-optimal result.

As an exception, the popular *landmark-sliding* algorithms [6, 3, 8, 9] revise the shape representation together with the transformation, both of which are established on a set of landmarks, thus has the potential to obtain the optimal shape model. However they have two limitations.

1. They require input shapes to be parameterized, so that the trajectories for landmarks can be manipulated. For 3D shapes or shapes with complicated topologies (multi-parts, self-intersection), this is difficult.
2. They are vulnerable to noise, because every piece of landmark trajectory is interpolated by a small number of points. Rough trajectories tend to make the cost function of registration much more nonlinear.

In this paper, we propose a new approach which **S**imultaneously **O**ptimizes **S**hape **R**epresentation and **T**ransformation (**SORT** for short) to overcome the limitations mentioned above. It formulates the shape registration problem in a Bayesian framework and solves the problem by an EM algorithm. SORT is similar to landmark-sliding algorithm in the sense of using a piece-wise linear representation for shapes, but instead of seeking for a set of landmarks, SORT searches for a set of short *segments*, on which the correspondence is established. SORT straightforwardly works on point sets without any need of parameterization, because a set of segments is able to model (represent) a point set directly in the 2D or 3D space.

2 Problem Statement

Given M shapes $\{S_m, m = 1 \dots M\}$ and each shape consists of a set of points $S_m = \{s_{mi}, i = 1 \dots N_m\}$, where s_{mi} is the coordinate of one point. Our goal is to find an appropriate piece-wise representation with L segments for each S_m . These segments are corresponding across the whole shape group. In 2D cases, each segment can be parameterized by a pair of points (u, v) , so S_m is represented by $X_m = \{(u_{mj}, v_{mj}), j = 1, \dots, L\}$. In 3D cases, each segment can be parameterized by three points.

Since all the shapes consist of corresponding segments, we assume there is a *prototype shape*, S_0 , that can generate all shapes in the group by certain transformations. S_0 is represented by $X_0 = \{(u_{0j}, v_{0j}), j = 1, \dots, L\}$. Then, the representation for each S_m is approximately recovered by $X_m = T_m(X_0) + \epsilon t$, where

T_m is the non-rigid transformation between X_0 and X_m , and ϵ_t is noise. Next, line segments are reconstructed from $X_m = \{(u_{mj}, v_{mj}), j = 1, \dots, L\}$ by a certain rule \mathcal{L} , e.g. if (u_{mj}, v_{mj}) represents two ends of a line segment, the rule is simply connecting them. Then all the points of $\{S_m\}$ can be obtained by uniformly sampling (denoted as \mathcal{U}) each segments plus a certain noise ϵ_s . The above process can be considered as a *generative model*:

$$S_m = \{s_{mi}\} = \mathcal{U} \cdot \mathcal{L} \cdot (T_m(X_0) + \epsilon_t) + \epsilon_s, \text{ where } \epsilon_t \sim N(0, I \cdot \sigma_t), \epsilon_s \sim N(0, I \cdot \sigma_s) \quad (1)$$

I is the 2×2 identity matrix, and σ_s is a scale indicating the data's *noise level*.

Now the shape registration problem can be formally stated as finding X_0 , $\{T_m\}$ and $\{X_m\}$, which makes the given $\{S_m\}$ most probable under the generative model. In a Bayesian framework, we infer them by maximizing the *a posterior* probability (MAP). Under the i.i.d. assumption, we have:

$$\begin{aligned} (\hat{X}_0, \{\hat{T}_m, \hat{X}_m\}) &= \arg \max_{X_0, \{T_m, X_m\}} p(X_0, \{T_m, X_m\} | S_m) \\ &= \arg \max \prod_{m=1}^M p(S_m | X_m, \theta_m) p(X_m | X_0, T_m) p(T_m), \end{aligned} \quad (2)$$

where θ_m represents the additional parameters may be involved. Note that there is no direct statistical relationship between S_m and (X_0, T_m) , neither between X_0 and $\{T_m\}$. θ_m and X_0 are simply assumed to be uniformly distributed. We discuss the *likelihood* and *prior* items in the following sections, which are corresponding to the representation and transformation models in shape registration.

2.1 Likelihood

It is natural to use mixture model as the likelihood model since we are going to represent the data by a set of segments. Each mixture component is devoted to one segment, with parameters $(u_{mj}, v_{mj}, \sigma_s)$. Recall there are N_m points in S_m , we have:

$$p(S_m | X_m, \theta_m) = \prod_{i=1}^{N_m} p(s_{mi} | X_m, \theta_m) = \prod_{i=1}^{N_m} \sum_{j=1}^L \pi_{mj} p(s_{mi} | u_{mj}, v_{mj}, \sigma_s) \quad (3)$$

where π_{mj} is component coefficient, and $\pi_{mj} \geq 0, \sum_{j=1}^L \pi_{mj} = 1$. According to the generative model (1), $p(s_{mi} | u_{mj}, v_{mj}, \sigma_s)$ may be further decomposed into a Gaussian and a uniform distribution between the two ends of the line segment. However, such model is not easy to be manipulated because of the uniform distribution involved. In this paper, we approximate this using a 2D Gaussian by imposing two constraints on the eigenvalues of its covariance matrix. Denoting the two eigenvalues as λ_1 and λ_2 and $\lambda_1 \geq \lambda_2$, the constraints are: (1) $\lambda_2 = \sigma_s^2$, which means the width of segment is equal to σ_s ; (2) $2\rho\sqrt{\lambda_1} = l$, where l is the length of line segment, and $\rho = \sqrt{2}$ in this paper. Because we prefer to use $(u_{mj}, v_{mj}, \sigma_s)$ as the parameters of likelihood model, and (u_{mj}, v_{mj}) should be coordinates of points, we need to establish the relationship between $(u_{mj}, v_{mj}, \sigma_s)$ and Gaussian's parameters (μ, Σ) . It is straightforward to set $u_{mj} = \mu$ as the center of Gaussian. As to v_{mj} , please refer to Fig. 1, we have:

$$\Sigma_{mj} = f(v_{mj}, \sigma_s) = R^T \Lambda R, \quad \text{where } R = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \quad \text{and } \Lambda = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma_s^2 \end{bmatrix}$$

$$\theta = \arctan \frac{y}{x}, \quad \sigma^2 = \frac{1}{\alpha}(x^2 + y^2), \quad [x, y]^T = v_{mj} - u_{mj}$$

which means v_{mj} is one point on the major axis of 2D Gaussian’s equalprob contours (ellipses), departing a certain distance from u_{mj} . The distance is set to be $\alpha\sigma^2$ where σ^2 the maximal eigen value of Σ_{mj} , and $\alpha \in (0, 1)$.



Fig. 1. Using 2D Gaussian to model a segment of shape. Left: Green dots represent the shape points, modeled by the blue line segment. The ellipse is one of the Gaussian’s equalprob contour, and (u, v) is located on its major axis; Right: using a set of Gaussians to model a shape of hand.

As v and v' shown in Fig. 1, there are two points along the major axis of Σ_{mj} that satisfy (4), on either sides of u_{mj} . The selection of v_{mj} is discussed in section 3.1.

We also expect a constraint on v_{mj} , which is that the ratio of σ to σ_s should not exceed a certain value β , $\beta > 1$. So finally in (3), $p(s_{mi}|X_m, \theta_m)$ is a constrained GMM, where $p(s_{mi}|u_{mj}, v_{mj}, \sigma_s)$ is a constrained Gaussian:

$$p(s_{mi}|u_{mj}, v_{mj}, \sigma_s) = \frac{1}{2\pi|\Sigma_{mj}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(s_{mi} - u_{mj})^T \Sigma_{mj}^{-1}(s_{mi} - u_{mj})\right)$$

$$\text{s.t. } \min(\text{eig}(\Sigma_{mj})) = \sigma_s, \Sigma_{mj} = f(v_{mj}, \sigma_s), \text{ and } \sigma_s^2 \leq \frac{1}{\alpha} \|v_{mj} - u_{mj}\|^2 \leq \beta \cdot \sigma_s^2 \quad (4)$$

2.2 Priors

According to the generative model (1), $p(X_m|T_m, X_0)$ is a Gaussian distribution:

$$p(X_m|T_m, X_0) = \frac{1}{2\pi\sigma_t} \exp\left(-\frac{1}{2\sigma_t} \|X_m - T_m(X_0)\|^2\right). \quad (5)$$

As to the transformations $\{T_m\}$, we assume them to be smooth, and we do not want any penalty on the *rotation*, *translation* and global *shear*, because we consider this is not the *intrinsic variation* of a shape. In this case the bending energy of Thin Plate Spline (TPS), E_{tps} , can be utilized to construct the prior, as successfully applied in many previous works [6,9], by which the density function of T_m can be defined as $p(T_m) = \frac{1}{Z_t} \exp(-\lambda_t E_{tps}(T_m))$. Then the total prior density function can be written as

$$p(X_m, T_m, X_0) = \frac{1}{Z} \exp\left(-\gamma(\|X_m - T_m(X_0)\|^2 + \lambda E_{tps}(T_m))\right), \quad (6)$$

where coefficient λ and γ are hyperparameters, and Z is a partition constant adopted to guarantee that $p(X_m, X_0, T_m)$ is a density function.

3 Optimization

The maximizing *a posteriori* (MAP) problem is now formulated as:

$$(\{\hat{\theta}_m\}, \{\hat{X}_m\}, \{\hat{T}_m\}, \hat{X}_0) = \arg \max \prod_{m=1}^M \prod_{i=1}^{N_m} \sum_{j=1}^L \pi_{mj} p_{mj}(s_{mi}|X_m, \sigma_s) p(X_m, T_m, X_0) \quad (7)$$

All the parameters, in both likelihood and prior items will be optimized together, or in other words, parameters for both representation and transformation models will be optimized simultaneously. Since the Gaussian mixture is involved, an EM algorithm is adopted for this task, where close-form solutions exist for π_{mj} , $\{u_{mj}\}$, X_0 and T_m , among which T_m is a regularized TPS [10]. There is no analytic solution for $\{v_{mi}\}$, and they are optimized by gradient-based method in each maximization iteration.

3.1 Implementation

The proposed representation and transformation models both have a parameter that can naturally control how fine the models are, so we embed EM iteration in a *coarse-to-fine* scheme to achieve a more stable optimization. For representation model, it is σ_s , which represents the noise level of shape data that the constrained GMM can cope with; For deformation model it is λ , which controls how nonrigid T_m is. The implementation in this paper starts with great σ_s^{max} and λ^{max} , then gradually decrease them by certain ratios r_σ and r_λ respectively, until $\sigma_s^{(k)}$ is less than σ_s^{min} . For each $\sigma_s^{(k)}$, the models are updated by the embedded EM algorithm, taking the result of previous step as the initial condition.

The very initial $\{T_m^{(0)}\}$ are set to be identity transforms plus a translation between each shape centroid and their mean. This assumes that at the coarsest stage the representation of all shapes are the same except for a translation. The initial $\{u_{mj}^{(0)}\}$ are set to be evenly placed on a small circle around shape centroid. Besides, we set the initial covariance matrix of each Gaussian mixture to be isotropic, thus the equalprob of Gaussian becomes a circle, so as the possible positions of v_{mj} . Then we pick the rightmost point as $v_{mj}^{(0)}$ for all the mixture components. After first update, $v_{mj}^{(k)}$ will always be the ones resulted from EM algorithm and no selection is needed. The initial $X_0^{(0)}$ is the mean of $\{X_m\}$.

In all the experiments presented in this paper, $\lambda_m^{max} = N_m \times M \times \sigma_s^{max}$, σ_s^{max} is set to be the square of the size of the shape, and the final σ_s^{min} is set to be the mean square of distance between neighboring points. Ratio r_σ and r_λ are both set to be 0.95. γ is difficult to decide and in this paper it is set to be $1/\sigma_s^{(k)}$. For other parameters, α is not sensitive and $\alpha = 0.5$; β decides how long the segments will be, and is the only parameter we adjust according to the data.

4 Experimental Results

4.1 Register 2D Shapes

We first demonstrate the registration results of SORT on 2D shapes with simple topologies. Typical landmark-sliding algorithms can also work on these shapes so a comparison can be conducted. SORT is compared with two of them. One is the MDL based algorithm implemented by Thodberg [8], and the other is arc-length parameterization. In all experiments, Thodberg’s algorithm will run with 8 active nodes optimized over 40 passes, using 16, 32, or 64 landmarks (SORT will work with 8, 16 or 32 segments accordingly). In addition, a popular point cloud matching algorithm – Super Clustering-Matching (SCM) algorithm proposed by Chui [7], is also included for comparison. All algorithms are implemented using MATLAB 7.1, and run on an 1.66 GHz Intel Centrino CPU.

Because our purpose for shape registration is to build shape models, we employ the *model generalization error* [3] as the criterion for evaluation. In particular, the *error* adopted here is the fitting error between the original shape *contour* and the contour generated by the resultant model, which is more meaningful in practice than the distance error between landmarks points. To evaluate the results of SORT by this criterion, we take $X_m = \{(u_{mj}, v_{mj}), j = 1, \dots, L\}$ as the “landmarks”, so the number of landmarks is twice the number of segments.

Shape data under experiments include 9 synthetic bump boxes, 9 shapes of femur, and 9 silhouette profiles. All of them are obtained from Thodberg’s package [8] and widely tested in literatures.

Fig. 2 shows the plain registration result for 9 bumpboxes and the variation of resulting shape model. The result is yielded by SORT using 8 segments. The equalprob ellipse of each segment is drawn on the shape (denoted by dense green dots), and the corresponding ones on different shapes have the same colors. The black and red dots on each segment denote $\{u_{mj}\}$ and $\{v_{mj}\}$ respectively. The variation of resulting model shows that SORT has correctly captured the variation of bumpboxes. The registration process takes 37 seconds.

The mean and standard deviation of shape generalization error for bumpboxes, are demonstrated in Fig. 3. They are compared with Thodberg’s algorithm, the

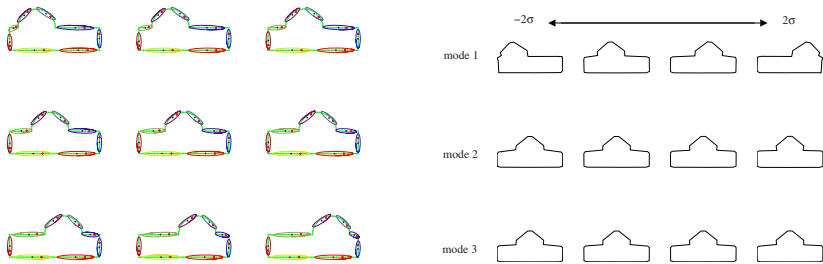


Fig. 2. Registration of 9 Bump boxes, using 8 segments. Left: Plain registration result; Right: the first three modes of variation.

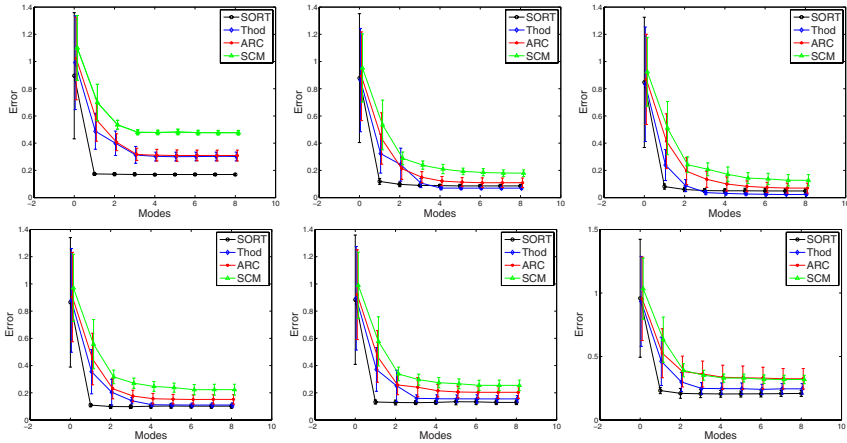


Fig. 3. Shape generalization errors on bump boxes. First row: From left to right, SORT uses 8, 16, and 32 segments, and other 3 algorithms use 16, 32, and 64 landmarks accordingly. Second row: Generalization error with noise of scale 0.1, 0.2, and 0.4 (from left to right). SORT uses 16 segments while other algorithms use 32 landmarks.

arc-length parameterization, and Chui’s SCM, denoted as “Thod”, “ARC” and “SCM” respectively in the figure. It is observed that SORT has considerable advantage when small number of segments is used for registration. Although SCM can find optimal transformations, its performance is not so good in terms of the shape generalization error. This is because the design of SCM has no intention to obtain an optimized representation of the original shapes. This is also true for most point cloud matching algorithms.

Fig. 3 also shows the shape generalization error of registration under different scale of Gaussian noise. The scales are defined as the ratio between noise magnitude and mean distance among neighboring points of the shape. It is observed that the performance of SORT is obviously less affected by data noise.

Similar results are also observed on Femur, and silhouette profiles. To demonstrate SORT’s capability of registering shapes of complicated topologies, which is difficult for landmark-sliding methods, a registration is conducted for shapes of 9 heart, each consisting of two chambers (shapes of multi-parts), and the results are given in Fig. 4. The registration process takes 173 seconds.

4.2 Register 3D Shapes

In the 3D space, the equalprob surfaces of Gaussian are ellipsoids, and the kernel for regularized TPS becomes $\phi_i(x) = \|x - X_{0i}\|$. Each segment is parameterized by 3 points (u, v, w) , where u is still the center of the Gaussian, while (v, w) are located on the first and second major axes of its equalprob ellipsoid. The

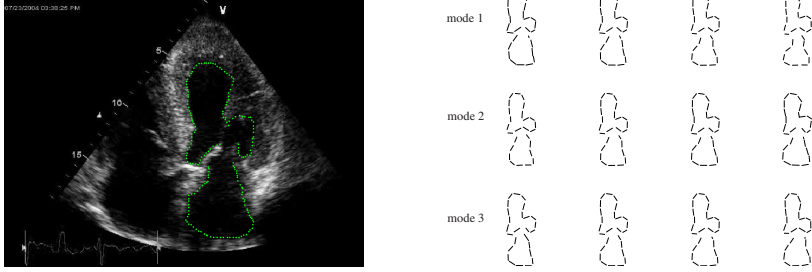


Fig. 4. Registration of 9 heart chamber shapes. SORT uses 26 segments. Left: a shape consisting of 2 chambers extracted from an echocardiogram image; Right: first 3 modes of shape variation.

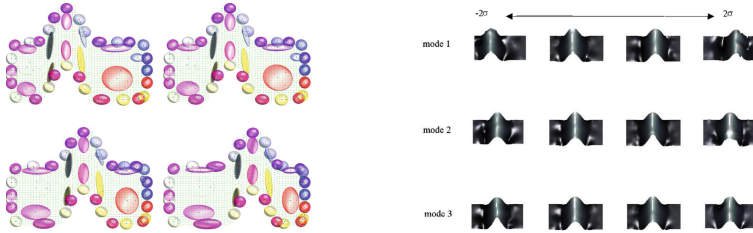


Fig. 5. Registration of 9 3D bumps. Left: Corresponding equalprob surfaces of segments on 4 bump shapes; Right: the shape variation (Surface shown here is interpolated from ends of the first two major axes of all the segments).

constraints for Gaussian are similar to 2D cases as in section 2.1, $\min(\text{eig}(\Sigma_{m_j})) = \sigma_s, \sigma_s^2 \leq \frac{1}{\alpha} \|v_{m_j} - u_{m_j}\|^2 \leq \beta \cdot \sigma_s^2, \sigma_s^2 \leq \frac{1}{\alpha} \|w_{m_j} - u_{m_j}\|^2 \leq \beta \cdot \sigma_s^2$, resulting a flattened equalprob ellipsoid, analogous to an elongated ellipse in 2D cases.

But attention is needed on the *ends* of those open shapes, *i.e.* the boundaries of open surfaces. In our experiments, we handle this by registering the boundaries separately as a 3D curve, whose results are then merged to the surface registration results. Fig. 5 demonstrates the registration results on a group of synthesized 3D bumps. 12 segments are used for each surface, and 30 segments are used for each surface boundary. As seen, SORT performs very well to capture the only shape variation for this dataset.

5 Summary and Future Work

In this paper, shape registration is formulated as a Bayesian inference problem with a constrained GMM coupled by a regularized TPS based prior. This problem is solved by an algorithm called SORT. SORT basically is an EM algorithm embedded in a coarse-to-fine scheme. Extensive experimental results demonstrate that SORT has a number of advantages compared with the popular algorithms

based on landmark-sliding. For future work, we plan to study some ways to improve the optimization, such as split and merge EM.

References

1. Dryden, I.L., Mardia, K.V.: *Statistical Shape Analysis*. John Wiley and Sons, West Sussex (1998)
2. Cootes, T.F., Taylor, C., Cooper, D., Graham, J.: Active shape models – their training and application. *Comput Vis Image Underst* 61, 38–59 (1995)
3. Davies, R.: *Learning Shape: Optimal Models for Analysing Shape Variability*. PhD thesis, University of Manchester (2002)
4. Meier, D., Fisher, E.: Parameter space warping: Shape-based correspondence between morphologically different objects. *IEEE Trans. Med. Imaging* 21(1), 31–47 (2002)
5. Huang, X., Paragios, N., Metaxas, D.N.: Shape registration in implicit spaces using information theory and free form deformations. *IEEE Trans. Patt. Anal. Mach. Intell.* 28(8), 1303–1318 (2006)
6. Bookstein, F.: Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Med. Image Anal.* 1, 225–243 (1997)
7. Chui, H., Zhang, J., Rangarajan, A.: Unsupervised learning of an atlas from unlabeled point-sets. *IEEE Trans. Patt. Anal. Mach. Intell* 26, 160–173 (2004)
8. Thodberg, H.H.: Minimum description length shape and appearance models. In: *Proc. IPMI, BMVA*, pp. 51–62 (2003)
9. Richardson, T., Wang, S.: Shape correspondence using landmark sliding, insertion and deletion. In: *Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749*, pp. 435–442. Springer, Heidelberg (2005)
10. Wahba, G.: *Spline models for observational data*. Society for Industrial and Applied Mathematics, Philadelphia (1990)