Statistical Analysis of Manual Segmentations of Structures in Medical Images

Sebastian Kurtek[§], Jingyong Su[*], Cindy Grimm[◇],
Michelle Vaughan[†], Ross Sowell[‡], Anuj Srivastava[*]

[§]*Department of Statistics, The Ohio State University*

[*]*Department of Statistics, Florida State University*

[◇]*School of Mechanical, Industrial & Manufacturing Engineering, Oregon State University*

[†]*Department of Computer Science and Engineering, Washington University in St. Louis*

[‡]*Department of Computer Science, Cornell College*

Abstract

The problem of extracting anatomical structures from medical images is both very important and difficult. In this paper we are motivated by a new paradigm in medical image segmentation, termed *Citizen Science*, which involves a volunteer effort from multiple, possibly non-expert, human participants. These contributors observe 2D images and generate their estimates of anatomical boundaries in the form of planar closed curves. The challenge, of course, is to combine these different estimates in a coherent fashion and to develop an overall estimate of the underlying structure. Treating these curves as random samples, we use statistical shape theory to generate joint inferences and analyze this data generated by the citizen scientists. The specific goals in this analysis are: (1) to find a robust estimate of the representative curve that provides an overall segmentation, (2) to quantify the level of agreement between segmentations, both globally (full contours) and locally (parts of contours), and (3) to automatically detect outliers and help reduce their influence in the estimation. We demonstrate these ideas using a number of artificial examples and real applications in medical imaging, and summarize their potential use in future scenarios.

Key words: medical imaging, segmentation uncertainty, non-expert image segmentation, statistical analysis of planar curves

1 Introduction

Segmentation of medical images to extract anatomical structures is arguably the most important problem in medical image analysis. A variety of techniques rang-

ing from variational methods to statistical methods to fuzzy logic have been applied to this problem but a general solution still remains elusive. The difficulty lies in capturing the vast observed variability, in the form of pixel values and anatomical shapes, into appropriate objective functions that enable high-quality segmentations. This is further compounded by low resolution, high noise and blurriness associated with common medical imaging modalities. The segmentation performance achieved by trained humans – physicians, clinicians, technicians, etc. – has been very difficult to match by automated systems. Since manual segmentations by skilled practitioners are generally expensive and the amount of imaging data being generated far out-paces the experts' availability, it is not practical to rely on these trained personnel for all our segmentation needs.

There are several directions for addressing this gap between the need and the available expertise. Firstly, motivated by the success of manual segmentations by *trained* experts, an important focus of current research in segmentation has been in training machine learning algorithms to improve their performance. A similar idea has been to develop structured prior models on shapes of interest and perform *Bayesian segmentations* – segmentations guided by prior knowledge of statistical shape variability associated with anatomical structures to be segmented. A second, completely different paradigm is to involve humans, albeit non-experts, who can volunteer their time and effort in segmenting medical images. This approach is part of a larger effort involving public participation in scientific data analysis. It is this second direction that motivates the methods presented in this paper. The last decade has seen a tremendous increase in what has been termed *Citizen Science* - non-experts helping scientists collect and/or analyze data. These projects range from Galaxy Zoo, where non-experts help classify galaxies [1], to Fold-It, where people help fold proteins [2], to measuring trees and counting birds in nests (Cornell Lab of Ornithology) [3]. These projects take advantage of human perceptual abilities and intuitions about shape (Galaxy Zoo/Fold-It) and the ability of volunteers to take measurements over a large spatial extent. We are currently trying to build the foundation for Citizen Science projects that further exploit the growth of digital imaging, both to support volunteers annotating 2D and 3D structures in imagery, and in building tools to allow volunteers to capture calibrated imagery to support more quantitative measurements. We believe these tools may support a diverse set of Citizen Science projects. This paper is about analyzing data arising from multiple manual segmentations of 3D medical image data, but could be applied to 2D photograph segmentations as well (e.g. finding birds, trees or cars in a scene). Shown in Fig. 1 is an illustration of a manual segmentation of the liver. The segmentation is performed using multiple oblique and parallel image slices and the resulting two-dimensional curves can be used for full three-dimensional surface reconstruction.

One challenge any data collection and analysis project faces is data verification and validation. This is particularly true in Citizen Science projects, where the participants can come from a wide variety of backgrounds and skill sets. To date, these projects have largely relied on carefully constructed training tutorials along with

2
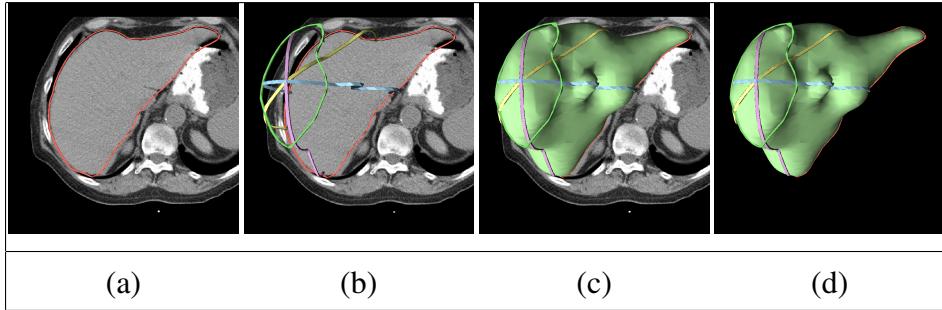
Fig. 1. Illustrations of manual segmentations of the liver in CT images. (a) 2D contour segmented from a planar image. (b) Contours segmented from different imaging planes. (c) Reconstructed surface imposed on the image. (d) Segmented surface.

simple statistics to filter out incorrect data (be it malicious or unintentional). Simple statistics work well in the case of category labels (e.g. Galaxy Zoo or bird counting) because any given dataset is viewed by multiple people and each person marks multiple datasets. This provides fairly reliable statistics for both the datasets and reliability data on the individuals marking the datasets.


Fig. 2. 2D CT image with a zoom-in on an area where the boundary of the brainstem is partly visible. Five expert segmentations with the one shown in the image in red. The contrast of CT images is very low making the segmentation problem difficult.

We would like to extend the same statistical approach to support Citizen Science projects where the volunteers contribute contour data rather than category labels. An example of such data (brainstem) is provided in Fig. 2. We show a 2D CT image with an example segmentation, a zoom-in on part of the image where the boundary of the brainstem is partly visible, and five expert segmentations that form the raw data for techniques developed in this paper. In particular, we would like to compute sample statistics for multiple contours that let us answer the following specific questions about the data:

(1) What is the mean or median contour? This helps filter out small noise and discrepancies in any individual person's contour.
(2) What is the magnitude of the standard deviation of a set of contours? This provides a measure of confidence for the segmentations.
(3) Is a contour an outlier?
(4) What is the average distance – under a shape metric – of an individual's contours from the average contour?
(5) Do the segmentations follow a Gaussian distribution or is the data better de-

scribed by a mixture of Gaussian distributions? (This happens quite often when there is more than one interpretation of the instructions. For instance, one group might contour an inner wall, while the other group contours an outer wall.)

In addition to answering the above questions, we are interested in providing the scientists (in this case physicians) with visualization tools that will help them analyze both inter and intra variation in their datasets, and to provide confidence analysis on their reconstructions. To this end, we provide visualizations of the major modes of variation, and their magnitudes, which provide some estimate of the reliability of different segments of the contour. It is important to note that these measures of confidence or reliability do not necessarily correlate with segmentation accuracy, but rather just consistency.

Historically, the problem of shape analysis of curves has been studied with a variety of different mathematical representations of curves. These representations include sample points or landmarks [4–6], level sets [7], deformable templates [8], medial axes [9,10], algebraic invariants, and others. However, the most natural representation – a parameterized curve – is relatively infrequent. The difficulty in studying shapes of parameterized curves lies in the parameterization variability: a curve can be re-parameterized in infinitely many ways but still have the same shape. Davies et al. [11,12] use landmark models and compute optimal re-parameterizations using a minimum description length (MDL) type energy. Although such models have proven very useful in medical image analysis and have good specificity, generalizability and compactness (as compared to manually selected landmarks and arclength parameterization), the method used to compute them has three main limitations. First, the optimization problem is defined in terms of ensembles and thus the distance between any two shapes depends on the other shapes in the ensemble. This contrasts the standard mathematical definition of a distance between objects. Second, the MDL energy does not preserve distances between shapes when they are re-parameterized in the same way. Intuitively, the shape metrics and statistical analysis should not change with a change in the parameterization, but rather only with a change in shape. Finally, the MDL driven optimization problem requires a pre-selection of a template, which can be rather arbitrary and choosing different templates may lead to different solutions.

An emerging body of work has proposed using an algebraic approach to handle the problem of parameterization. In this approach, one unifies all possible re-parameterizations of a curve by forming equivalence relations, and each shape is denoted uniquely by an equivalence class. Shape metrics are defined and computed between equivalence classes, rather than individual curves. A very important advantage of this framework is that the process of comparing any two shapes, i.e. two equivalence classes, involves finding optimal registrations between the corresponding curves. (Note that the registration of points along curves corresponds to re-parameterizations of curves.) Thus, this framework, termed *elastic shape anal-*

ysis, leads to a simultaneous registration and comparison of shapes of curves, under a single metric, and to a principled approach to shape analysis. Notably, the methods involving landmarks, level sets, medial axes, and others, do not provide this ability to simultaneously register and analyze shapes. Several papers, including [13–15], have utilized an elastic framework, but we will use the approach presented in [16,17] for this paper. An important advantage of this approach is that amongst all methods for elastic shape analysis of contours, this is the only one so far that extends to contours in higher dimensions.

Thus, the main contributions of this work are:

(1) The application of elastic analysis to modeling contour segmentation uncertainty in medical images.
(2) A method for visualization of segmentation reliability based on principal component analysis.
(3) Extension of shape analysis methods presented in [16,17] that enables one to model rotation, scale and position in addition to shape.
(4) Definition and computation of the elastic median based on a general algorithm presented in [18].
(5) A formal statistical procedure to identify outliers in datasets based on elastic distances from the median.

The rest of this paper is organized as follows. In Section 2, we summarize a mathematical/statistical framework presented in past papers, such as [17], and adapt these tools for use in the current problem. Specifically, we introduce an algorithm for estimating the sample median in the space of contours and use that median estimate for studying dominant modes of data variability. Experimental results involving brainstem and prostate images are presented in Section 3, and the paper concludes with a short summary in Section 4.

2 Mathematical Framework

The basic problem we face requires a technique for statistical analysis of planar, closed contours. There are several theories available for this task. In this section we summarize a recent approach for elastic analysis of contours that is particularly attractive for the current problem.

2.1 Elastic Analysis of Curves

Let a parameterized planar, closed curve be denoted as $\beta(t) \in \mathbb{R}^2$, where t is the parameter. Since the curve is closed, it is more natural to parameterize it using the

domain $\mathbb{S}^1$, instead of an interval, since there are no natural end points on a circle to match the two end points on an interval. There are several possibilities for mathematically representing β. One can simply use the x and y coordinate functions of β, as shown in Fig. 3(b). Another possibility is to parameterize β using arc-length and compute the angle $\dot{\beta}$ makes with the x axis, as shown in (c). Finally, one can take the derivative of this previous angle function to obtain the curvature function, shown in (d). Although arc-length parameterization (with fixed seed placement)
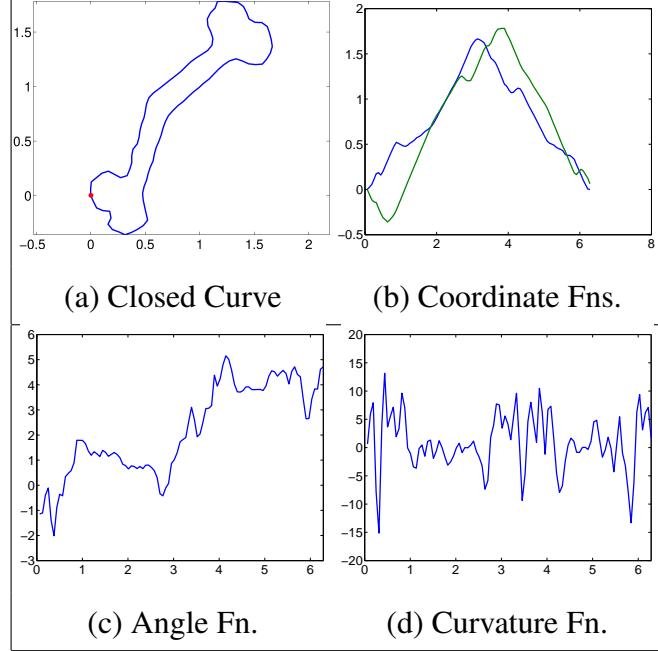

Fig. 3. (a) A simple closed curve β with the starting point in red, (b) the two coordinate functions x and y plotted against the arc-length, (c) the angle function θ, and (d) the curvature function κ.

removes the variability associated with the parameterization of curves, it suffers from the problems of suboptimal registration due to a linear registration of points across curves. It is more natural to include arbitrary parameterizations of curves in the analysis, and to seek optimal re-parameterizations during pairwise matching of curves. This allows for the possibility of nonlinear registrations of points, an aspect that is central to elastic shape analysis.

Define the group of re-parameterization functions as:

$$\Gamma = \left\{ \gamma : \mathbb{S}^1 \to \mathbb{S}^1 | \gamma \text{ is an orientation-preserving diffeomorphism} \right\}.$$

The re-parameterization of a curve β, termed the action of Γ on the space of curves, is given by composition, $(\beta, \gamma) = \beta \circ \gamma$. One also needs a metric for comparing shapes of curves and the Euclidean metric is the most common choice in the literature. A major problem in methods that use Euclidean distances to compare shapes is that $\|\beta_1 - \beta_2\| \neq \|\beta_1 \circ \gamma - \beta_2 \circ \gamma\|$, for a general $\gamma \in \Gamma$, where $\| \cdot \|$ is the $\mathbb{L}^2$ metric for functions on $\mathbb{S}^1$. This means that a comparison of two curves depends

6

on their parameterizations! A solution suggested by Srivastava et al. [16,19,17] is to use a new mathematical representation of curves, called the square-root velocity function (SRVF), given by $q(t) \equiv \frac{\dot{\beta}(t)}{\sqrt{\|\dot{\beta}(t)\|}}$. If a curve β is re-parameterized to $\beta \circ \gamma$, then its SRVF changes from q to $(q \circ \gamma)\sqrt{\dot{\gamma}}$; we will use the notation (q, γ) to denote the new SRVF. One of the reasons for using this representation is that $\|q_1 - q_2\| = \|(q_1, \gamma) - (q_2, \gamma)\|$, for all $\gamma \in \Gamma$.

In the case where the curve is closed, the corresponding SRVF satisfies the condition $\int_{\mathbb{S}^1} q(t)\|q(t)\|dt = 0$. Thus, the space of all planar, closed curves, represented by their SRVFs, is given by $\mathcal{C} = \{q : \mathbb{S}^1 \to \mathbb{R}^2| \int_{\mathbb{S}^1} q(t)\|q(t)\|dt = 0\}$. $\mathcal{C}$ is a nonlinear manifold because of the closure condition. In a general shape analysis framework one often removes the variability due to scale, rotation and position of the curves, but in this work these quantities are informative and are thus included in the analysis. The position variable is, unfortunately, lost in the SRVF representation, since it is based on the derivative $\dot{\beta}$. We reinstate it in the analysis as a separate variable as described in Sec. 2.2. $\mathcal{C}$ is a Riemannian manifold with the standard $\mathbb{L}^2$ metric, $\langle v_1, v_2 \rangle = \int_{\mathbb{S}^1} \langle v_1(t), v_2(t) \rangle dt$, where the inner product in the integrand is the standard Euclidean inner product in $\mathbb{R}^2$. The task of computing geodesic paths between any two elements $q_1, q_2 \in \mathcal{C}$ is accomplished numerically, using an algorithm called the *path straightening* algorithm, introduced first in [20] but adapted to the SRVF representation in [17]. This algorithm initializes a path in $\mathcal{C}$ connecting q_1 and q_2, and iteratively "straightens" it until it becomes a geodesic. Let $\alpha : [0, 1] \in \mathcal{C}$ denote the resulting geodesic path. Then, the length of this geodesic path provides a geodesic distance between q_1 and q_2 in $\mathcal{C}$: $d_c(q_1, q_2) = L[\alpha] = \int_0^1 (\langle \dot{\alpha}(\tau), \dot{\alpha}(\tau) \rangle)^{1/2} d\tau$.

Curves that are within a re-parameterization of each other result in different elements of $\mathcal{C}$. The unification of such curves is performed using an equivalence relation under the re-parameterization group action. That is, for all elements in $\mathcal{C}$ that denote re-parameterizations of the same curve, we put them in the same equivalence class given by $[q] = \text{closure}\{(q \circ \gamma)\sqrt{\dot{\gamma}}|\gamma \in \Gamma\}$. Each such equivalence class $[q]$ is associated with a shape uniquely and vice versa. The set of all these equivalence classes is a quotient space and is denoted by $\mathcal{S}$. The distance d_c can be used to define a distance on $\mathcal{S}$ according to $d_s([q_1], [q_2]) = \inf_{\gamma \in \Gamma} d_c(q_1, (q_2, \gamma)) = \inf_{\gamma \in \Gamma} d_c((q_1, \gamma), q_2)$. This minimization is performed in practice by sampling each curve with some large number of points and then applying dynamic programming with an additional seed search.

2.2 Statistical Summaries

An important part in our analysis of contours is generating their statistical summaries – in the form of means and covariances – and using these sample statistics to analyze contours. With the possibilities of bad segmentations in mind, we also

want the statistical summaries to be robust to outliers. Specifically, we are interested in a robust estimation of a representative contour using the concept of a *median*. Next, we provide the definition and algorithms to compute statistical summaries of closed, planar curves in the form of a mean, median and covariance. Furthermore, we provide a simple yet effective procedure for removal of outliers.

Mean Estimation: Once we have defined a distance d_s on $\mathcal{S}$, the mean can be defined as follows. Let $\beta_1, \beta_2, \ldots, \beta_n$ denote a set of given contours and $q_1, q_2, \ldots, q_n$ be the corresponding SRVFs. Then, the Karcher (or Frechet) mean curve (actually an equivalence class of curves) is $[\bar{q}] = \operatorname{argmin}_{[q] \in \mathcal{S}} \sum_{i=1}^n d_s([q], [q_i])^2$. A gradient-based approach for finding this mean is given in several places [21,4], and is repeated here for convenience as Algorithm 1. The algorithms for computing the exponential map and its inverse – $\exp$ and $\exp^{-1}$ – are similar to the technique for finding geodesics and are presented in [17].

Algorithm 1 (Karcher Mean): Let $\bar{q}_0$ be an initial estimate of the Karcher mean. Set $j = 0$ and ϵ_1, ϵ_2 to be small positive values.
(1) For each $i = 1, \ldots, n$, compute $v_i = \exp_{\bar{q}_j}^{-1}(q_i)$.
(2) Compute the average direction $\bar{v} = (1/n) \sum_{i=1}^n v_i$.
(3) If $\|\bar{v}\| < \epsilon_1$, then stop. Else, update using $\bar{q}_{j+1} = \exp_{\bar{q}_j}(\epsilon_2 \bar{v})$.
(4) Set $j = j + 1$ and return to Step 1.

The main problem with the mean estimator is that it is susceptible to noise, especially outliers. To handle that problem, one often uses the median value instead. Next, we introduce the concept of a median estimator for a collection of contours under the framework of elastic analysis.

Median Estimation: Again, let $\beta_1, \beta_2, \ldots, \beta_n$ denote a set of given contours and $q_1, q_2, \ldots, q_n$ be the corresponding SRVFs. Then, the median curve (again an equivalence class of curves) is $[\tilde{q}] = \operatorname{argmin}_{[q] \in \mathcal{S}} \sum_{i=1}^n d_s([q], [q_i])$. A gradient-based approach for finding the geometric median on general Riemannian manifolds is given in [18]. A version of this procedure particularized to our framework is given as Algorithm 2. The only difference here from the mean algorithm is in the weights applied to the shooting vectors v_i to obtain the weighted average $\tilde{v}$. It is important to note that in some cases the mean and median algorithms may converge to a local solution. Also, in both algorithms we set ϵ_1 to 1% of the initial gradient and $\epsilon_2 = 0.3$. Once we have a median contour we can address the problem of outlier detection.

Algorithm 2 (Median): Let $\tilde{q}_0$ be an initial estimate of the median. Set $j = 0$ and ϵ_1, ϵ_2 to be small positive values.
(1) For each $i = 1, \ldots, n$, compute $v_i = \exp_{\tilde{q}_j}^{-1}(q_i)$.
(2) For each $i = 1, \ldots, n$, compute $d_i = d_s(\tilde{q}_j, q_i)$.
(3) Compute the update direction $\tilde{v} = \sum_{i=1}^n (v_i/d_i)(\sum_{i=1}^n 1/d_i)^{-1}$.
(4) If $\|\tilde{v}\| < \epsilon_1$, then stop. Else, update using $\tilde{q}_{j+1} = \exp_{\tilde{q}_j}(\epsilon_2 \tilde{v})$.
(5) Set $j = j + 1$ and return to Step 1.

Outlier Detection: We will take a distance-based approach to outlier detection. For

this purpose we will utilize the median rather than the mean because it is a more robust estimate in the presence of outliers. Given the sample median of the data, we proceed by computing the geodesic distances between the estimated median and each of the curves in the data ($d_s([\tilde{q}], [q_i]),\ i = 1, \ldots, n$). Using these distances we can compute the quartiles (Q_1, Q_3) and the interquartile range ($IQR = Q_3 - Q_1$) and label observations corresponding to distances that are greater than $Q_3 + kIQR$ as outliers. The choice of this cutoff is based on a pre-determined expected probability of type one error, and can be adjusted for different desired error probabilities by varying k. For the purpose of this paper, we use the standard $k = 1.5$.

Covariance Estimation and Random Sampling: In order to estimate the covariance structure of our data we will again utilize the estimated median rather than the mean. The covariance can be estimated with respect to any point on a Riemannian manifold but since we argue that the median provides a robust estimate of the center of mass of the data, that is the point we choose [22]. The evaluation of the covariance around the mean using elastic analysis has been previously discussed in the context of shapes of closed curves [23,17], 3D open curves [24] and surfaces [25]. The general computation of the covariance around the median is as follows. Let $v_i = \exp_{\tilde{q}}^{-1}(q_i),\ i = 1, 2, \ldots, n,\ v_i \in T_{\tilde{q}}(\mathcal{S})$. Then, the covariance kernel can be defined as a function $K_q : [0, 1] \times [0, 1] \rightarrow \mathbb{R}$ given by $K_q(\omega, \tau) = (1/(n-1)) \sum_{i=1}^{n} \langle v_i(\omega), v_i(\tau) \rangle$. In practice, since the curves have to be sampled with a finite number of points, say m, the resulting covariance matrices are finite-dimensional. Often the observation size n is much less than m and, consequently, n controls the degree of variability in the stochastic model. In the case of learning statistical models from the observations, one can reach an efficient basis for $T_{\tilde{q}}(\mathcal{S})$ using the traditional principal component analysis (PCA) as follows. Let $V \in \mathbb{R}^{2m \times n}$ be the observed tangent data matrix with n observations and m sample points in $\mathbb{R}^2$ on each tangent. Let $K \in \mathbb{R}^{2m \times 2m}$ be the Karcher covariance matrix and let $K = U\Sigma U^T$ be its SVD. The submatrix formed by the first r columns of U, call it $\tilde{U}$, span the principal subspace of the observed data and provide the observations of the principal coefficients as $C = \tilde{U}^T V \in \mathbb{R}^{r \times 2m}$. We can validate the computed models through random sampling. This can be done using a wrapped Gaussian distribution as follows. A multivariate Gaussian model for a tangent vector re-arranged as a long vector $v \in \mathbb{R}^{2m}$ is given by $v = \sum_{i=1}^{n} z_i \sqrt{\Sigma_{ii}} U_i$, where $z_i \stackrel{iid}{\sim} N(0, 1)$ and Σ_{ii} is the variance of the ith principal component. One can restructure the elements of v to form a matrix of size $\mathbb{R}^{2 \times m}$ and obtain a random SRVF using $q = \tilde{q} + v$. Finally, this SRVF can be mapped to a parameterized curve using integration.

Position Variability: One issue that we encounter in our framework is that the SRVF representation automatically removes the translation variability of the given contours (see the definition of SRVF given earlier). In order to include it in the analysis, we treat it separately by first computing the mean position of the given contours and then performing a joint computation of the covariance (including the contour information and an additional translation vector). This gives us a natural
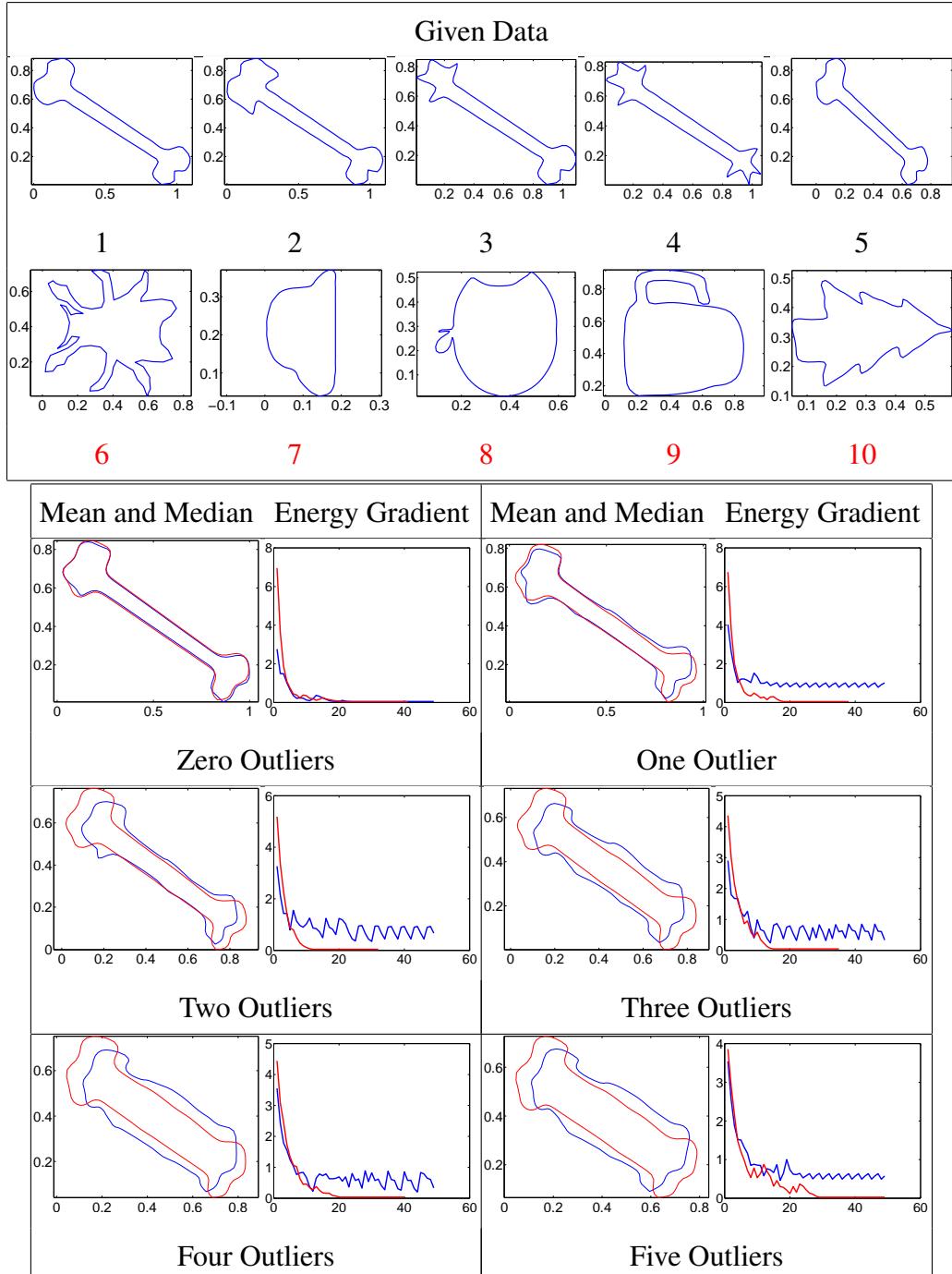
Fig. 4. Simulated dataset. Mean (blue) and median (red) estimation in the presence of out-liers. Gradient of the energy with respect to number of iterations (blue=mean, red=median).

way of including translation in summarizing the variability of our data using PCA.
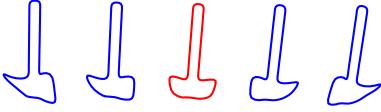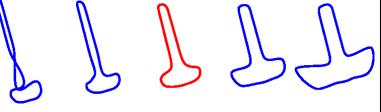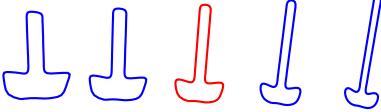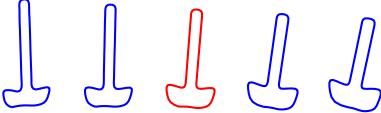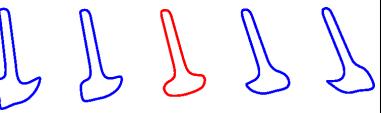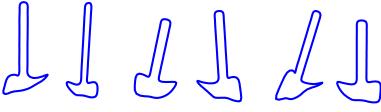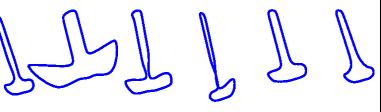
Fig. 5. Covariance structure and Gaussian random samples for a dataset of 20 hammers (left=elastic analysis, right=arc-length analysis).
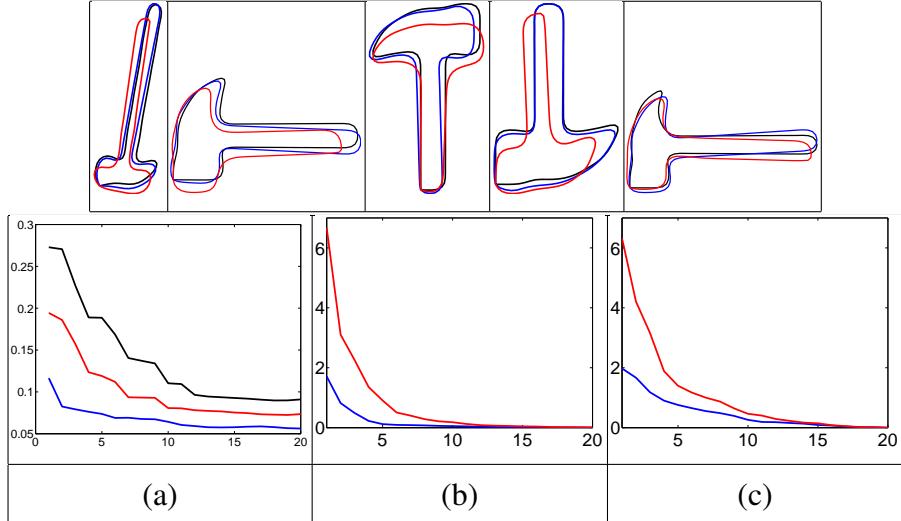

Fig. 6. Top: Example reconstructions using the first five PCs (black=original curve, blue=elastic analysis, red=arc-length analysis). Average leave-one-out distances between true and reconstructed curves as a function of the number of PCs used (blue=elastic analysis and d_s, red=arc-length analysis and d_s, black=arc-length analysis and d_c), for the hammer data (a). Variance of each principal component computed using elastic analysis (blue) and arc-length analysis (red) for the hammer example (b) and the heart example (c).

2.3 Illustration of the Framework

In this section we present multiple artificial examples to showcase the methods described in this paper. Each example presents a different type of result that emphasizes part of the described methodology. In most of these examples we compare
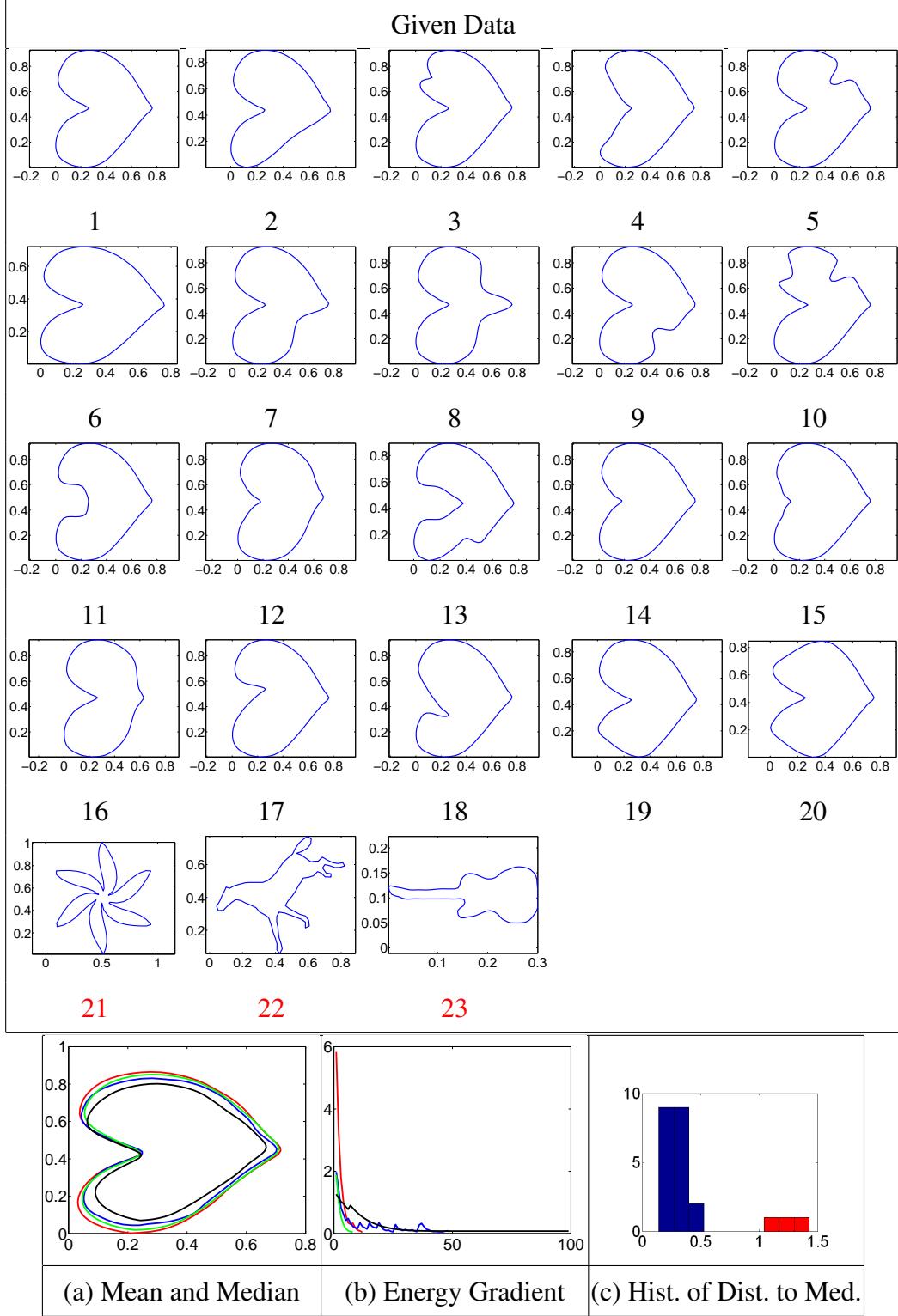
Fig. 7. Simulated dataset. (a) Estimated elastic mean (blue) and median (red), arc-length mean (black) and median (green). (b) Gradient of the energy (blue=elastic mean, red=elastic median, black=arc-length mean, green=arc-length median). (c) Histogram of distances from the median to the given data (red indicates the outliers).
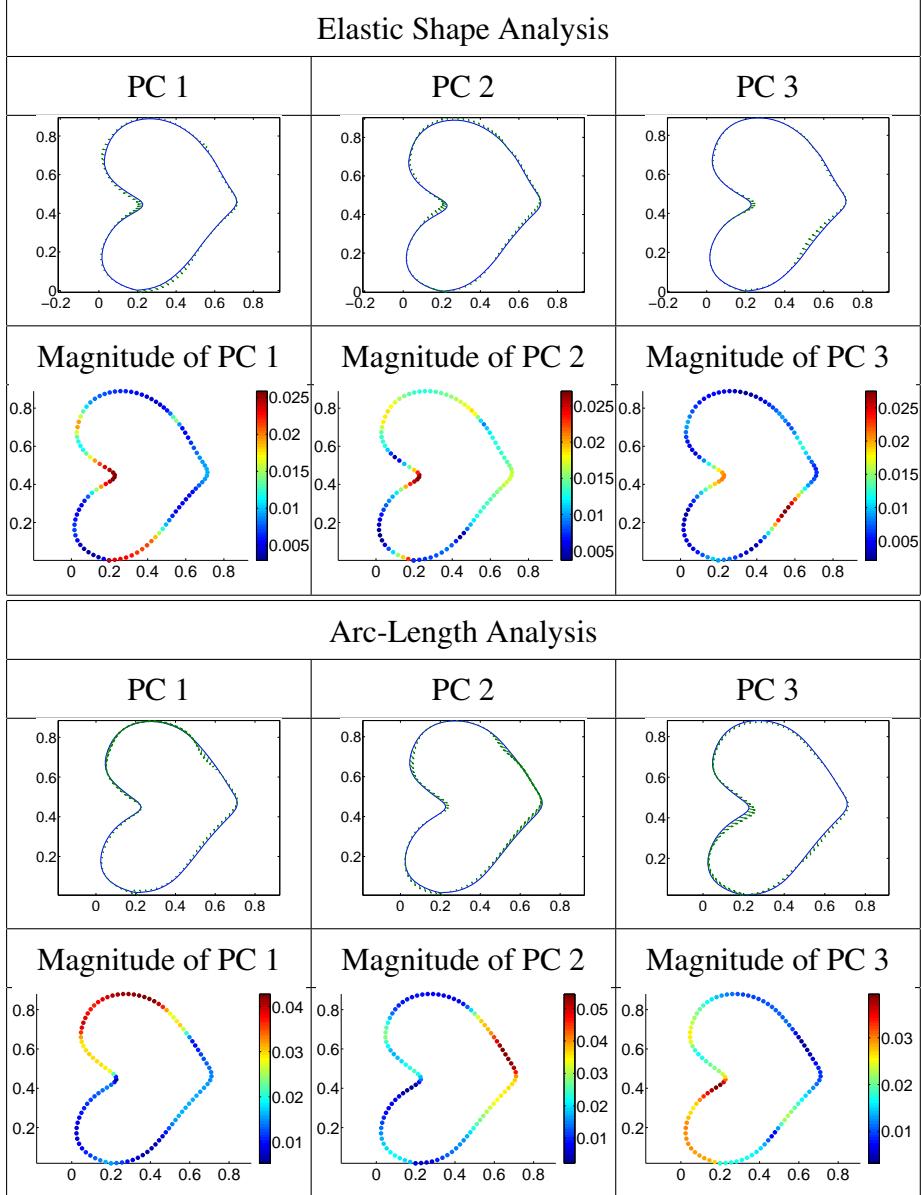
Fig. 8. Covariance structure in the data displayed using the first three principal components (top=elastic analysis, bottom=arc-length analysis).

the generated models to those computed using arc-length parameterization while removing the seed variability (using SRVFs).

Mean and Median Comparison: We begin with an example that showcases the robustness of the median estimation compared to the Karcher mean in the presence of clear outliers. In Fig. 4, we display ten contours that come from the MPEG 7 database; the first five come from the same class, while the next five are outliers. We performed the mean and median estimation six times, first using the first five curves with no outliers, second using the first five curves with one outlier (curve six), then with two outliers (curves six and seven), etc. We expect the estimated median to be more robust to outliers than the mean. We see a clear deterioration

13

of the shape of the estimated Karcher mean (blue) as we increase the number of outliers in the data. The median contour (red) is a much better estimate, although it also begins to deteriorate when the number of outliers becomes equal to the number of curves coming from the same class.
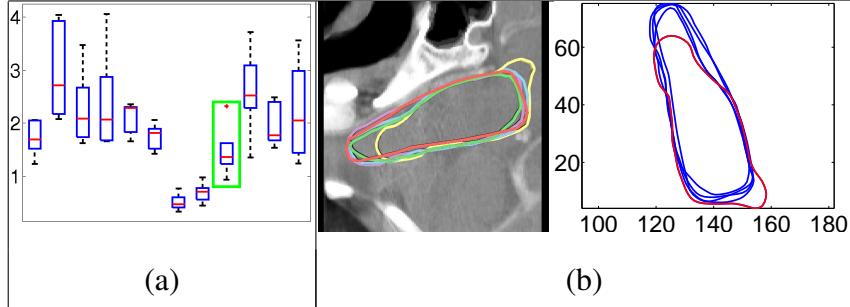

Fig. 9. (a) Boxplots of elastic distances to the median for all medical image examples (with Ex. 9 highlighted in green). (b) Data in Ex. 9 with the outlier highlighted in red.

Statistical Modeling, Example 1: In this example we would like to address the compactness, generalizablity and specificity of our models. For this purpose, we use a dataset of 20 hammer contours (from MPEG 7). For the sake of brevity we do not display the original dataset. It is important to note that in this example we remove the rotation variability of the data (in a pair-wise manner using Procrustes alignment). First, in Fig. 5, we display the first three modes of variation (within 2 standard deviations around the median drawn in red) in the computed models for both elastic and arc-length analysis as well as random samples from the Gaussian distribution. The shown elastic analysis principal directions of variation represent the given data well and thus the computed models appear to have good specificity. This can also be observed in the generated random samples, which are all valid. These properties are not true in the case of arc-length analysis. In fact, one of the random curves has self-crossings representing an invalid instance. Next, we want to quantify the generalizability of our models using leave-one-out reconstructions of the given data. For each of the curves in the data we compute its reconstruction with a certain number of PCs and compare it to the true curve using d_s defined in Section 2.1. In Fig. 6(a) we plot the distances (averaged over all curves in the data) as a function of the number of PCs used for reconstruction. The blue curve represents elastic analysis and the red one represents arc-length analysis. We note that the elastic models have significantly better generalizability than the arc-length ones. For each number of PCs we also performed a paired t-test at the 0.05 significance level, which rejected the null hypothesis that the difference between mean distances generated by the two methods is zero. In addition, for arc-length analysis, we computed the same result using d_c (also defined in Section 2.1). We plot this result in black in the same figure. As expected, these average distances are significantly larger than those shown in red. Additionally, we provide five qualitative reconstruction results where we plot the true curve in black, the elastic reconstruction in blue and the arc-length reconstruction in red. Finally, in Fig. 6(b), we plot the variance of each PC (elastic=blue, arc-length=red). We note that the elastic models

14

are much more compact than the arc-length ones.

Ex.	Data	Mean/Median
1		
	(a) Brainstem Oblique View 1	
2		
	(b) Brainstem Oblique View 2	
3		
	(c) Brainstem Oblique View 3	
4		
	(d) Brainstem Oblique View 4	
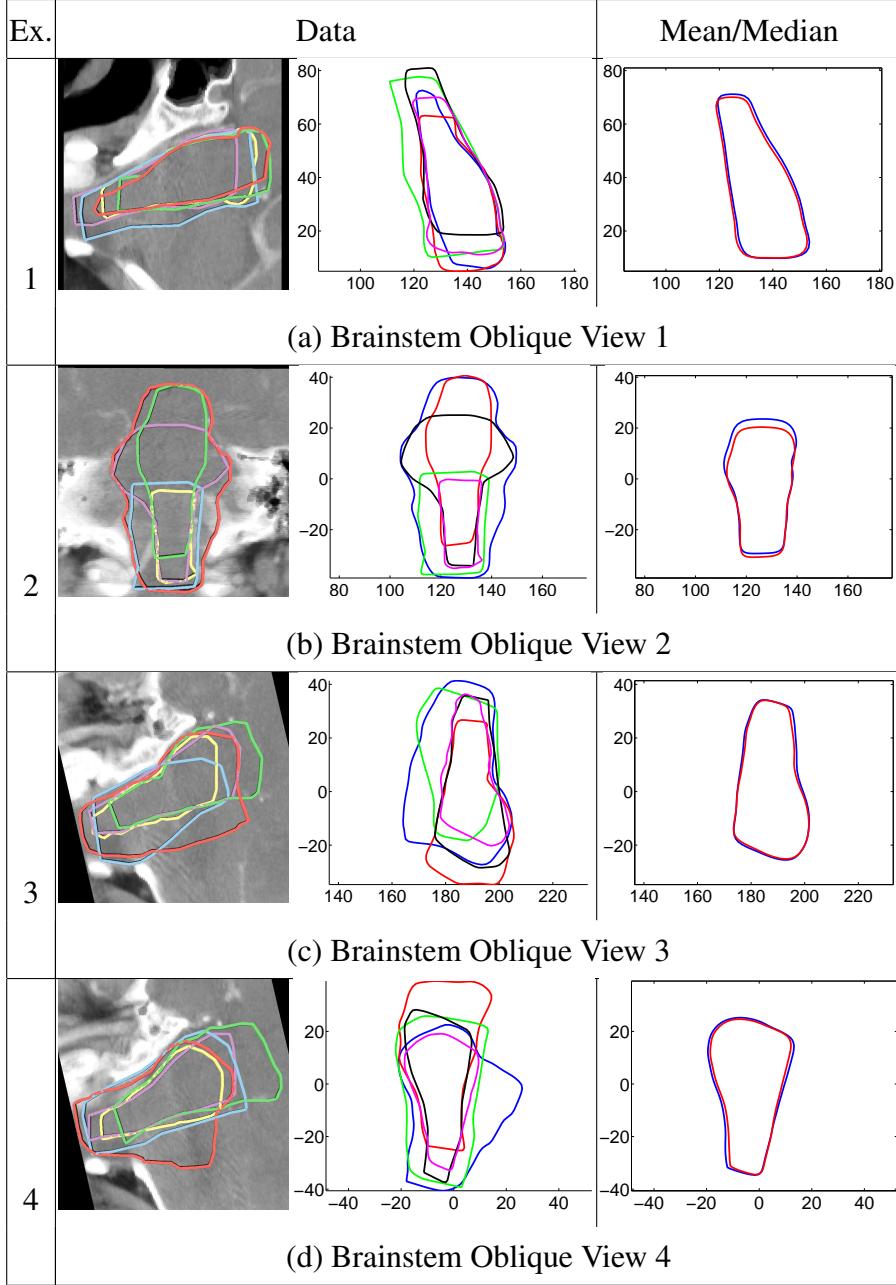
Fig. 10. Mean and median for five expert segmentations of the same brainstem structure using four oblique views.

Statistical Modeling, Example 2: In Figs. 7 and 8, we display results of using the above described statistical modeling techniques to analyze artificial heart contours (from MPEG 7). In this dataset, we have included three clear outliers to test whether our method can detect them. We began by computing the elastic mean (blue) and median (red), and the arc-length mean (black) and median (green) of the data. Again, we see that the medians provide a more robust statistical summary of the data, although the differences here are smaller due to an overwhelming number
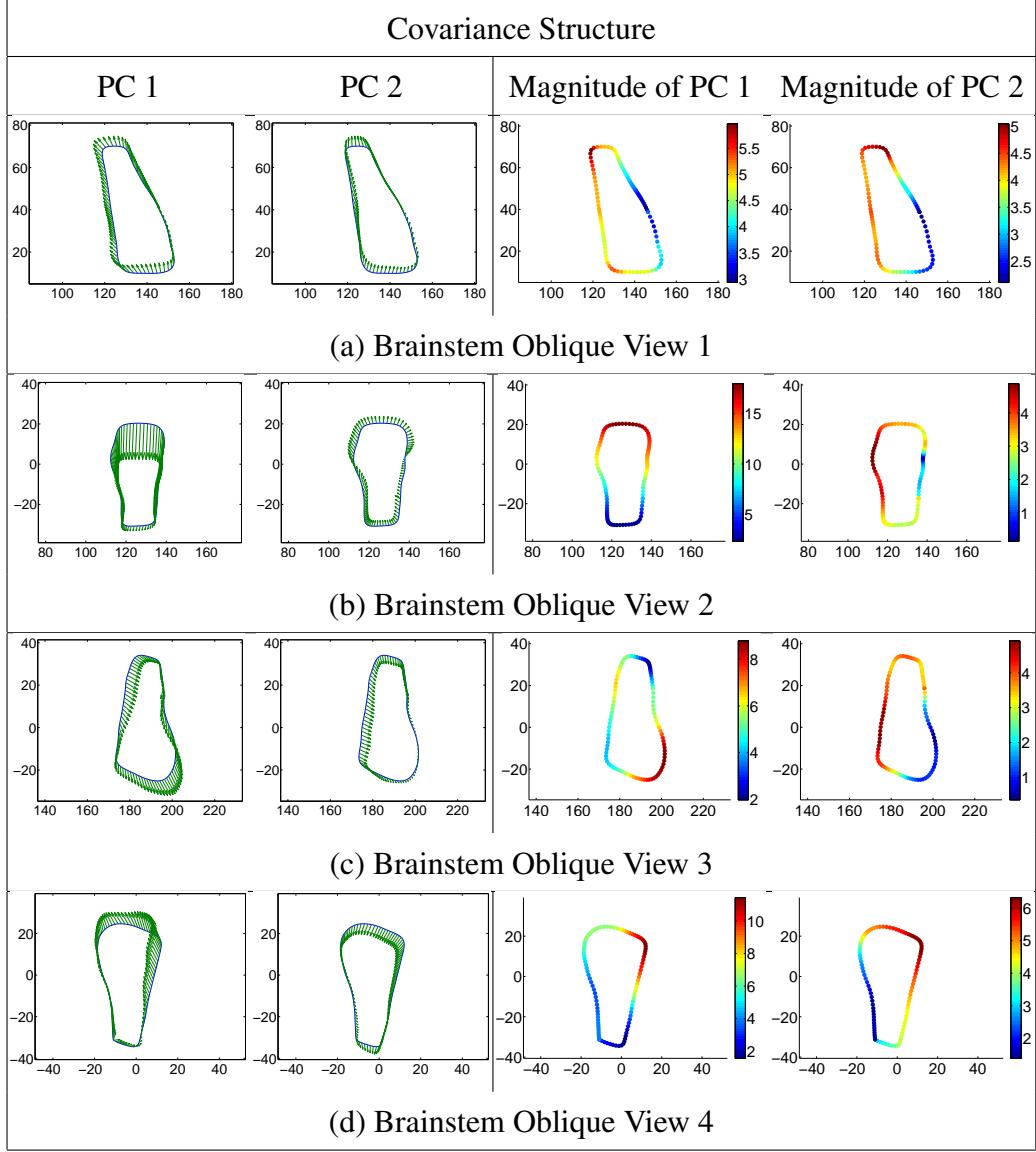
Fig. 11. Estimated covariance (in the form of principal components) for five expert segmentations of the same brainstem structure using four oblique views.

of inliers in the data. The difference here is mainly seen in the size of the estimate. The means are always smaller than the medians, which reflects the difference in scale of the inliers and the outliers. Furthermore, the arc-length mean and median are smaller than the elastic ones due to additional averaging out of features. The distance based outlier cutoff in this example is 0.5863 and we identified three clear outlying observations, corresponding to curves 21, 22 and 23. We removed these outliers from the data and computed the covariance structure using the re-estimated medians. It is important to note that after re-estimating the median and checking for outliers again, none of the contours were flagged. We display the resulting principal components as vector fields on the median curve (these are scaled according to the standard deviation of each principal component). The corresponding magnitude of each vector is also displayed using colors, where red areas correspond

to higher variability in the given data. We see that the main variation, computed using elastic analysis, matches the variability observed in the original data (overall shape, indent at top of the heart, indents on the sides of the heart). This implies that our models are good at capturing the underlying data structure. On the other hand, when arc-length analysis is used this trend is not as clear. Also, the main variation accounted for by PC1 corresponds to a nearly tangential vector field along the boundary of the median heart. This mainly represents a misalignment of points, and can cloud the overall interpretation of the model. Also, as shown in Fig. 6(c), the variance of PCs computed using elastic analysis (blue, cumulative variance=9.80) is much lower than that computed using arc-length analysis (red, cumulative variance=22.72) suggesting a more compact model.
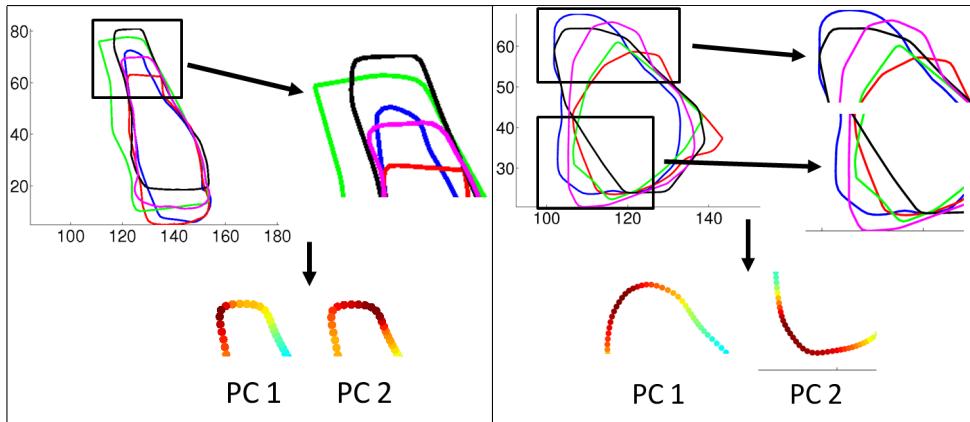

Fig. 12. Zoom-in on specific contour areas to highlight the appropriateness of elastic models for examples in Figs. 10(a) and 13(a).

3   Experimental Results on Medical Images

In this section we display examples of elastic statistical models computed for real segmentation data. The data used here comes from oblique slices of images of the brainstem and prostate, and a few parallel image slices of the same anatomical structures. In some examples we also compare expert segmentations to those performed by non-experts. In each example, we estimate both the mean and median of the manual segmentations. We also compute the covariance structure at the median. In computing these models, we first consider the outlier detection problem. Fig. 9 shows the boxplots of elastic distances to the median for each of the 12 presented examples. As can be seen, we identified an outlier in the non-expert segmentations in Ex. 9. This outlier was removed from the data before computing the elastic models.

In many of the presented examples we note that the median is, as stated earlier, a more robust estimate of the center of mass of the data than the mean. We also see that the computed models efficiently summarize the given data and that in most
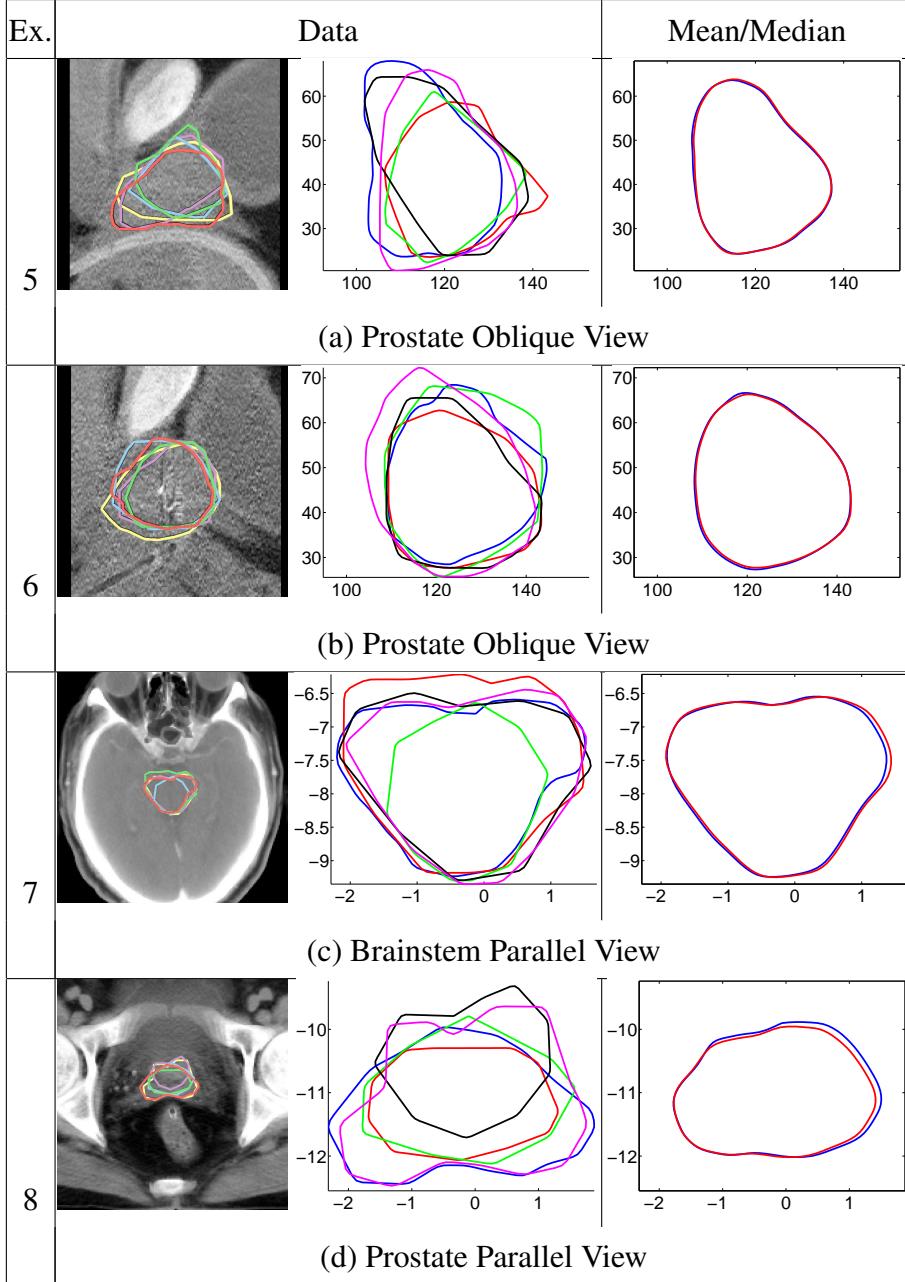
Fig. 13. Mean and median for five expert segmentations of structures coming from four different images.

cases only two principal components can account for most of the variation. The principal components, which represent the covariance structure we estimated are displayed as vector fields on the median curve. One important fact to note here is that we are only displaying the positive direction (median+PC) and the vectors were scaled according to one standard deviation from the median. We also provide a corresponding magnitude plot for each of the principal components. These visualizations can be used to identify areas on the segmented curves that the segmenters were unsure about (red colors correspond to more uncertainty/variability).

18
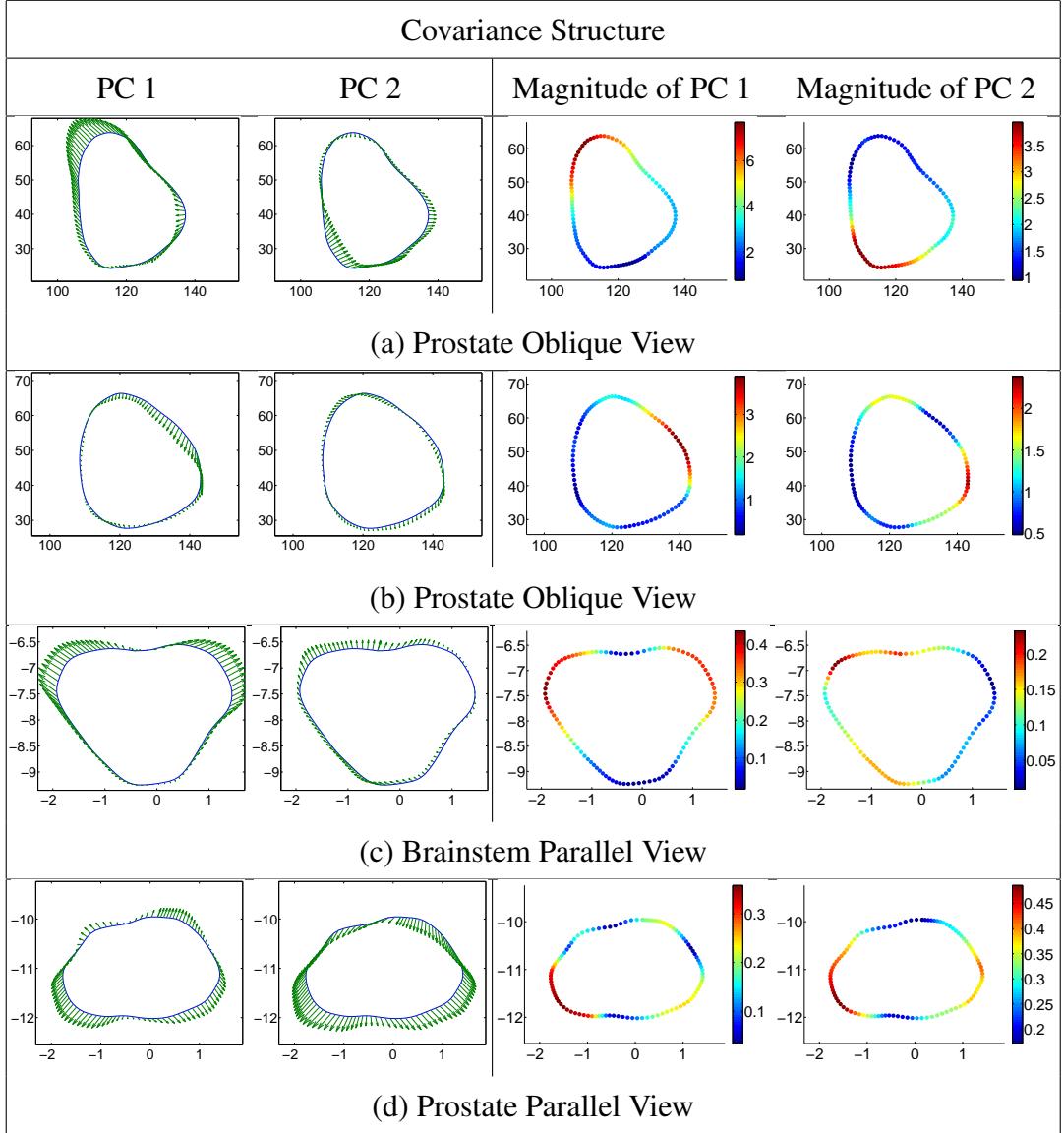
Fig. 14. Estimated covariance (in the form of principal components) for five expert segmentations of structures coming from four different images.

The computed uncertainty models can later be used as additional information in surface reconstruction from 3D images and image registration.

In the first four examples provided in Figs. 10 and 11 we consider four oblique views of the same brainstem structure. We are interested in modeling five manual expert segmentations for each view. We see that our estimates of the mean and median are valid representatives of the given data. In addition, the computed principal components efficiently reflect the variability in the given segmentations. In Figs. 13 and 14, we display some additional statistical models for expert segmentations of the brainstem and prostate structures using different oblique and parallel views. Again, we note that the models constructed using our method describe the given data well. In Fig. 12 we zoom-in on the segmentation variability presented in Figs.
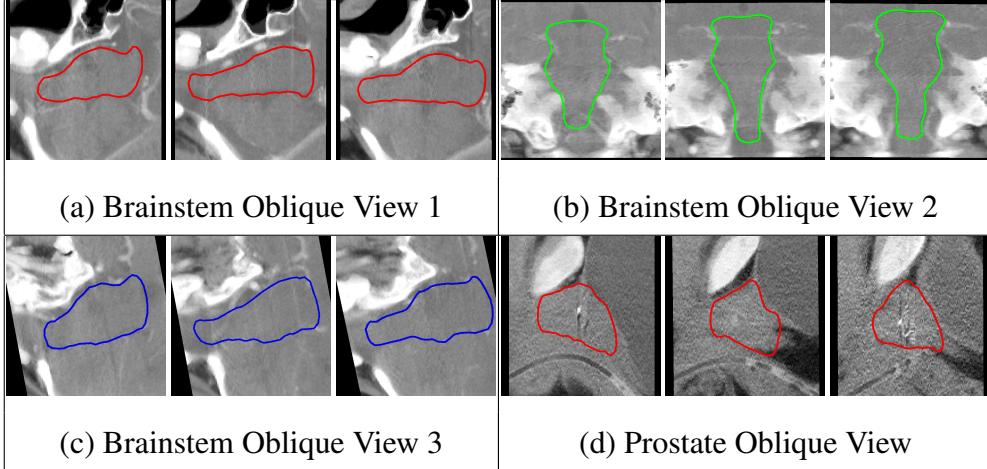
Fig. 15. Example images with segmentations shown to the non-experts.

11(a) and 14(a) for improved display. We also highlight areas of the segmentations where most variability is observed. It is clearly evident that the magnitude of the principal components in those areas is high (marked in red), which implies that our models are good at capturing such variability.

Finally, in Figs. 16, 17, we provide statistical models for segmentations of the brainstem and prostate performed by non-experts. In two of the examples we are modeling ten contours. These segmentations were performed on the same images as those shown in Figs. 10, 11(a-c) and Figs. 13, 14(a) for expert segmenters. This is a view angle that the experts had never seen before when contouring (they always contour on transverse slices) and they were only indirectly aware of where the slice was [26]. In contrast, the non-experts were given three example segmentations on a different image (same view) as references, which are displayed in Fig. 15. They were also presented with simple instructions to look for specific anatomical cues in the image in order to accurately delineate the structure of interest. The contours in the example images were made by intersecting the nonparallel planes with a surface model reconstructed from many physician-reviewed contours drawn on the parallel cross-sections of the dataset. We notice a very interesting result in these two models. The non-experts had lower variability in their segmentations than the experts. We hypothesize that this trend is due to the given example segmentations. Perhaps in some cases the non-experts were trying to emulate the given example images by finding similar shapes in the image data.

4 Summary

We have presented a comprehensive set of methods for statistical analysis of closed, planar contours. These methods are especially useful in what is termed Citizen Science, where non-experts help scientists collect and analyze real world data. In our framework, we consider modeling uncertainty of manual segmentations performed
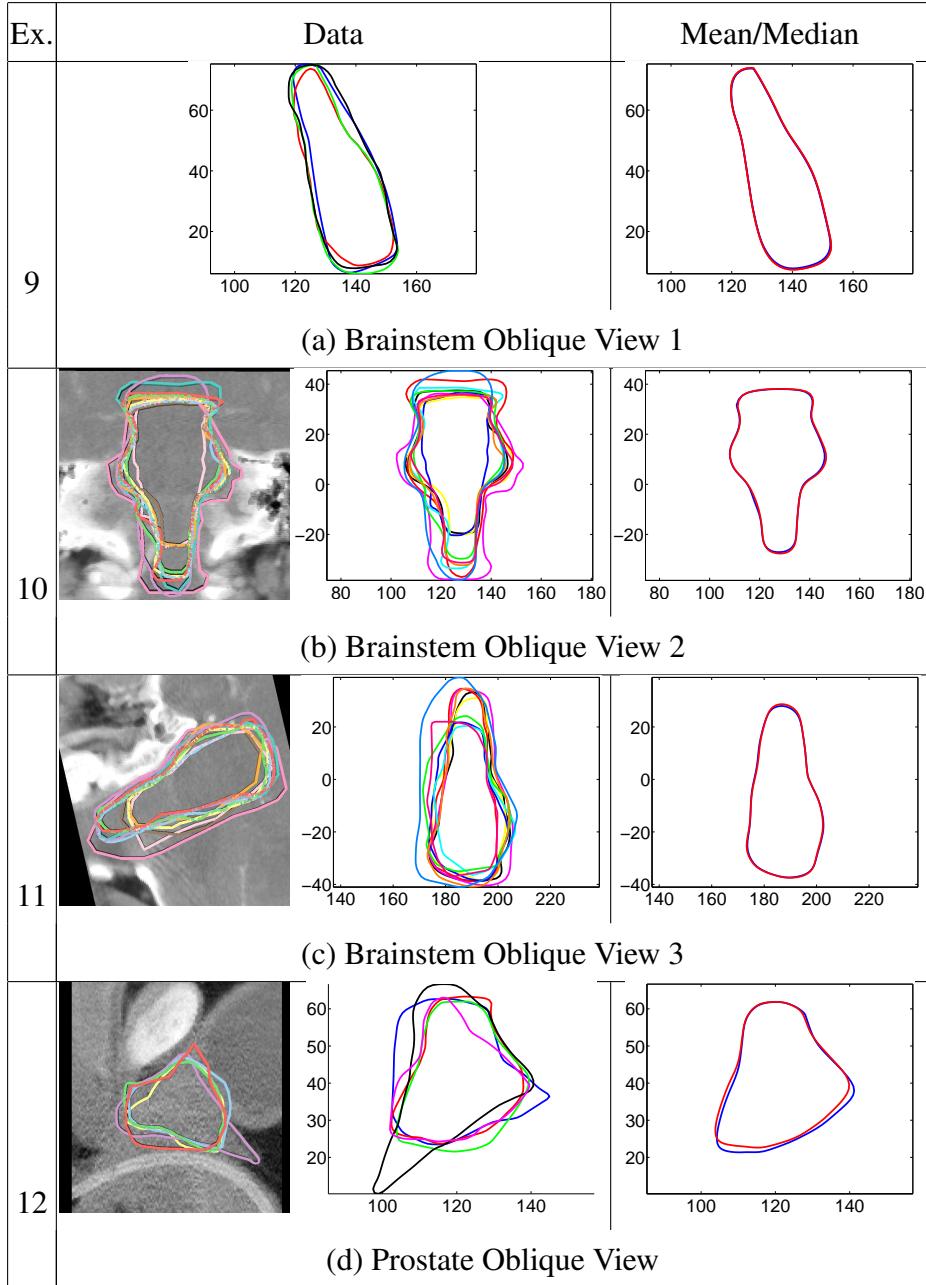
Fig. 16. Mean and median for non-expert segmentations. The corresponding statistical models for expert segmentations are presented in Figs. 10, 11(a-c) and Figs. 13, 14(a).

on medical images. These types of statistical analyses of contours are very important in medical surface reconstruction and medical image registration. We have provided multiple examples on simulated data in order to validate our methods. Furthermore, we have provided results on real segmentations of the brainstem and prostate.

In the future, we would like to explore how influential the example images shown to non-experts are in their segmentations. We would like to design the Citizen Sci-
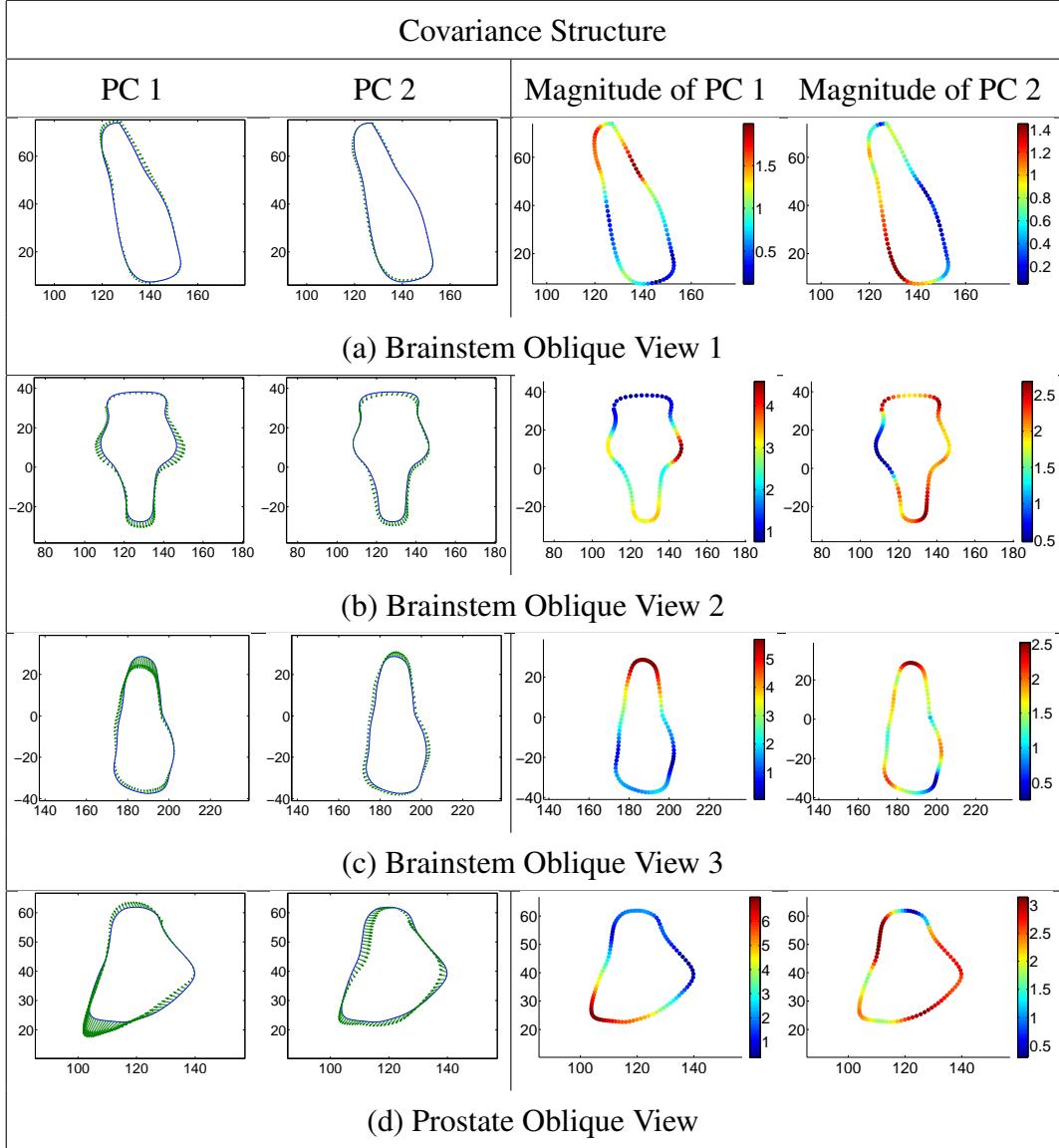
Fig. 17. Estimated covariance for non-expert segmentations of the brainstem and prostate structures.

ence studies such that there is an equilibrium between the non-expert segmenters' own knowledge and the guidance provided by the example images. Furthermore, we would like to analyze the effectiveness of different segmenters based on inter and intra-operator variability. An interesting additional future problem is the reconstruction of 3D objects, i.e. surfaces, using estimated 2D contours generated from different slices of volume image data.

References

[1] C. Lintott, K. Schawinski, A. Slosar, K. Land, S. Bamford, D. Thomas, M. Raddick, R. Nichol, A. Szalay, D. Andreescu, P. Murray, J. Vandenberg, Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey, Monthly Notices of the Royal Astronomical Society 389 (3) (2008) 1179–1189.

[2] S. Cooper, F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beenen, A. Leaver-Fay, D. Baker, Z. Popovic, Predicting protein structures with a multiplayer online game, Nature 466 (7307) (2010) 756–760.

[3] C. L. of Ornithology, Citizen science, `http://exoplanet.eu/catalog.php`.

[4] I. L. Dryden, K. Mardia, Statistical Shape Analysis, John Wiley & Son, 1998.

[5] D. G. Kendall, D. Barden, T. K. Carne, H. Le, Shape and shape theory, Wiley, 1999.

[6] C. G. Small, The Statistical Theory of Shape, Springer, 1996.

[7] S. Osher, R. Fedkiw, Level Set Methods and Dynamic Implicit Surfaces, Springer Verlag, 2003.

[8] U. Grenander, M. I. Miller, Computational anatomy: An emerging discipline, Quarterly of Applied Mathematics LVI (4) (1998) 617–694.

[9] P. T. Fletcher, S. Joshi, C. Lu, S. M. Pizer, Principal geodesic analysis for the study of nonlinear statistics of shape, IEEE Transactions on Medical Imaging 23 (8) (Aug. 2004) 995–1005.

[10] K. Gorczowski, M. Styner, J. Jeong, J. Marron, J. Piven, H. Hazlett, S. Pizer, G. Gerig, Multi-object analysis of volume, pose, and shape using statistical discrimination, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (4) (2010) 652–666.

[11] R. Davies, C. Twining, T. Cootes, C. Taylor, Building 3-d statistical shape models by direct optimization, IEEE Transactions on Medical Imaging 29 (4) (2010) 961–981.

[12] R. Davies, C. Twining, T. Cootes, J. Waterton, C. Taylor, A minimum description length approach to statistical shape modeling, IEEE Transactions on Medical Imaging 21 (2002) 525–537.

[13] W. Mio, A. Srivastava, S. H. Joshi, On shape of plane elastic curves, International Journal of Computer Vision 73 (3) (2007) 307–324.

[14] L. Younes, P. W. Michor, J. Shah, D. Mumford, R. Lincei, A metric on shape space with explicit geodesics, Matematica E Applicazioni 19 (1) (2008) 25–57.

[15] G. Sundaramoorthi, A. Mennucci, S. Soatto, A. Yezzi, Tracking deforming objects by filtering and prediction in the space of curves, in: Proceedings of IEEE Conference on Decision and Control, 2009.

[16] S. H. Joshi, E. Klassen, A. Srivastava, I. H. Jermyn, A novel representation for Riemannian analysis of elastic curves in $\mathbb{R}^n$, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–7.

[17] A. Srivastava, E. Klassen, S. H. Joshi, I. H. Jermyn, Shape analysis of elastic curves in Euclidean spaces, IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (7) (2011) 1415–1428.

[18] P. Fletcher, S. Venkatasubramanian, S. Joshi, The geometric median on Riemannian manifolds with application to robust atlas estimation, Neuroimage 45 (1) (2009) S143–S152.

[19] S. H. Joshi, E. Klassen, A. Srivastava, I. Jermyn, Removing shape-preserving transformations in square-root elastic (SRE) framework for shape analysis of curves, in: Proceedings of International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition, 2007, pp. 387–398.

[20] E. Klassen, A. Srivastava, Geodesics between 3D closed curves using path-straightening, in: Proceedings of European Conference on Computer Vision, 2006, pp. 95–106.

[21] H. Le, Locating frechet means with application to shape spaces, Advances in Applied Probability 33 (2) (2001) 324–338.

[22] X. Pennec, Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements, Journal of Mathematical Imaging and Vision 25 (1) (2006) 127–154.

[23] A. Srivastava, S. H. Joshi, W. Mio, X. Liu, Statistical shape anlaysis: Clustering, learning and testing, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (4) (2005) 590–602.

[24] S. Kurtek, A. Srivastava, E. Klassen, Z. Ding, Statistical modeling of curves using shapes and related features, Journal of the American Statistical Association 107 (499) (2012) 1152–1165.

[25] S. Kurtek, E. Klassen, Z. Ding, M. Avison, A. Srivastava, Parameterization-invariant shape statistics and probabilistic classification of anatomical surfaces., in: G. Szkely, H. K. Hahn (Eds.), Proceedings of Information Processing in Medical Imaging, Vol. 6801 of Lecture Notes in Computer Science, Springer, 2011, pp. 147–158.

[26] R. Sowell, L. Liu, T. Ju, C. Grimm, C. Abraham, G. Gokhroo, D. Low, Volume viewer: an interactive tool for fitting surfaces to volume data, in: Proceedings of the 6th Eurographics Symposium on Sketch-Based Interfaces and Modeling, 2009, pp. 141–148.