

# Unsupervised learning of an atlas from unlabeled point-sets

Haili Chui	Anand Rangarajan, Jie Zhang	Christiana Leonard
Medical Imaging Group	Dept. of CISE	Dept. of Neuroscience
R2 Technologies	Univ. of Florida	Univ. of Florida
Sunnyvale, CA	Gainesville, FL	Gainesville, FL
hailic@r2tech.com	{anand,jiezhang}@cise.ufl.edu	leonard@ufbi.ufl.edu

Corresponding author: Anand Rangarajan  
Department of Computer and Information Science and Engineering  
University of Florida  
P. O. Box 116120  
Gainesville, FL 32611-6120, USA  
Phone: +1 352 392 1507  
Fax: +1 352 392 1220  
E-mail: *anand@cise.ufl.edu*

## **Abstract**

One of the key challenges in deformable shape modeling is the problem of estimating a meaningful average or mean shape from a set of unlabeled shapes. We present a new joint clustering and matching algorithm that is capable of computing such a mean shape from multiple

shape samples which are represented by unlabeled point-sets. An iterative bootstrap process is used wherein multiple shape sample point-sets are non-rigidly deformed to the emerging mean shape, with subsequent estimation of the mean shape based on these non-rigid alignments. The process is entirely symmetric with no bias toward any of the original shape sample point-sets. We believe that this method can be especially useful for creating atlases of various shapes present in medical images. We have applied the method to create mean shapes from nine hand-segmented 2D corpus callosum data sets and ten hippocampal 3D point-sets.

## **1 Introduction**

The study of deformable shapes has recently been a very active area in computer vision and medical imaging. Most of the effort has been primarily focused on understanding deformable shapes using a statistical approach. Analysis of deformable shapes has in turn helped create automated segmentation tools. A statistical understanding of shapes, their representations and deformations is therefore vitally important in many segmentation tasks. Statistical shape analysis using active shape models (Cootes et al., 1995) is a representative example in this category. After the shapes of a certain single structure but from multiple subjects are extracted, much information can be gained by identifying, measuring and characterizing the shapes. Through such analysis, it is hoped that the subtle differences between different groups or populations, e.g. a normal group vs a diseased group, can be discovered and made available to aid clinical diagnosis. Recent work in brain image analysis (Guimond et al., 2000; Davatzikos, 1997; Thompson et al., 1997) can be seen as examples in this category.

## 1.1 Shape representations

Deformable shapes can have different representations. Curves and surfaces of the shape boundary can obviously be used (Sebastian et al., 2000; Tagare, 1999). Despite being intuitively natural choices for shape representation, curves and surfaces are somewhat difficult to directly use in statistical analysis. Using an intrinsic parameterization is difficult when the goal is atlas creation. Using one of many extrinsic parameterizations typically leads to a mixed representation defined on point locations and spline coefficients. Statistical shape analysis in this representation requires learning a density function defined on locations *and* coefficients.

Other than curves or surfaces, another natural choice is to use point location information. A point-set representation has the main advantage that statistical shape analysis can be well formulated (Cootes et al., 1995) in that space. Not surprisingly, many recent statistical methodologies for shape analysis (Bookstein, 1989; Cootes et al., 1995; Duta et al., 1999) use the point-set representation. Our work also focuses on using points to study deformable shapes.

## 1.2 Statistical shape analysis with unknown correspondence

Given a set of deformable shapes represented by point-sets, basic statistical measures such as the mean and the covariance enable us to quantify the set of shapes at the very least up to second-order statistics. The mean point-set is usually a placeholder for a normal representative shape. The covariance information, usually of the form of a high dimensional covariance matrix, further provides us with information about how the shapes can deform and vary from the mean shape (in a second-order statistical sense). By examining the covariance matrix's eigen-vectors and eigen-values, we can also observe the dominant variations or deformations present in the shape group (Cootes et al., 1995). If higher-order statistical information is desired, recent techniques such as Independent Component Analysis (ICA) can be pressed into service.

The primary technical challenge in using point-set representations of shapes is the *correspondence* problem. Computing a meaningful mean point-set (and then the covariance matrix) from multiple shape point-sets is only possible when the correspondences between all the shape point-sets are known. Automated correspondence estimation in shape point-sets is a non trivial task because of several reasons.

Typically, correspondences can be estimated once the point-sets are properly aligned with appropriate spatial transformations. An adequate transformation would obviously be a deformation, i.e., a non-rigid spatial mapping, since the objects at hand are deformable. Solving for non-rigid deformations between point-sets with unknown correspondence is a hard problem (Chui and Rangarajan, 2000). In fact, many current methods only attempt to solve for rigid transformations, e.g. affine transformations, for the alignment (Duta et al., 1999). The correspondences resulting from these methods are, therefore, only rough approximations since they ignore the non-rigid aspect of the transformation. However, when dealing with point-sets, there is another aspect of the atlas problem, which we believe is more fundamental, and has largely been neglected (in the point-set literature).

### **1.3 Joint clustering and matching**

The problem lies in the fact that the shape sample point-sets may not be *consistent* in the first place, i.e., points in each set may not be positioned at corresponding locations. Due to the vagaries of feature extraction, points in one set may not have exact counterparts in another point-set. Without taking into account this feature extraction and sampling issue, the resulting correspondence achieved by any alignment method (rigid or non-rigid) is doomed from the start.

We face a dilemma here. On the one hand, we need to compare the point-sets, possibly through alignment, to each other in order to establish correspondence. On the other hand, the alignment

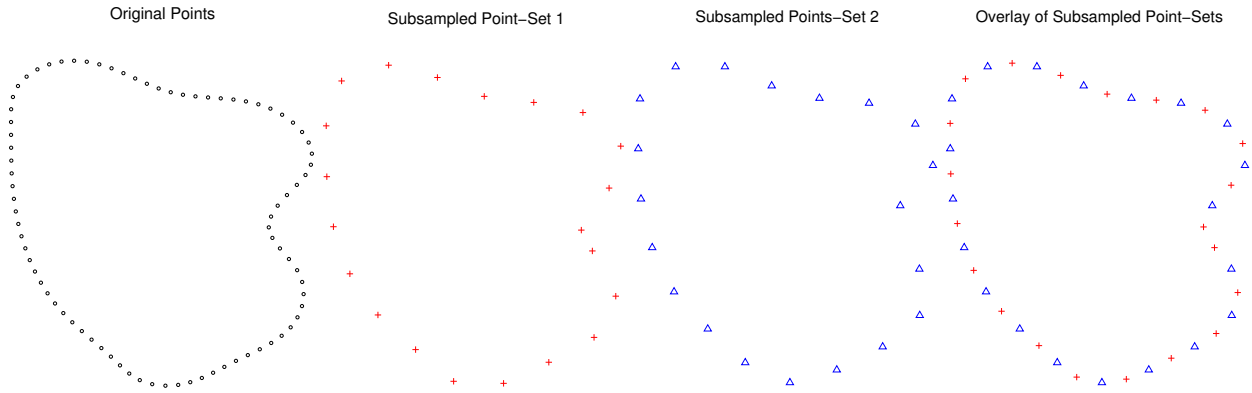


Figure 1: The consistency problem. From left to right: i) the original point-set with densely distributed points; ii, iii) two subsampled point-sets; iv) the overlay of the subsampled point-sets. Note the difference between the two subsampled point-sets in the overlay.

won't be entirely successful unless the sample point-sets have already satisfied the consistency criterion. The dilemma clearly shows that any process of using fixed sample point-sets to do the alignment is flawed. The problem is graphically depicted in Figure 1.

Finally, we also encounter the bias problem in atlas creation. Since we have more than two sample point-sets to align, a question that arises is: how do we align all the point-sets in a symmetric manner so that there is no bias toward any particular point-set? One way to circumvent this problem is to define a mean point-set and then align every sample point-set to the mean. We see this as an opportunity for mutual improvement instead of a fixed unidirectional process. As explained later, our approach consists of mutual refinement of the mean shape and the non-rigid deformations from the point-sets to the emerging mean.

After surveying the fundamental problems in atlas creation, we summarize our approach. The basic idea is to insist that point to point correspondences are only valid for cluster centers in each point-set and not for the original points. However, if we separately cluster each point-set, then we are not guaranteed that the cluster center placements will be in accord. Consequently, we use

a *joint* clustering and matching approach to overcome the consistency problem. The matching is performed on *movable* cluster centers and not on the original points. The cluster centers are also used to estimate and re-estimate the atlas or mean shape through non-rigid matching and averaging. The estimated variables comprise (i) the cluster centers, (ii) the atlas (mean shape), and (iii) the non-rigid transformations from (to) the cluster centers and to (from) the atlas. We have designed a joint clustering and matching algorithm to carry out this iterative process. Embedded within a *deterministic annealing* scheme, the algorithm gradually refines all three sets of variables until convergence.

## 2 Review

After discussing the underlying motivations and rationale behind our method, we take a step back and briefly summarize *some* representative previous research work in this area. We also focus on the differences between these methods and ours.

We first draw a fundamental distinction between intensity-based and feature-based methods. Intensity-based methods using mutual information (Rohlfing et al., 2001), (Rueckert et al., 2001) or distance transforms (Golland et al., 1999) as a non-rigid registration measure between the emerging atlas and a training set of images are very different from the feature-based atlas estimation method presented in this paper. Intensity-based methods do not face the correspondence problem. However, these methods are forced to deal with variability in illumination.

The work presented in (Sebastian et al., 2000) is a representative method using an intrinsic curve parameterization to analyze deformable shapes. It is very close to previous efforts in (Tagare, 1999) and (Davatzikos, 1997) where intrinsic curve properties, like the curvature and the arc length, are used to align curves. In (Sebastian et al., 2000), the alignment is done in a pairwise manner. To somewhat reduce the bias introduced by this non-symmetrical procedure, the sample curves

take turns in being the reference curve (atlas). The authors considered the symmetric alignment of multiple sample shapes to an emerging mean shape to be “intractable.” We are going to argue the opposite view in this paper. As with other methods using intrinsic curve or surface representations, further statistical analysis on these representations is much more difficult than analysis on the point representation *but the payoff may be greater* due to the use of intrinsic higher-order representations.

The active shape model proposed in (Cootes et al., 1995) utilized points to represent deformable shapes. Their work successfully initiated the research efforts in building point distribution models to understand deformable shapes (Cootes et al., 1995; Wang and Staib, 2000). To be able to build such a model, a set of training data, which are a group of shape samples with known correspondences, are required. The training data are normally acquired through a more or less manual landmarking process where a person goes through all the samples and attempts, to the best of his/her knowledge, to mark corresponding points on each sample. It is a rather tedious process and the accuracy is limited. Since the derived point shape distribution is only going to be as good as the training data, it is highly desirable to automate the training data generation process and to make it more accurate. In recent work (Hill et al., 2000), the authors attempt to overcome this limitation by attempting to automatically solve for the correspondences in a non-rigid setting. The resulting algorithm is very similar to the earlier work in (Tagare, 1999) and is restricted to curve correspondence. Finally, this work, while still being limited to curve correspondence, has been extended to atlas building (Davies et al., 2002).

Bookstein pioneered the usage of non-rigid spatial mapping functions - specifically *thin-plate* splines, to analyze deformable shapes (Bookstein, 1989). Since this method is landmark-based, it avoids the correspondence problem but suffers from the typical problems besetting landmark methods. Note that landmark-based methods largely overcome the consistency problem since

the placement of corresponding points is driven by the visual perception of experts and not by an automated algorithm. In our previous work (Chui et al., 2003), we have used a very similar approach to the one presented here to register two point-sets (with different point counts) using thin-plate splines. In this previous work, we have also conducted an extensive review of methods that attempt to simultaneously solve for correspondences and deformations.

The work in (Duta et al., 2001) also uses 2D points to learn shape statistics. The method is quite similar to the active shape model method (Cootes et al., 1995) except that more attention has been paid to the training data (or shape sample point-sets) generation process. The shape samples are first represented as curves and aligned with a rigid transformation. One curve is chosen as the reference and points are uniformly placed on that curve to form a reference point-set. To determine the location of corresponding points on the other curves, the extrinsic curvature information is compared. Points are allowed to slide along the curves to improve the agreement between the curvatures. This shape learning method improves the consistency of the shape samples under the assumption that the curvature information can be reliably computed for the shapes at hand. Since the extrinsic curvature is a rigid invariant, only a rigid mapping can be used. The process is not symmetric.

There is a long tradition in the statistics literature of statistical shape analysis (Kendall, 1984), (Small, 1996). In this literature, the main emphasis is in defining shape spaces in such a way that further statistical analysis is well defined. To some extent, we have sidestepped these problems by using point-set representations of shapes rather than curves or surfaces. Also, the main drawback of this earlier work is the lack of attention paid to the correspondence problem. We hope that we have convincingly argued the case for the importance of simultaneously addressing the correspondence problem, non-rigid deformations and atlas estimation.

## 3 Methodology

### 3.1 Intuitive description of our approach

The point correspondence problem, the shape sample consistency problem and the mean shape estimation problem are all inter-related to each other. Rather than treating each of them as a separate problem, we propose to regard them as three interlocking steps in a more general framework wherein we can simultaneously achieve meaningful answers to all three problems. We also believe that directly solving for non-rigid transformations, rather than using indirect information criteria or relying on curvature information, has many advantages in terms of efficiency and accuracy. By directly modeling the deformation, the results will be intuitively easier to evaluate. Using non-rigid transformations, rather than rigid transformations is also one of the key aspects in improving the estimation of correspondence.

The three inter-locking steps in our atlas estimation framework are (i) clustering, (ii) non-rigid mapping estimation and (iii) atlas (mean shape) estimation. The clustering step allows the cluster centers to be repositioned (to slide along the original points) so that they have a better chance of being consistent. The overall Euclidean distance between the cluster centers and the original point-set is the criterion for shape representation. The cluster center sets are associated with the emerging mean shape using non-rigid deformations. Since there is only one mean, this also provides an indirect connection between the sample cluster center sets. A spline deformation energy is the criterion for shape matching. With consistent sample cluster center sets positioned at corresponding locations, and with multiple non-rigid transformations accounting for the shape differences between the samples and the emerging mean, we apply the transformations and recompute the atlas (mean shape) by averaging over the warped cluster center sets. The average Euclidean distance of all warped cluster center sets to the mean shape is the criterion for atlas estimation.

A good way to understand this process is through the following: if the shape sample of interest looks like a curved structure, the sample points would then all fall on this “imaginary curve.” After clustering, the cluster centers from the sample points will also be constrained to roughly lie on the imaginary curve because otherwise they would not be a good shape representation of the original sample points. So the curve-like shape structure will always be maintained. However, this doesn’t prevent the cluster centers from having the extra freedom of being able to slide along the “imaginary curve” so that they can be better positioned to satisfy the consistency criteria.

### 3.2 Learning an atlas as a density estimation problem

Our approach to learning an atlas is based on density estimation. The basic idea is that the atlas is that point-set which best explains all the data point-sets. Since the atlas and the data are not in the same space, the above statement is unpacked to mean that the atlas best explains each data point-set once it has been registered and warped into the space of the data point-set. Consequently, the non-rigid registration parameters which take the atlas onto the data have to be included in the estimation. The overall estimation problem can be situated within a Bayesian maximum *a posteriori* (MAP) framework.

A Gaussian mixture model (McLachlan and Basford, 1988) is used to model each data point-set. Since the feature point-sets are usually highly structured, we can expect them to cluster well. Each point-set has the same number of cluster centers. Correspondence is automatic since the same index is used for the cluster centers in all point-sets and the atlas. While the indices are common, the locations of the cluster centers will obviously be different.

Our notation is as follows. The data point-sets are denoted by  $\{X^p, p \in \{1, \dots, P\}\}$ . Each point-set  $X^p$  consists of points  $\{x_i^p \in \mathcal{R}^D, i \in \{1, \dots, N_p\}\}$ . Since a Gaussian mixture model is used to model each point-set, we define a set of cluster centers, one for each point set. The cluster

center point-sets are denoted by  $\{V^p, p \in \{1, \dots, P\}\}$ . Each  $V^p$  consists of cluster locations  $\{v_a^p \in \mathcal{R}^D, a \in \{1, \dots, K\}\}$ . Note that there are  $K$  points in each  $V^p$ . The number of clusters are chosen to be the same in each  $V^p$  and in the atlas. The atlas point-set is denoted by  $Z$  and consists of points  $\{z_a \in \mathcal{R}^D, a \in \{1, \dots, K\}\}$ . We begin by specifying the density function of each point-set.

$$p(X^p|V^p, \pi^p) = \prod_{i=1}^{N_p} \sum_{a=1}^K \pi_a^p p(x_i^p|v_a^p). \quad (1)$$

In (1),  $p(X^p|V^p, \pi^p)$  is a mixture model containing component densities  $p(x_i^p|v_a^p)$ . The occupancy probability which is different for each data point-set is denoted by  $\pi^p$ . Later, we specialize the component density to a Gaussian and set the occupancy probability to uniform in order to simplify (the gargantuan) atlas estimation procedure. Having specified the density function of the data, we turn to the specification of the density function of the individual cluster centers  $\{V^p, p \in \{1, \dots, P\}\}$ .

$$p(V^p|Z, f^p, g^p) = \frac{1}{Z_{\text{partition}}^V} \exp\left(-\sum_{a=1}^K |v_a^p - g^p(z_a)|^2 - \sum_{a=1}^K |z_a^p - f^p(v_a)|^2\right). \quad (2)$$

The density function in (2) does not have a straightforward interpretation in terms of canonical parametric densities. Nonetheless, the basic idea in (2) is that each  $V^p$  is related to  $Z$  via a pair of functions  $(f^p, g^p)$ . The function  $f^p$  models the mapping from the cluster center set  $V^p$  to the atlas  $Z$  with the function  $g^p$  in the reverse mapping role. (The reason for employing two functions  $f^p$  and  $g^p$  will become clearer as we proceed.) Since the points in each  $V^p$  and in  $Z$  are in correspondence, the density function in (2) defines a pair of *landmark matching* problems (one for  $f^p$  and one for  $g^p$ ). We now specify regularization priors on  $f^p$  and  $g^p$ .

$$p(f^p) = \frac{1}{Z_{\text{partition}}^f} \exp(-\|Lf^p\|^2), \quad p(g^p) = \frac{1}{Z_{\text{partition}}^g} \exp(-\|Lg^p\|^2). \quad (3)$$

In (3), the operator  $L$  determines the kind of regularization imposed. For example,  $L$  could correspond to a thin-plate spline, a Gaussian radial basis function, etc. Each choice of  $L$  is in turn related to a *kernel* and a *metric* of the deformation from and to  $Z$ . Having specified the density functions of  $X^p$ ,  $V^p$  and the pair  $(f^p, g^p)$ , we may use Bayes' theorem and obtain the posterior.

$$p(Z, \{V^p, f^p, g^p\} | \{X^p\}) = \frac{\prod_{p=1}^P p(X^p | V^p) p(V^p | Z, f^p, g^p) p(f^p) p(g^p)}{p(\{X^p\})}. \quad (4)$$

Some liberties have been taken in deriving (4). We assume that there is a “uniform” prior on  $Z$  which is technically impossible since each point  $z_a$  is defined on  $\mathcal{R}^D$ . It is possible to work around this problem by using a Gaussian density with a very large variance. We now specialize the density  $p(X^p | V^p)$  to a Gaussian mixture model:

$$p(X^p | V^p, \pi^p, \Sigma) = \prod_{i=1}^{N_p} \sum_{a=1}^K \pi_a^p p(x_i^p | v_a^p, \Sigma_a) \quad (5)$$

where

$$p(x_i^p | v_a^p, \Sigma) = \frac{1}{(2\pi)^{\frac{D}{2}} \Sigma_a^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_i^p - v_a^p)^T \Sigma_a^{-1} (x_i^p - v_a^p)\right). \quad (6)$$

where  $\{\Sigma_a, a \in \{1, \dots, K\}\}$  is the set of cluster covariance matrices. For the sake of simplicity and ease of implementation, we immediately specialize to the case where the occupancy probabilities are uniform ( $\pi_a^p = \frac{1}{K}$ ) and the covariance matrices  $\Sigma_a$  are isotropic, diagonal and identical  $[(\Sigma_a = \sigma^2 I_D)]$ . Clearly, these choices can be questioned. The occupancy probability contains valuable information regarding the number of members in a given cluster. And, the covariance matrix gives us valuable information regarding the principal direction (tangent vector) at each cluster center. Since we are already estimating the cluster centers, the deformation between the point-sets and the atlas and finally the atlas, we have elected not to excessively burden the computation

(in terms of speed and quality of solutions) by also estimating the occupancy probability and the covariance matrix of each cluster.

The MAP estimation problem can be finally written as

$$\begin{aligned}
(\hat{Z}, \{\hat{V}^p, \hat{f}^p, \hat{g}^p\}) &= \arg \min_{Z, \{V^p, f^p, g^p\}} -\log p(Z, \{V^p, f^p, g^p\} | \{X^p\}) \\
&= -\sum_{p=1}^P \sum_{i=1}^{N_p} \log \frac{1}{K} \sum_{a=1}^K \frac{1}{(2\pi\sigma^2)^{\frac{D}{2}}} \exp\left(-\frac{1}{2\sigma^2} |x_i^p - v_a^p|^2\right) \\
&\quad + \sum_{p=1}^P \left[ |v_a^p - g^p(z_a)|^2 + |z^p - f^p(v_a^p)|^2 + \lambda \|Lf^p\|^2 + \lambda \|Lg^p\|^2 \right]
\end{aligned} \tag{7}$$

As it stands, minimizing (7) is awkward due to the  $\sum_{p=1}^P \sum_{i=1}^{N_p} \log \sum_{a=1}^K \exp$  form appearing in the first term of the objective. This is a well known problem in Gaussian mixture modeling (Redner and Walker, 1984). The expectation-maximization (EM) algorithm (McLachlan and Basford, 1988) has been quite popular since it avoids a direct minimization of the above cost function. Since the mixture likelihood is non-convex, the EM algorithm is usually executed many times with varying random initial conditions. We instead adopt a deterministic annealing approach (Rose et al., 1990; Yuille et al., 1994; Hofmann and Buhmann, 1997) which (as we shall argue) is especially well suited to the atlas estimation task.

The main difference between the traditional EM algorithm for mixtures and a deterministic annealing algorithm is in the treatment of the isotropic variance parameter  $\sigma^2$ . In deterministic annealing, the variance parameter is externally imposed rather than being estimated from within. A temperature parameter  $T = 2\sigma^2$  as in simulated annealing (or MCMC) is gradually lowered from high values to low values. When the temperature  $T$  is high, the cluster centers congregate around the center of mass of the point-sets. As the temperature is lowered, a series of symmetry-breaking “phase transitions” (Rose et al., 1990) occur during which the cluster centers progressively move

away from the center of mass and toward their more local members.

Consider the following Gaussian mixture likelihood objective function:

$$E_{\text{mix}}(v) = - \sum_{i=1}^{N_x} \log \sum_{a=1}^K \exp \left( -\frac{1}{2\sigma^2} |x_i - v_a|^2 \right). \quad (8)$$

The objective function in (8) is a straightforward mixture objective without any deformation prior.

Now consider

$$E_{\text{cmp}}(v, m^x) = \sum_{i=1}^{N_x} \sum_{a=1}^K m_{ai}^x |x_i - v_a|^2 + T \sum_{i=1}^{N_x} \sum_{a=1}^K m_{ai}^x \log m_{ai}^x \quad (9)$$

The objective function in (9) has a new variable  $m^x$  and the temperature parameter  $T$ . It turns out that

$$\min_{m^x} E_{\text{cmp}}(v, m^x) = E_{\text{mix}}(v) \quad (10)$$

when  $m^x$  satisfies  $m_{ai}^x > 0$  and  $\sum_{a=1}^K m_{ai}^x = 1$  (Hathaway, 1986; Yuille et al., 1994) and when  $T$  is identified with  $2\sigma^2$ . The new variable  $m_{ai}^x$  is a *membership* variable indicating the degree to which each point feature  $x_i$  belongs to cluster center  $v_a$ . The main convenience resulting from using (9) rather than (8) is that (9) does not have the  $\sum \log \sum \exp$  form in it. Also, the term  $\sum_{ai} m_{ai}^x \log m_{ai}^x$  is an *entropy* barrier function with  $T$  being the temperature.

We now specify the overall cost function for atlas estimation. We convert the variance parameter  $\sigma^2$  into a deterministic annealing parameter and rewrite the cost function using the new membership variables  $\{M^p, p \in \{1, \dots, P\}\}$ .

$$E(Z, \{V^p, M^p, f^p, g^p\}) = \sum_{p=1}^P E^p(Z, V^p, M^p, f^p, g^p), \quad (11)$$

where

$$\begin{aligned}
E^P(Z, V^p, M^p, f^p, g^p) &= \sum_{i=1}^{N_p} \sum_{a=1}^K m_{ai}^p |x_i^p - v_a^p|^2 \\
&+ \sum_{a=1}^K |z_a - f^p(v_a^p)|^2 + \sum_{a=1}^K |g^p(z_a) - v_a^p|^2 \\
&+ \lambda \|Lf^p\|^2 + \lambda \|Lg^p\|^2 + T \sum_{i=1}^{N_p} \sum_{a=1}^K m_{ai}^p \log m_{ai}^p
\end{aligned} \tag{12}$$

where the membership matrix entries  $m_{ai}^p \in [0, 1]$  and satisfy the constraint  $\sum_{a=1}^K m_{ai}^p = 1$ . We briefly go over the individual terms in this energy function and explain some of the new variables. The first term  $\sum_{i=1}^{N_p} \sum_{a=1}^K m_{ai}^p |x_i^p - v_a^p|^2$  measures the average distance between the sample points and the cluster centers. Minimization of this term will essentially force all the cluster centers to be as close as possible to the original sample points, thus maintaining the original shape. The second and third terms  $\sum_{p=1}^P \sum_{a=1}^K |z_a - f^p(v_a^p)|^2$ ,  $\sum_{p=1}^P \sum_{a=1}^K |g^p(z_a) - v_a^p|^2$ , model the deformations (both the forward  $f^p$ , which warps the cluster center sets  $V^p$  to  $Z$ , and the reverse  $g^p$ , which does the opposite) to align the cluster centers with the mean. We think it is possible to force the deformation to be consistent by requiring  $g^p$  to be  $(f^p)^{-1}$  as in (Christensen, 1999) but have not done so here. The fourth term  $\|Lf^p\|^2 + \|Lg^p\|^2$  measures the amount of distortion or bending introduced to the space by the non-rigid warps  $f^p$  and  $g^p$ . By penalizing this bending measure, the algorithm will effectively place the mean shape points to be somewhat similar to all shape samples, since otherwise it will require a larger amount of bending to warp any of the shape samples onto the mean. Since all cluster center sets share a common mean point set  $Z$ , consistency between the cluster centers is also achieved. The fifth term  $T \sum_{i=1}^{N_p} \sum_{a=1}^K m_{ai}^p \log m_{ai}^p$  arises from deterministic annealing (Chui and Rangarajan, 2000) as explained above. The parameter  $T$ , termed the temperature, is used to control the fuzziness of the clustering membership matrices  $\{M^p\}$ : higher the temperature, greater the fuzziness. The  $x \log x$  form of the barrier function effectively leads to the formation of

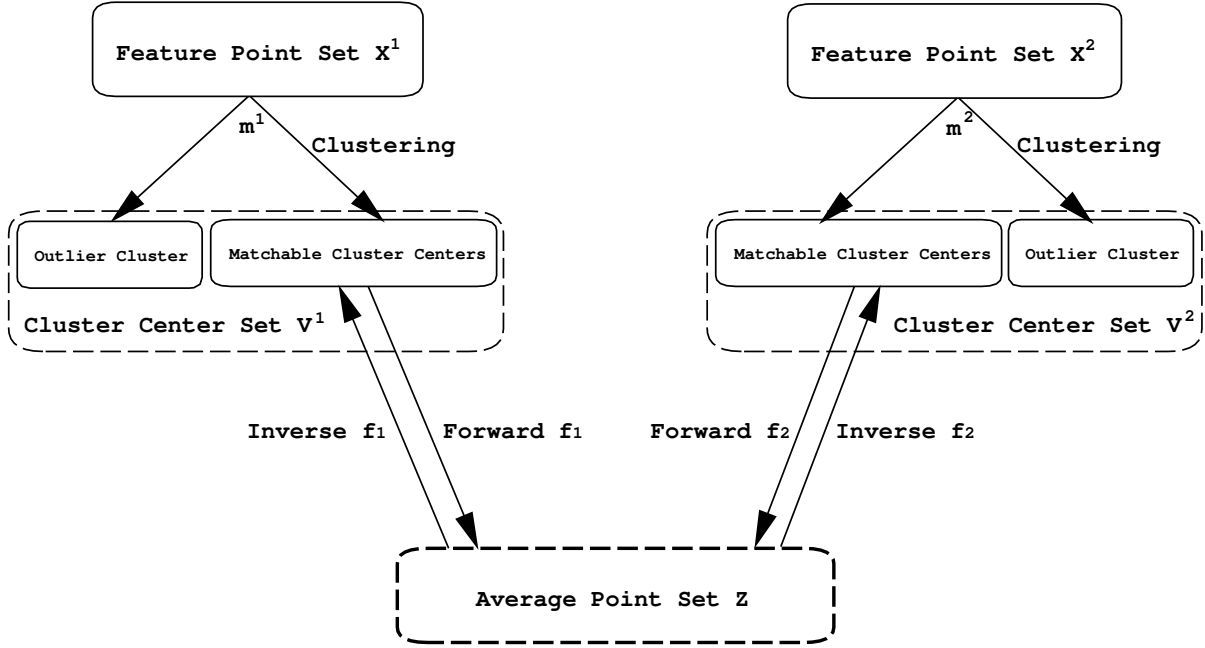


Figure 2: The super clustering-matching algorithm. Each original point-set ( $X^1$  and  $X^2$ ) is clustered down to a set of cluster centers ( $X^1 \rightarrow V^1$  and  $X^2 \rightarrow V^2$ ). Each cluster set has an outlier cluster ( $v_{K+1}^1$  in  $V^1$ ,  $v_{K+1}^2$  in  $V^2$ ) to account for the possible spurious points in each point-set. The rest of the cluster centers ( $\{v_a^1, a = 1, 2, \dots, K\}$  and  $\{v_a^2, a = 1, 2, \dots, K\}$ ) are matched to the average point-set  $Z$ .

Gaussian clusters. The temperature  $T$  can be interpreted as a common gaussian mixture variance parameter (Chui and Rangarajan, 2000; Yuille et al., 1994). However, it is manually controlled as opposed to being automatically adjusted. The fuzziness of the membership matrices is gradually reduced by slowly annealing  $T$  in a pre-defined linear scheme. In the experiments, we have empirically observed that annealing the regularization parameter  $\lambda$  leads to improved solutions. This allows us to focus on first estimating the rigid transformation parameters and then the non-rigid transformations.

The whole setup is demonstrated in Figure 2. For simplicity, it is illustrated with only two sample point-sets ( $P = 2$ ). Extending the setup to incorporate multiple point-sets is obvious

since there is no bias toward either point-set in Figure 2. The figure also depicts a possible outlier estimation extension which has not been implemented in this work.

Having specified this objective function with three groups of unknown variables, we conduct a grouped coordinate descent procedure to achieve the energy minimization. Essentially, an alternating algorithm is used to cycle between the updates of each of the three groups of variables.

The update steps can be summarized as the following:

(i) Update the cluster center sets: We first estimate the membership matrices  $\{M^p\}$

$$m_{ai}^p = \frac{q_{ai}^p}{\sum_{a=1}^K q_{ai}^p}, \forall a, i \quad (13)$$

where

$$q_{ai}^p = e^{-\frac{|x_i^p - v_a^p|^2}{T}}, \forall a \in \{1, \dots, K\}, \text{ and } \forall i \in \{1, \dots, N_p\}. \quad (14)$$

Then we compute the cluster centers,

$$v_a^p = \frac{\sum_{i=1}^{N_p} m_{ai}^p x_i^p + g^p(z_a)}{\sum_{i=1}^{N_p} m_{ai} + 1} \forall a \in \{1, \dots, K\}, \text{ and } \forall p \in \{1, \dots, P\}. \quad (15)$$

Note that the cluster centers are determined both by the original sample points  $\{x_i^p\}$  as well as the common mean  $Z$ . As mentioned before, the cluster center sets are linked to each other through the mean set  $Z$  which enables them to achieve consistency. The update equation for  $\{M^p\}$  is *exact* whereas the update equation for  $\{V^p\}$  is *approximate*. We have neglected the term  $\sum_{a=1}^K |z_a - f^p(v_a^p)|^2$  in computing  $\{V^p\}$ . This is because we believe that the “inverse” term  $\sum_{a=1}^K |g^p(z_a) - v_a^p|^2$  is a sufficient constraint for estimation. A closed-form solution for  $\{V^p\}$  cannot be obtained if the term  $\sum_{a=1}^K |z_a - f^p(v_a^p)|^2$  is taken into account.

(ii) Update the atlas (mean shape) point-set:

$$z_a = \frac{1}{P} \sum_{p=1}^P f^p(v_a^p) \forall a \in \{1, \dots, K\}. \quad (16)$$

This update equation for  $Z$  is approximate since we have neglected the term  $\sum_{a=1}^K |g^p(z_a) - v_a^p|^2$  in computing the mean  $Z$ . This is because we believe that the forward term  $\sum_{a=1}^K |z_a - f^p(v_a^p)|^2$  is a sufficient constraint for estimation. A closed-form solution for  $Z$  cannot be obtained if the term  $\sum_{a=1}^K |g^p(z_a) - v_a^p|^2$  is taken into account. To get an initial estimate of  $Z$ , we initialize the algorithm by setting  $\{f^p, g^p\}$  to be identity transformations.

(iii) **Landmark matching:** Update the non-rigid transformations:

$$f^p = \arg \min_{f^p} \left( \sum_{a=1}^K |z_a - f^p(v_a^p)|^2 + \lambda \|L f^p\|^2 \right) \forall a \in \{1, \dots, K\} \text{ and } \forall p \in \{1, \dots, P\} \quad (17)$$

with a similar update for the “inverse” transformation  $g^p$ . Since this is equivalent to a landmark-based approach, we can solve for  $f^p$  and  $g^p$  in closed form. Our approach can therefore be viewed as an automated landmarking process. If diffeomorphisms are required for  $f^p$ , a single closed form solution is usually not possible. Instead, we would have to introduce a “time” parameter and construct an ODE to take each  $V^p$  onto  $Z$  as in (Camion and Younes, 2001; Joshi and Miller, 2000). The pseudo-code of the algorithm is as follows.

**The Super Clustering-Matching Algorithm:**

Initialize parameters  $T = T_0$  and  $\lambda$ .

Initialize all membership matrices  $\{M^p\}$  (e.g., use uniform matrix).

Initialize all transformation matrices  $\{f_p\}$  (e.g., use identity transformation).

Initialize the average point-set  $Z$  (e.g., use the center of mass of all point-sets).

**Begin A: Deterministic Annealing.****Begin B: Alternating Update.**

Step 1: Update cluster centers  $\{V^p\}$  based on current  $\{M^p\}$ ,  $\{g^p\}$  and  $Z$ .

Step 2: Update average point-set  $Z$  based on cluster centers  $\{V^p\}$  and  $\{f_p\}$ .

Step 3a: Update transformations  $\{f^p, g^p\}$  based on current  $V^p$  and  $Z$ .

Step 3b: Update cluster memberships  $\{M^p\}$  based on current  $\{V^p\}$ .

**End B**

Decrease  $T \leftarrow T \cdot r$  until  $T_{final}$  is reached.

**End A**

Instead of having the flavor of essentially a gradient descent algorithm, which can be plagued by many local minima, our algorithm utilizes deterministic annealing (DA) with its proven benefit of robustness especially in clustering problems (Yuille et al., 1994). The working mechanism of deterministic annealing has been discussed extensively in many other papers (Yuille et al., 1994; Chui and Rangarajan, 2000). We would like to point out here that utilizing annealing enables us to compute the cluster centers in a coarse to fine fashion, and hence also the computation of the mean. In all experiments, the annealing is also applied to the regularization parameter  $\lambda$  of the non-rigid transformation. The annealing technique is the key to solving this difficult deformable shape averaging problem.

## 4 Results

After discussing the details of our algorithm and its implementation, we begin with a simple but demonstrative example of the joint clustering and matching algorithm for 2D atlas estimation. After presenting this example, we describe a 3D implementations on a real hippocampal dataset.

The structure that we are interested in is the corpus callosum as it appears in MR brain images. The corpus callosum structure normally has an elongated shape. Constructing an atlas for the corpus callosum and subsequently analyzing the individual shape variation from “normal” anatomy has been regarded as potentially valuable for the study of brain diseases such as agenesis of the corpus callosum (ACC), and fetal alcohol syndrome (FAS).

We manually extracted points on the outer contour of the corpus callosum from nine normal subjects (as shown in Figure 3). The resulting atlas (mean point-set) is shown in Figure 6. The cluster center sets, which are our newly refined shape samples, are shown in Figure 4. The correspondences and the transformations between the cluster centers and the mean are also shown in Figure 5. As we described earlier, all these results are computed simultaneously and automatically. This example clearly demonstrates that our joint clustering and matching algorithm can simultaneously align multiple shape sample point-sets and compute a meaningful atlas (mean point-set) while automatically refining the sample points’ location to achieve better consistency. To anecdotally test the influence of the free parameter  $K$  (number of clusters), we estimated the atlas with  $K = 20, 40, \text{ and } 60$ . The results are shown in Figure 7. The atlas estimate improves as  $K$  is increased. Unfortunately, all that this experiment shows is an underfitting effect caused by too few cluster centers. The more subtle over-fitting effect of having too many cluster centers has not yet been studied.

Next, we present results on 3D hippocampal point-sets. Ten 3D point-sets were extracted from epilepsy patients with left anterior temporal lobe foci identified with EEG. An interactive seg-

mentation tool was used to segment the hippocampus in the 3D anatomical brain MRI of the ten subjects. The point-sets differ in shape. The number of points in each point-set varies from 350-450. Nine of the ten hippocampal point-sets are shown in Figure 8. In this experiment, we modified the annealing scheme somewhat from the corpus callosum example. At high temperature, we kept the thin-plate spline regularization parameter fixed at a very high value. As the temperature was decreased, we did not change the value of  $\lambda$ . Only after the rigid (affine) parameters were approximately computed did we lower  $\lambda$ . This allowed for the non-rigid transformations to be gradually introduced after prior computation of the rigid transformation. In Figure 10, the estimated atlas is shown along with a superimposition on all of the original datasets. We used  $K = 150$  cluster centers for the atlas.

Once the atlas has been estimated, we may warp the original point-sets into the atlas space. Since there is a one-to-one correspondence between the cluster centers for each point-set and the atlas, we can estimate a thin-plate spline mapping *using the same limiting value of  $\lambda$*  as per the atlas estimation. In Figure 11 we show the scatter plot of all ten point-sets after warping into the atlas space. For the sake of comparison, in Figure 12 we also show the scatter plot of all ten point-sets after rigid (affine) warping into the atlas space. An examination of the two scatter plots clearly shows the efficacy of the non-rigid warping. The overall shape of the atlas is only visible in the former but not in the latter, which clearly demonstrates the need for using non-rigid deformations between the individual point-sets and the emerging atlas.

## 5 Conclusion

One of the features of the atlas estimation method presented here is the low overall computational complexity. Since, we are dealing with an iterative algorithm, it only makes sense to talk about the per-iteration complexity. The clustering step is  $O(NKP)$  and the thin-plate spline fitting step

is  $O(K^3P)$  and consequently, if  $K$  is  $O(N)$ , then the thin-plate spline fitting step is the most expensive. As  $K$  increases, direct matrix inverses become infeasible and we have to resort to numerical approximations of the inverse.

There are two important parameters in our cost function which need to be estimated; the regularization parameter  $\lambda$  and the number of clusters  $K$ . The annealing schedule, initial temperature and final temperature are not as critical since they can be readily set as in any other scale-space strategy. We wish to point out an interesting relationship between these two unknown parameters and a connection to entropy-based approaches. If the number of cluster centers is increased, then it is possible to better approximate the shape of a point-set. However, if the number of clusters exceeds a limit, its effectiveness decreases since the inherent feature “jitter” increases (resulting in increased cluster center variance). Increased cluster center variance negatively impacts on the deformation. If the regularization parameter is too high, then the deformation cannot compensate for the cluster center “jitter” resulting in a bad spline fit. If the regularization parameter is too low, then the spline happily follows the cluster center jitter and overfits the deformation. Obviously there is going to be a tradeoff regime for both parameters.

A reasonable criterion for atlas estimation is the minimization of the overlap entropy of the original point-sets. The downside is that the entropy can be trivially sent to zero by collapsing all the original point-sets into the same point. However, in this case the warping between the individual point-set cluster centers and the atlas becomes very large. Clearly there is a tradeoff between the entropy (of the overlaid point-sets) and the degree of deformation. Regardless of what criterion is chosen, the number of cluster centers and the regularization parameter must be chosen to balance i) clustering the original point-sets, ii) minimizing the overlap entropy and iii) penalizing both too large and too small a deformation. Once the atlas is estimated (with minimal reliance on free parameters), the next step is to use the atlas on a test set. We deemed this issue to

be beyond the scope of this paper but it is an interesting area for future research.

We have presented anecdotal evidence demonstrating that it is indeed possible to simultaneously estimate correspondences, deformations and an atlas from unlabeled point-sets. And we have argued, we hope convincingly, that correspondence, sample consistency, deformations and atlases belong together and should be simultaneously estimated in order to avoid bias, landmark inconsistency and a host of other problems. Obviously a good validation effort is required in order to show the effectiveness of our approach.

## Acknowledgments

This work is supported by NSF IIS 0196457. An abbreviated version of this paper first appeared in (Chui and Rangarajan, 2001).

## References

- Bookstein, F. L. (1989). Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. Patt. Anal. Mach. Intell.*, 11(6):567–585.
- Camion, V. and Younes, L. (2001). Geodesic interpolating splines. In *Energy Minimization Methods for Computer Vision and Pattern Recognition*, pages 513–527. Springer, New York.
- Christensen, G. (1999). Consistent linear-elastic transformations for image matching. In *Proceedings of Information Processing in Medical Imaging—IPMI 99*, pages 224–237. Springer-Verlag.
- Chui, H. and Rangarajan, A. (2000). A new feature registration framework using mixture models.

- In *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis–MMBIA 2000*.  
IEEE Press.
- Chui, H. and Rangarajan, A. (2001). Learning an atlas from unlabeled point-sets. In *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA)*, pages 179–186.  
IEEE Press.
- Chui, H., Win, L., Duncan, J., Schultz, R., and Rangarajan, A. (2003). A unified non-rigid feature registration method for brain mapping. *Medical Image Analysis*, 7:112–130.
- Cootes, T., Taylor, C., Cooper, D., and Graham, J. (1995). Active shape models: Their training and application. *Computer Vision and Image Understanding*, 61(1):38–59.
- Davatzikos, C. (1997). Spatial transformation and registration of brain images using elastically deformable models. *Computer Vision and Image Understanding: Special Issue on Medical Imaging*, 6(2):207–222.
- Davies, R. H., Twining, C., Cootes, T. F., and Taylor, C. J. (2002). A minimum description length approach to statistical shape modelling. *IEEE Trans. on Medical Imaging*, 21:525–537.
- Duta, N., Jain, A. K., and Dubuisson-Jolly, M. (1999). Learning 2D shape models. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 8–14.
- Duta, N., Jain, A. K., and Dubuisson-Jolly, M.-P. (2001). Automatic construction of 2D shape models. *IEEE Trans. Patt. Anal. Mach. Intell.*, 23:433–446.
- Golland, P., Grimson, E., and Kikinis, R. (1999). Statistical shape analysis using fixed topology skeletons: Corpus callosum study. In *Information Processing in Medical Imaging (IPMI)*, volume 1613, pages 382–388. Springer.

- Guimond, A., Meunier, J., and Thirion, J.-P. (2000). Average brain models: A convergence study. *Computer Vision and Image Understanding*, 77(2):192–210.
- Hathaway, R. (1986). Another interpretation of the EM algorithm for mixture distributions. *Statistics and Probability Letters*, 4:53–56.
- Hill, A., Taylor, C. J., and Brett, A. D. (2000). A framework for automatic landmark identification using a new method of nonrigid correspondence. *IEEE Trans. Patt. Anal. Mach. Intell.*, 22(3):241–251.
- Hofmann, T. and Buhmann, J. M. (1997). Pairwise data clustering by deterministic annealing. *IEEE Trans. Patt. Anal. Mach. Intell.*, 19(1):1–14.
- Joshi, S. and Miller, M. (2000). Landmark matching via large deformation diffeomorphisms. *IEEE Trans. Image Processing*, 9:1357–1370.
- Kendall, D. G. (1984). Shape-manifolds, Procrustean metrics and complex projective spaces. *Bulletin of the London Mathematical Society*, 16:81–121.
- McLachlan, G. J. and Basford, K. E. (1988). *Mixture models: inference and applications to clustering*. Marcel Dekker, New York.
- Redner, R. A. and Walker, H. F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, 26(2):195–239.
- Rohlfing, T., Brandt, R., Maurer Jr, C. R., and Menzel, R. (2001). Bee brains, B-splines and computational democracy: Generating an average shape atlas. In *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA)*, pages 187–194. IEEE Press.

- Rose, K., Gurewitz, E., and Fox, G. (1990). Statistical mechanics and phase transitions in clustering. *Physical Review Letters*, 65(8):945–948.
- Rueckert, D., Frangi, A., and Schnabel, J. (2001). Automatic construction of 3D statistical deformation models using non-rigid registration. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pages 77–84.
- Sebastian, T. B., Crisco, J. J., Klein, P. N., and Kimia, B. B. (2000). Construction of 2D curve atlases. In *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis—MMBIA 2000*, pages 70–77.
- Small, C. (1996). *The statistical theory of shape*. Springer, New York, NY.
- Tagare, H. (1999). Shape-based nonrigid correspondence with application to heart motion analysis. *IEEE Transactions on Medical Imaging*, 18(7):570–579.
- Thompson, P., MacDonald, D., Mega, M., Holmes, C., Evans, C., and Toga, A. (1997). Detection and mapping of abnormal brain structure with a probabilistic atlas of cortical surfaces. *Journal of Computer Assisted Tomography*, 21(4):567–581.
- Wang, Y. and Staib, L. H. (2000). Boundary finding with prior shape and smoothness models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7):738–743.
- Yuille, A. L., Stolorz, P., and Utans, J. (1994). Statistical physics, mixtures of distributions, and the EM algorithm. *Neural Computation*, 6(2):334–340.

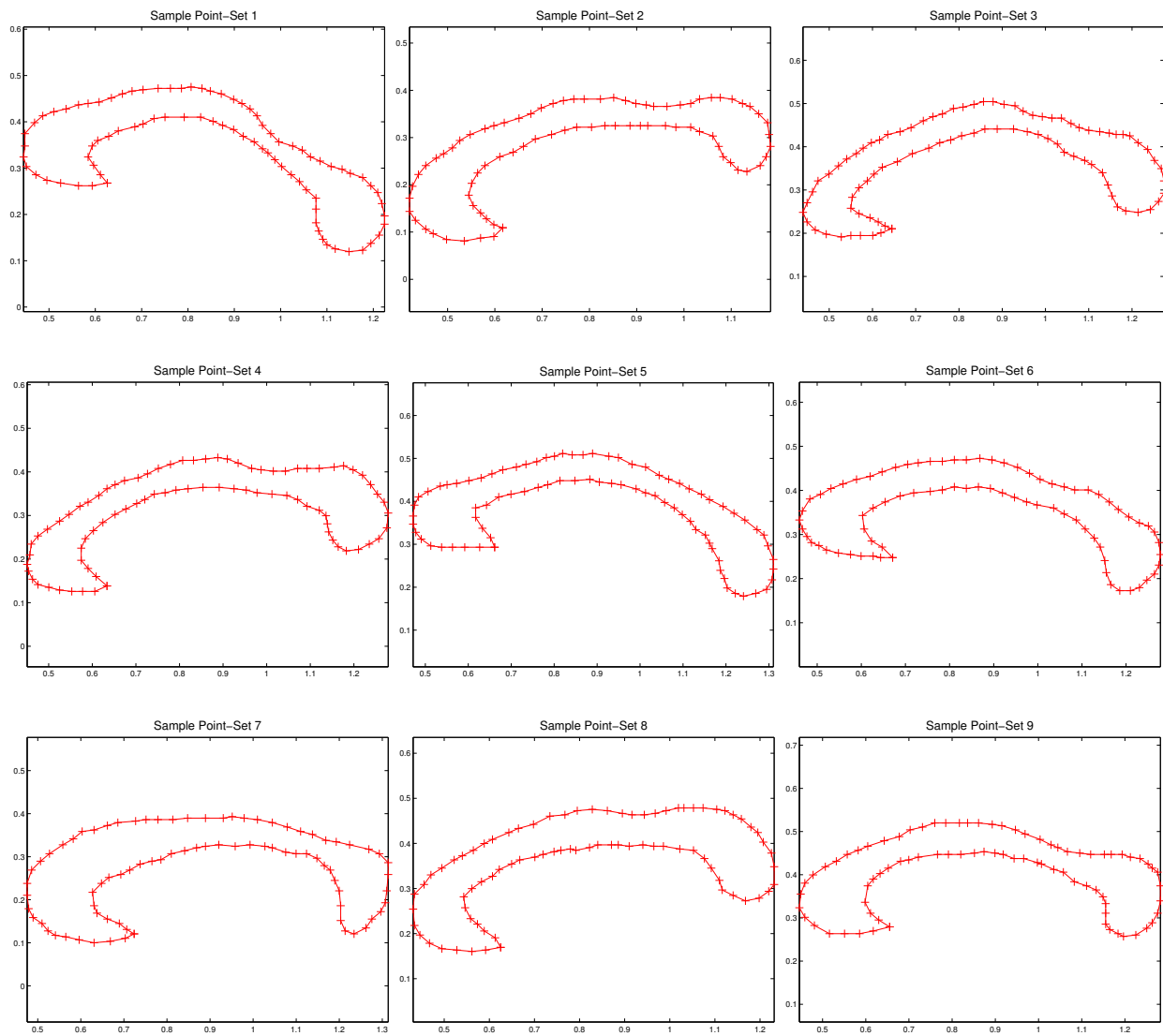


Figure 3: Nine corpus callosum 2D shapes are shown. Care was taken to approximately extract 2D contours at the same orientation.

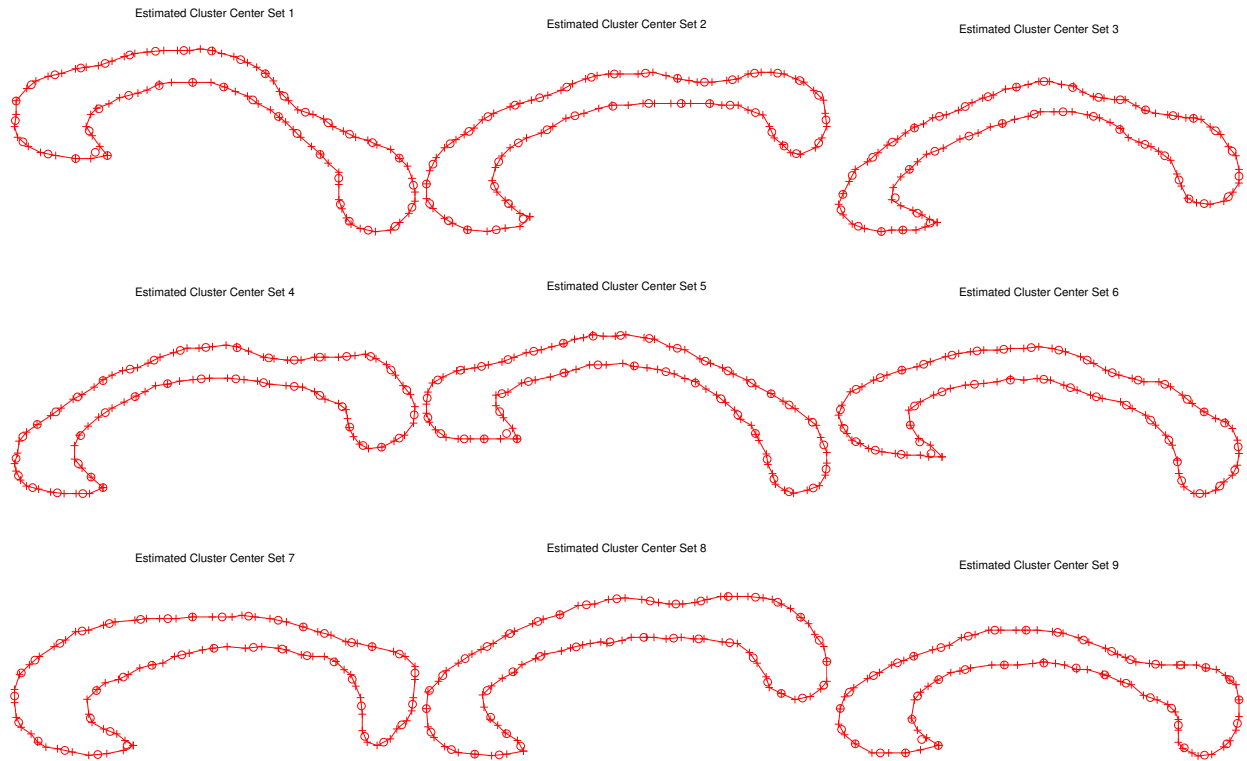


Figure 4: The estimated cluster centers for all nine corpus callosum shapes are shown. Note that the cluster centers for the most part lie on the shape outline. The cluster centers are in correspondence because they share a common index.

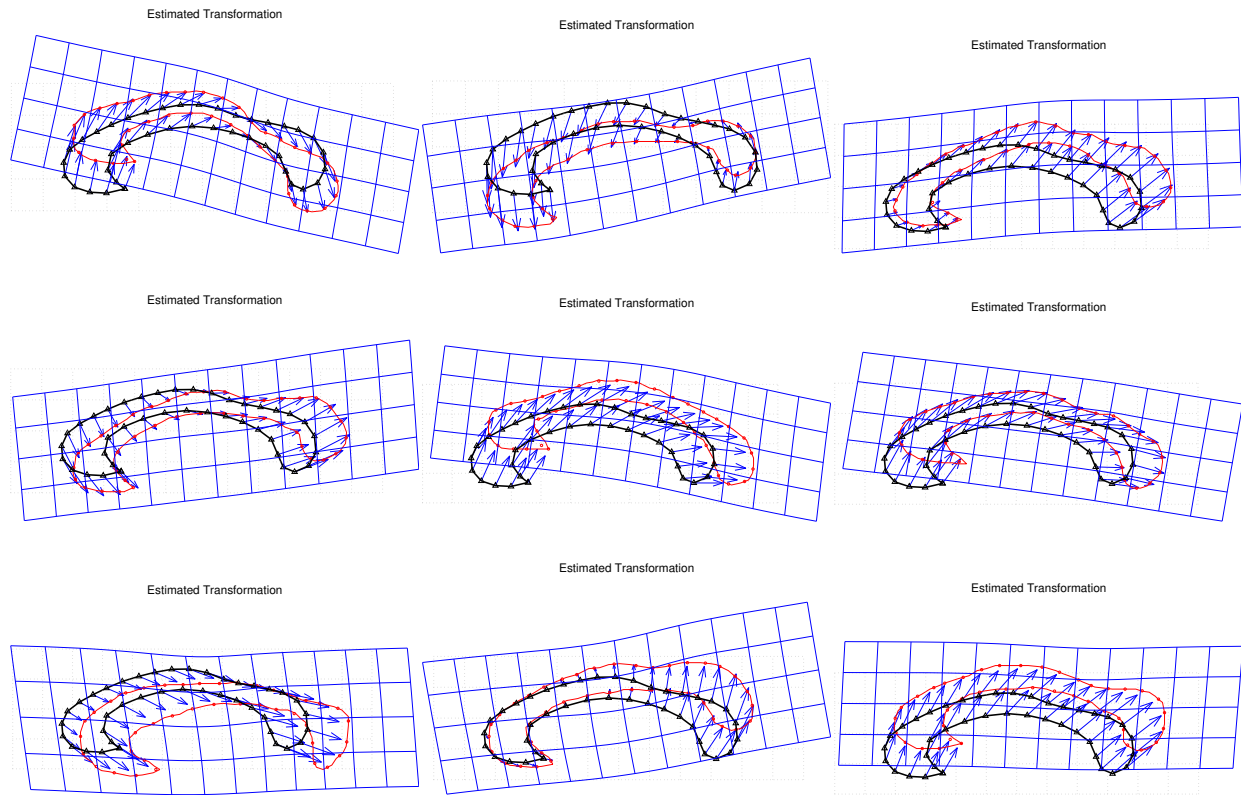


Figure 5: The deformations of each set of clusters onto the atlas are shown.

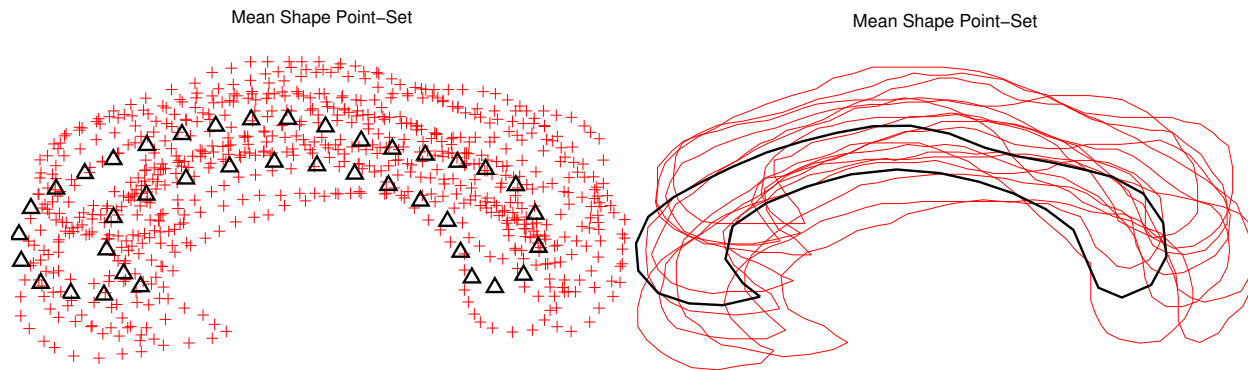


Figure 6: Left: The estimated atlas is shown superimposed over all the point-sets. Right: An atlas contour is traced and shown superimposed over all the original contours.

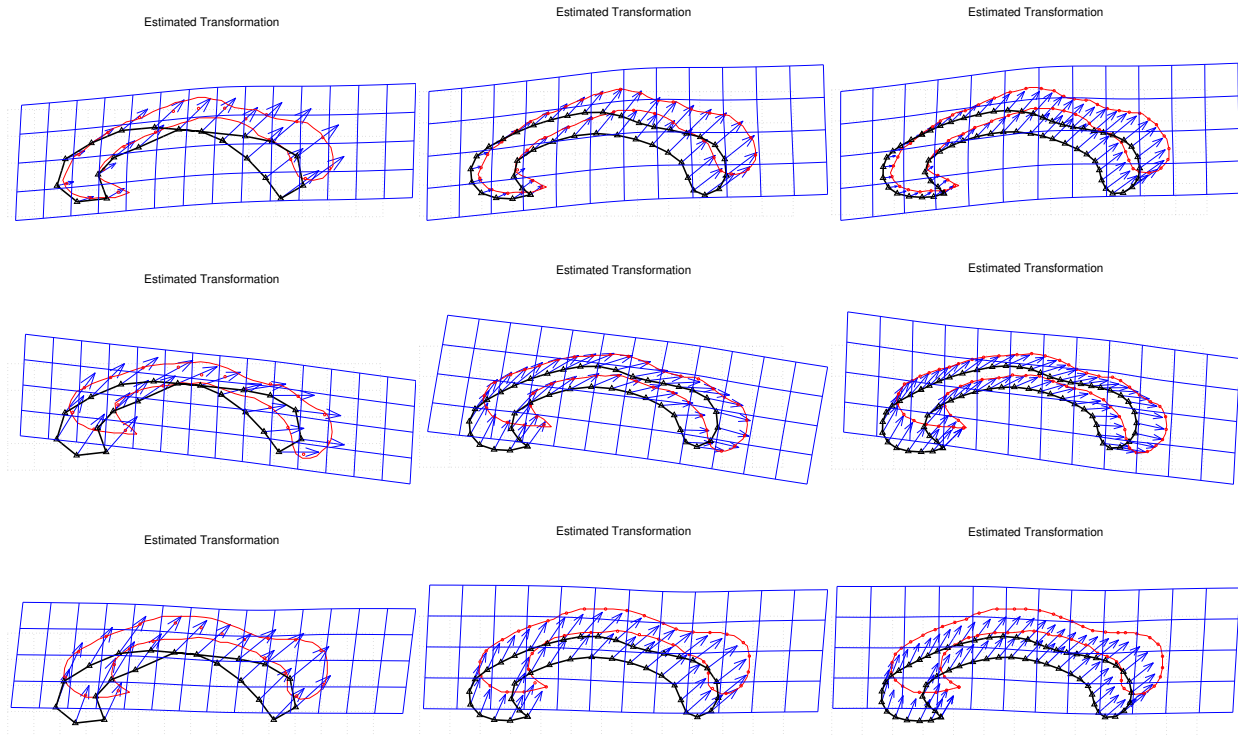


Figure 7: **Left to Right:** Variation of the number of cluster centers. The estimated atlas is shown for a different number  $K$  of cluster centers. **Left:**  $K = 20$ , **Middle:**  $K = 40$ , **Right:**  $K = 60$ . Note the improvement in atlas estimation as the number of cluster centers is increased from left to right. Three different corpus callosum training examples are shown from top to bottom.

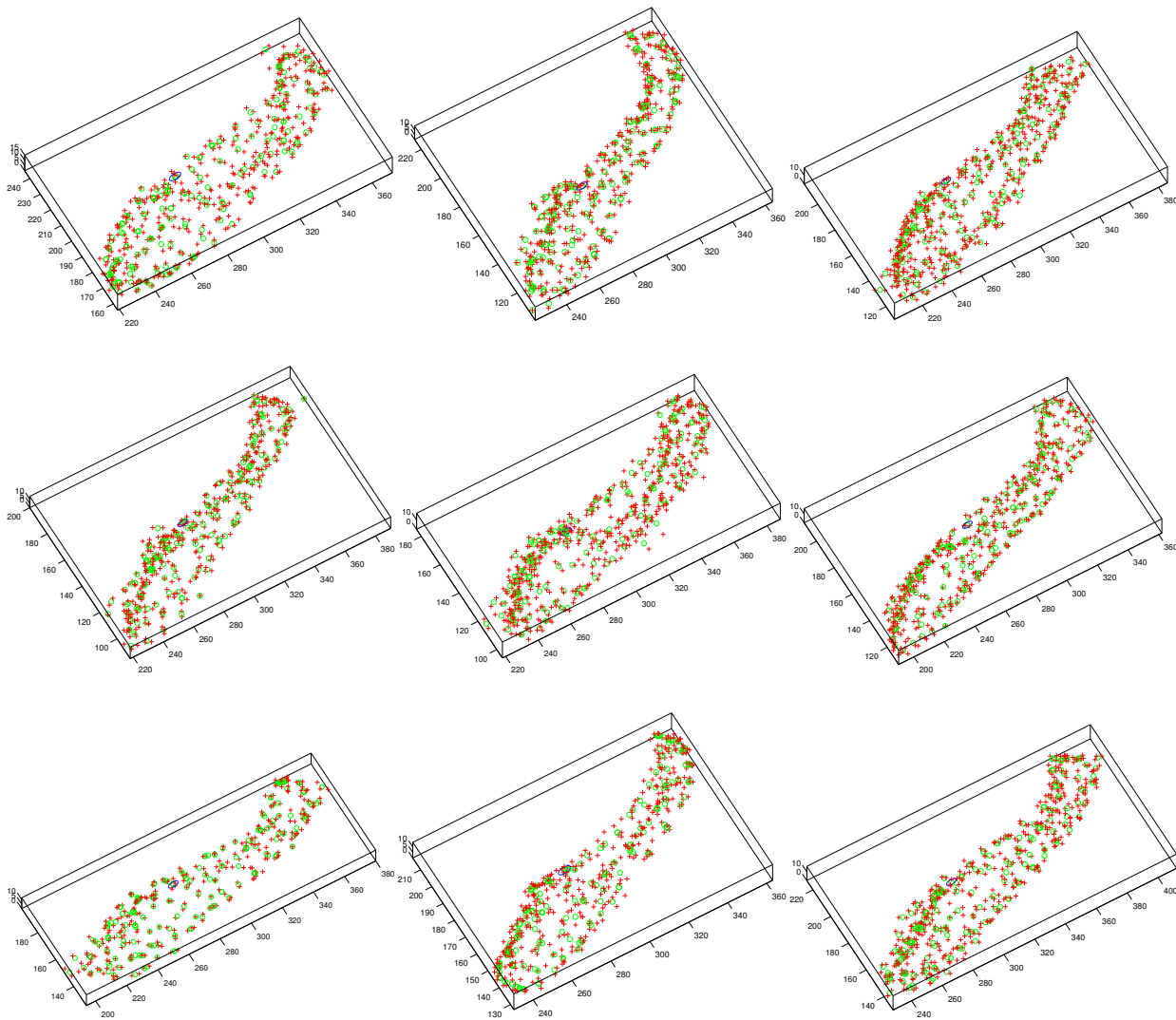


Figure 8: **3D Hippocampal point-sets:** Nine (of the ten) hippocampal point-sets are shown. The point-sets correspond to preoperative data from nine subjects with epileptic foci in the left anterior temporal lobe (LATL) as verified by seizure elimination after left anterior lobectomy. The point-sets are shown in roughly the same orientation. The cluster centers (circles) for each point-set are shown superimposed on the original data.

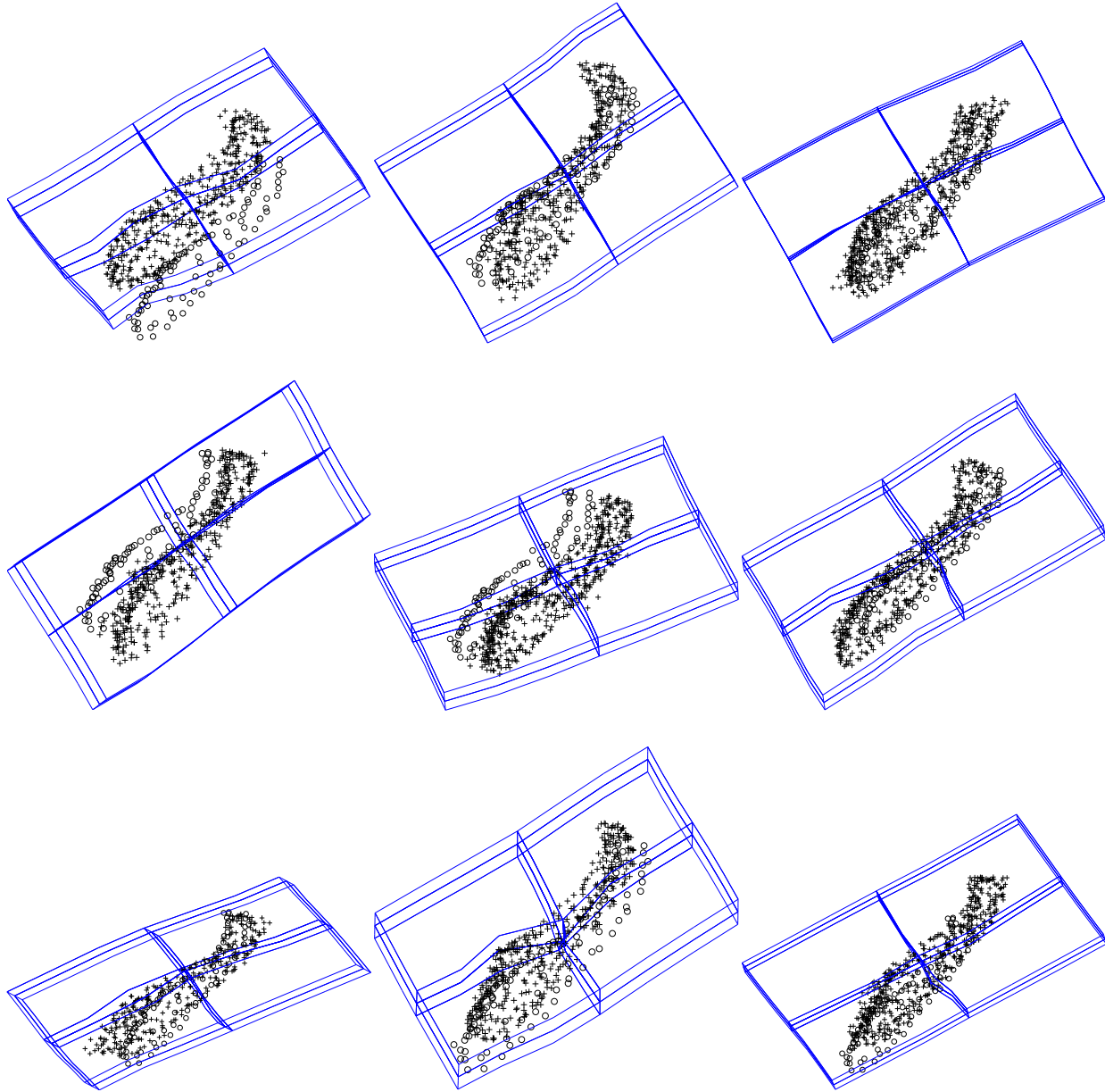


Figure 9: **3D Hippocampal point-sets:** Nine (of the ten) hippocampal point-sets are shown. The point-sets correspond to preoperative data from nine subjects with epileptic foci in the left anterior temporal lobe. The atlas (circles) is shown superimposed on the data along with the deformation that takes the atlas onto each point-set.

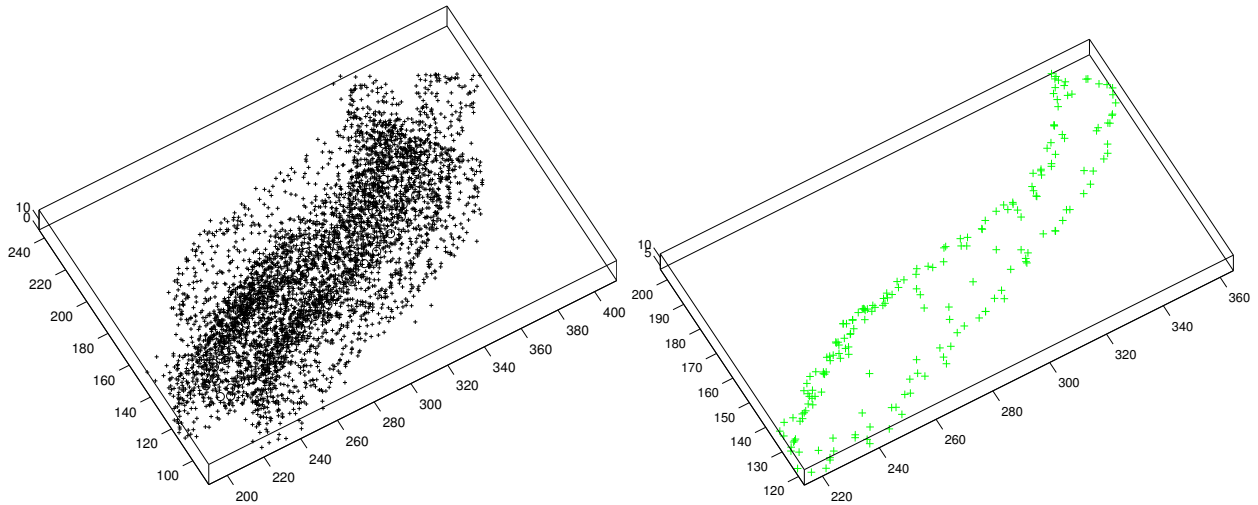


Figure 10: **Ten hippocampal point-sets.** **Left:** Scatter plot of ten 3D hippocampal point-sets. **Right:** The estimated atlas is shown at roughly the same orientation as the point-sets.

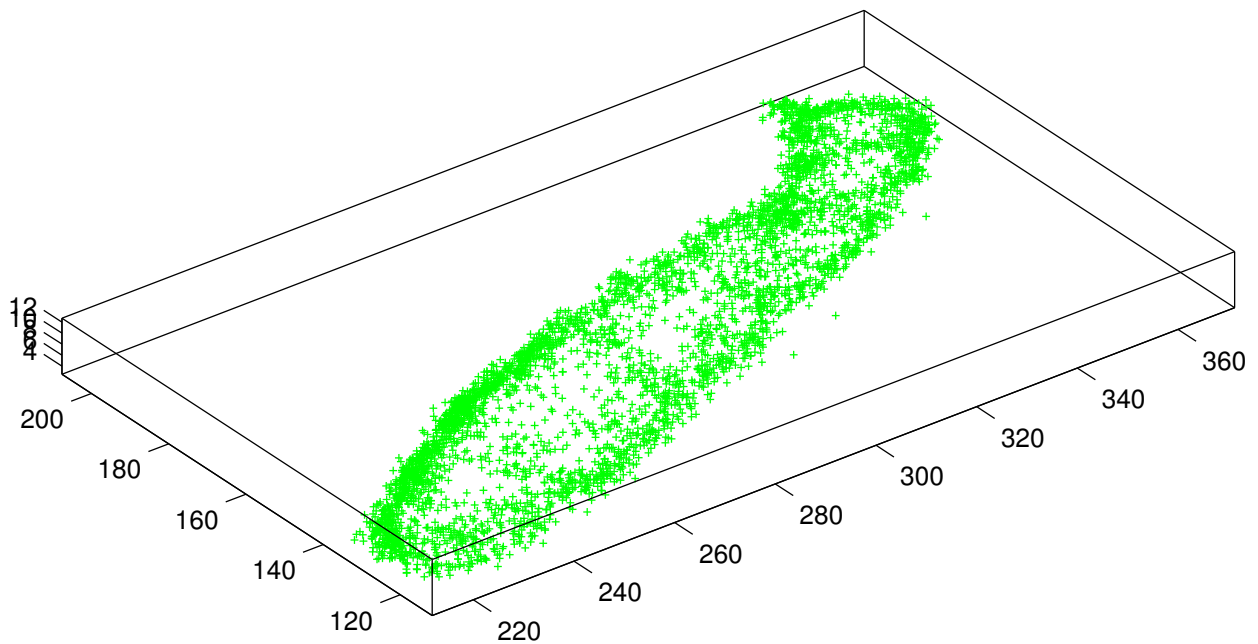


Figure 11: **Warping the original point-sets into the atlas space:** All the hippocampal point-sets were warped onto the atlas space. The scatter plot obviously shows the reduction in entropy.

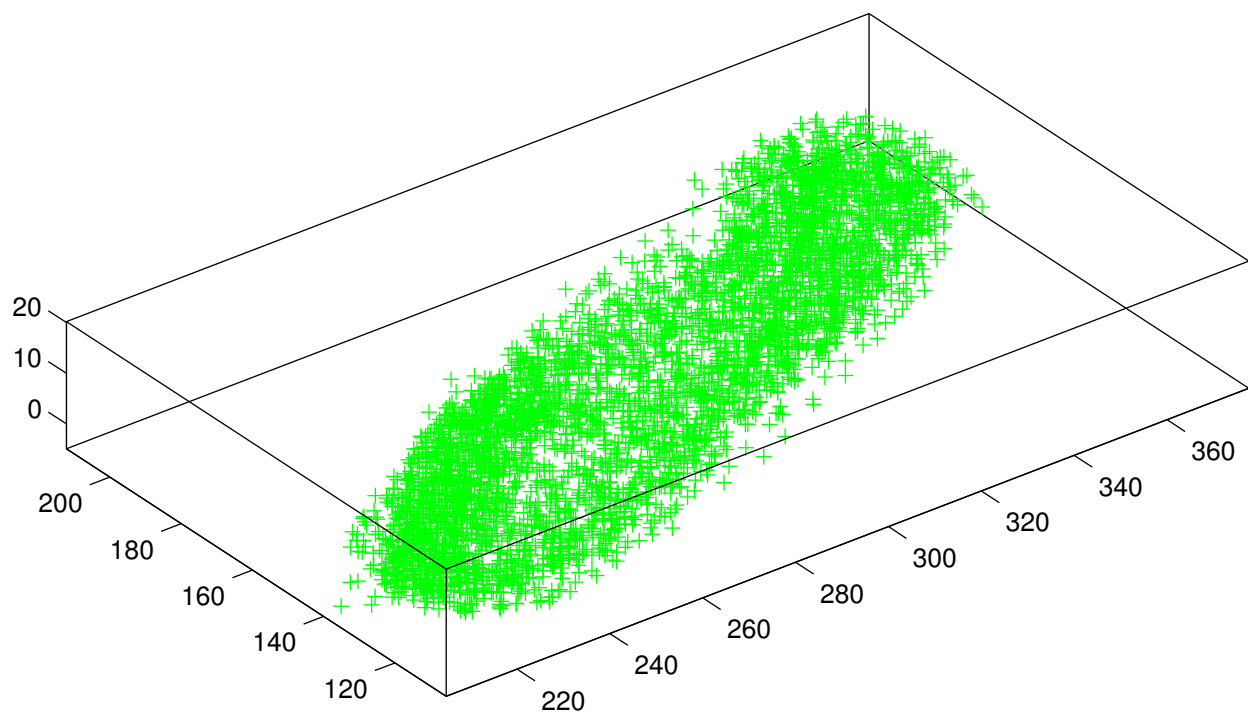


Figure 12: **Affine mapping of the original point-sets into the atlas space:** All the hippocampal point-sets were affine mapped onto the atlas space. The scatter plot shows the lack of shape definition by mere affine mapping.