Training Models of Anatomic Shape Variability
Medical Physics MS #07-339

Derek Merck, Gregg Tracton, Rohit Saboo, Joshua Levy,
Edward Chaney, Stephen Pizer, Sarang Joshi

Medical Image Display & Analysis Group, University of North Carolina at Chapel Hill

Abstract

Learning probability distributions of the shape of anatomic structures requires fitting shape representations to human expert segmentations from training sets of medical images. The quality of statistical segmentation and registration methods is directly related to the quality of this initial shape fitting, yet the subject is largely overlooked or described in an *ad hoc* way. This paper presents a set of general principles to guide such training. Our novel method is to jointly estimate the shape distribution and the best geometric model for any given image by iteratively relaxing purely geometric constraints in favor of the converging shape probabilities as the fitted object converges to its target segmentation. These geometric constraints are carefully written to obtain non-self-interpenetrating shapes and to impose the model-to-model correspondences required for useful statistical analysis. The paper closes with example applications of the method to synthetic and real patient CT image sets, including same patient male pelvis and head and neck images, and cross patient kidney and brain images.

1. Introduction

Our target application, adaptive and image-guided radiotherapy (ART [1], IGRT [2]), requires a framework for accurately mapping anatomical objects from inter-fractional images into the same coordinate system as the planning or other reference image. Current research activity is aimed at developing practical and reliable methods for space –filling non-rigid mapping [3,4,5,6,7,8] to overcome the shortcomings of current clinical methods based largely on rigid registration. This work focuses on the framework required for probabilistic image segmentation as a means for automatically computing non-rigid mappings. Probabilistic segmentation can also provide a basis for image guided surgery (e.g., [9,10]) and diagnosis [11].

Probabilistic segmentation is based on understanding the shape variability of the anatomic structures found in medical images. Robust probability density estimates of shape have been shown to be effective for object-based methods of probabilistic segmentation [9,12,13] for two reasons: 1) their relatively low dimensionality allows us efficiently to compute optimal solutions, and 2) the optimal solutions yield anatomically credible objects. Geometrically legal space-filling non-rigid mappings can be derived from differential equations (e.g., optical flow [14,15], elastic flow [16], visco-elastic flow [6,17,18] or finite-element models [19,20]) and then reduced to lower dimension for statistical analysis by means of control points or decomposition via basis functions. However, these methods cannot be restricted to credible shapes because one must trade off spatial resolution with the fact that the extremely high dimensions prevents learning the feature space of credible shapes from the limited number of training cases available [21]. Furthermore, while these methods have been applied to images with similar topology such as brain or inter-fractional or 4D scans, they have not been shown to be applicable to cross-patient images. This paper therefore restricts its attention to models of one or more anatomic objects, e.g., of their surfaces or skeletons.

The method presented here is based on statistical deformable shape models (SDSMs), which are object-based parametric shape model that characterizes shape changes relative to a "typical" instance. The typical shape usually is taken to be the mean shape, and relative changes are encoded as a limited number of important eigenmodes of shape variability. Using an SDSM, any particular shape in the space defined by the training set can be completely and uniquely identified by a few coefficients to within some small truncation error. Once trained, SDSMs can be used as the basis of numerical methods for shape discrimination, comparison, and interpolation, for longitudinal shape studies [11], and for deformable image segmentation and registration [22, esp. 23,24]. (See [25] for a recent summary of clinical applications of both space filling and shape based image registrations.) Fig. 1 diagrams the process of training and use of the estimated probability distributions on the object(s). This paper is concerned with defining a framework for fitting parametric models to training images while maintaining credible, legal shapes with locational correspondence across the training population.
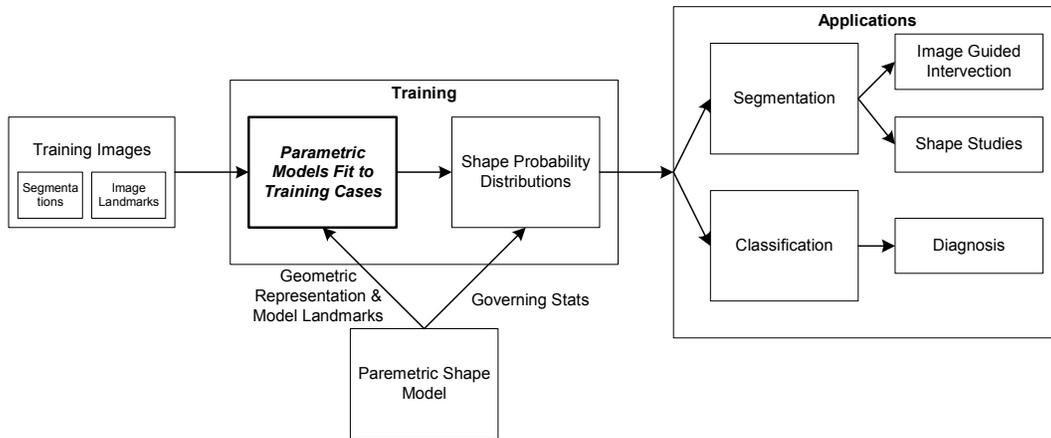
Fig. 1. The process of training and using probability distributions on object shapes in clinical applications. The step outlined in bold, fitting training images with parametric models as input to statistical analysis, is the main subject of this paper.
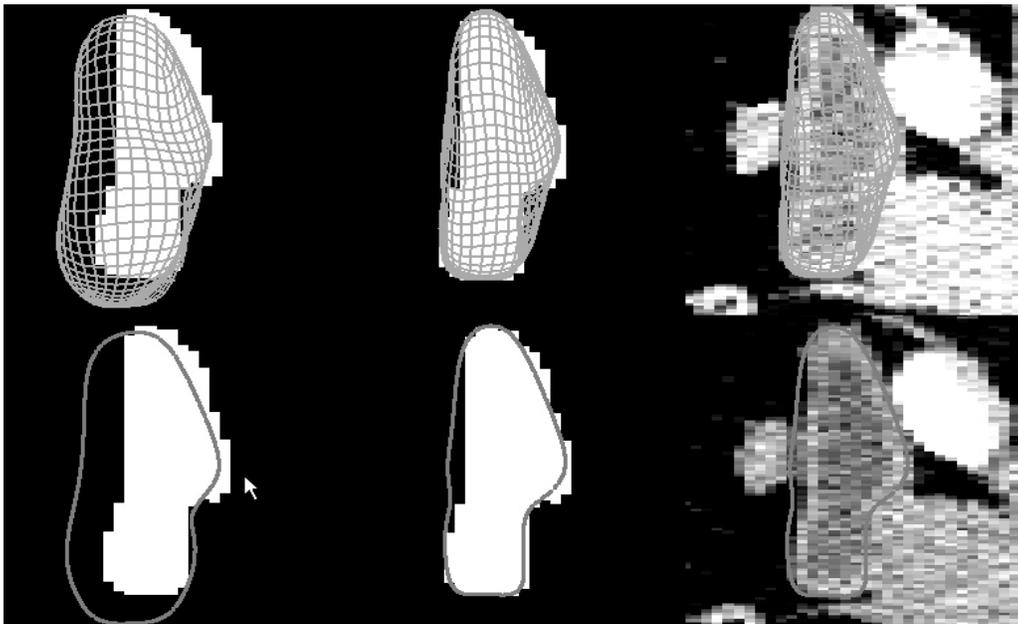
Fig. 2. Fitting a statistical deformable model to a target training image. (Top) 3D surface views and (bottom) single sagittal slice views of bladder template geometry (left) coarsely aligned to a target training image, (middle) deformably fit to that image, and (right) in the context of the actual grayscale data.

We consider shape training as a special case of Bayesian image segmentation [26]. In SDSM-based segmentation, a starting shape is initialized to a target image and automatically deformed based on the eigenmodes of shape variation to optimally match the image data, as illustrated in fig. 2. The deformation is driven numerically through the SDSM parameter space by minimizing an objective function including a term reflecting the probability of the deformed shape and a term measuring

how well the shape matches the image. However, by using a shape parameterization that supports reasonable physical deformations, it is possible to define purely geometric conditions which establish and maintain legality and the locational correspondence over the training population we need to compute correct statistics. After the training population has been fit, these first-round shape estimates are used to compute coarse statistics. The training images are then refit about a new mean, relaxing the purely geometric constraints as the shape statistics are iteratively refined. This recomputation of an atlas, unbiased to a particular subset of the training instances, by computing a successively better estimate of a mean as the fits improve is comparable to the iterative computation of a Fréchet mean and fits to it used in the non-rigid registration framework [18,27]. However, in our method not only the mean but also the variability is being recomputed at each iteration.

Our automatic method for shape training has been applied to a variety of anatomic structures from both synthetic and real 3D medical images, such inter-fractional and cross-patient male pelvis, inter-fractional head and neck, and cross-patient brain and kidney images. Whereas inter-fractional SDSMs cover day-to-day anatomic shape change, such as the bladder filling and emptying, etc., cross-patient SDSMs must be able to account for much broader anatomic shape variability, such as how prostates differ in both shape and relationship to the neighboring bladder, rectum, and pubic bones from patient to patient.

2. Method

Extraction of shape representations of training cases has been accomplished in the following ways for the boundary-only shape models that are common in the statistical shape analysis literature. For these, the model is intiialized by sampling the boundary voxels of each training sample to produce a non-folding tiled surface. Then each case's surface is reparameterized to give correspondence according to geometric, mechanical, or local image properties. As examples, [28] uses no image match grounds, but requires a regular sampling with tight geometric distributions. Active appearance models [12,29] require both a tight geometric distribution as well as similar local image intensities. The training method used in [30,31] reparameterizes the sampled surface according to an orthogonal decomposition based on spherical harmonics, then resamples the surface at locations corresponding to equally-distributed points on the parameterizing sphere in that framework. [32] uses a finite element model to track voxel correspondences for a retrospective dose study. [33]

describes parametric topological constraints for active contour models (snakes) [24,34] or active shape models. Many of these approaches can obtain good results, but none directly constrain their shape statistics to the generation of only credible shapes or, for most, even of non-self-interpenetrating shapes.
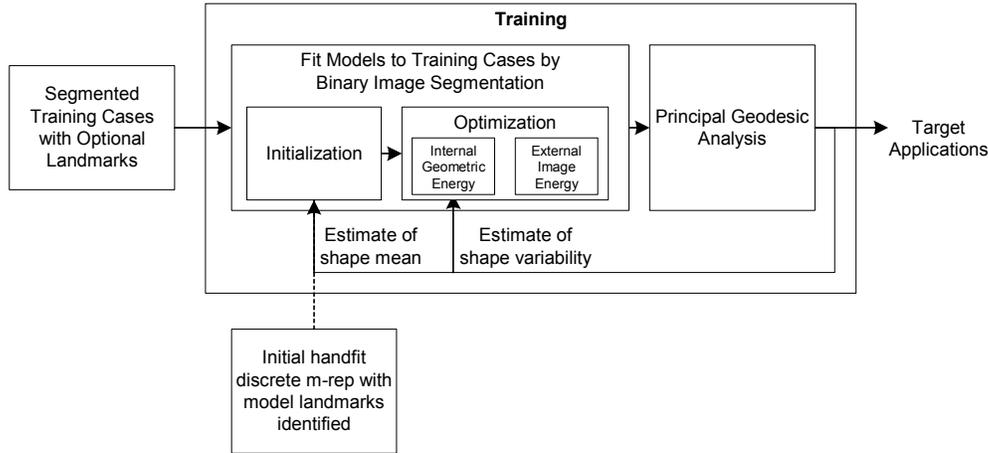
Fig. 3. A detailed view of our method for the training step shown in fig. 1. Shapes are fit to the training images iteratively according to a binary image segmentation algorithm, with purely geometric terms relaxed in favor of converging group statistics. See Appendix I for a summary of the discrete medial representation and principal geodesic analysis.

To obtain correspondence while maintaining a space of credible objects, our process is designed as a special case of binary image segmentation within the probabilistic segmentation framework. The goal of our training process is therefore to jointly estimate both a parametric shape space that covers the organ shape over all *n* images, and at the same time to obtain the best set of corresponding descriptors of parametric models for each organ in each image. As illustrated in fig. 3 and overviewed in algorithm 1, in this framework an energy function composed of internal geometric forces and external image match forces is optimized over a shape space. Typically this space comes from the trained probability distributions, but in their absence, we rely on a shape parameterization that supports reasonable physical changes and a carefully written internal geometric term to constrain possible shape changes.

The parameterization that we use is a discrete medial representation called *m-reps*, although the ideas can be generalized to other SDSMs with a well defined volumetric coordinate system and an explicit encoding of volumetric legality. See Appendix I for a summary of m-reps [12] and its

governing statistical model, principal geodesic analysis [35]. Since m-reps capture not only

positional information (i.e., of the skeleton and the implied boundary) but also orientational

information (of spokes between the skeleton and the boundary), the parameterization handles

plausible physical changes such as local twisting and bending in a way that can be easily captured

by the statistical model. Further, we can leverage differential geometry directly in our legality term

to enforce non-self-interpenetration.

Algorithm 1: Iteratively Training Models of Shape Variability

```
Input:  I, a collection of image segmentations i defining geometric truth

1. R₀ = HandFit( i in I s.t. i is reasonably typical )
2. Until difference(Rₖ,Rₖ₊₁) < convergence error:
   2.1.   For each i in I:
      2.1.1.  Find mᵢ s.t. mᵢ = Arg Min₍ₛᵢₘ ₜᵣₐₙₛ ₒf Rk₎ Energy(Rₖ,i)
      2.1.2.  Find mᵢ s.t. mᵢ = Arg Min₍ₘₗₒbₐₗ dₑfₒᵣₘₐₜᵢₒₙₛ ₒf Rk₎ Energy(Rₖ,i)
      2.1.3.  Find mᵢ s.t. mᵢ = Arg Min₍ₗₒcₐₗ dₑfₒᵣₘₐₜᵢₒₙₛ ₒf Rk₎ Energy(Rₖ,i)
   2.2.   (Rₖ₊₁, Shape stats) = Principal Geodesic Analysis( M={mᵢ} )
```

This method section of this paper is organized to follow fig. 3. After discussing the inputs to the

process in section 2.1, we describe the initial model-to-image alignment in section 2.2. The energy

function is overviewed in 2.3, the internal geometric term is explained in 2.4 and the external image

term is explained in 2.5. The iterative optimization framework is described in 2.6. Algorithm 1

gives an additional overview of the method. A major contribution of this paper is the iterative

fitting of models and shape probability estimation indicated in step 2 of algorithm 1. Additional

novel contributions are describing the principles and design of shape terms for the energy function

used in steps 2.1.1-2.1.3 that maintain object correspondence and legality during the training fitting,

and describing the iterative shape estimation where purely geometric terms are relaxed as shape

statistics become available.

2.1 Materials

For inter-fractional shape study, our method begins with a collection of CT scans from a single patient over the course of a multi-day treatment regimen. Typical images are 512×512×40 voxels in dimension and have 1×1×3 mm voxel resolution. The CT images include both treatment and non-treatment organs. Between 12 and 18 scans are collected for inter-fractional studies. Other studies have slightly different image characteristics, but the method remains the same; details are given in the results section for the various case studies presented.

Expert raters segment structures of interest in each image by drawing contours of target and non-target organs on transaxial slices. Methods to generate ground truth from these segmentations comprise a separate area of research beyond the scope of this paper [36]. Because the quality of the manual segmentations is critical to the quality the derived model, at the present time we train against individual raters whose segmentations have been reviewed and, if necessary, edited by other experts. Precautions are taken to control for intra-rater variability across images of the same patient. For example, segmentations from one image are transferred to the next, coarsely registered with the corresponding objects, and then edited to match the objects in that day's image. Moreover the prostate volume is not allowed to change from day to day. Based on anecdotal evidence these precautions are believed to control for random variations that would be observed when segmenting each image *de novo*. The segmentations are then scan-converted to a label image per organ. These label images have the same extent and resolution as the source images, but all voxel values are set to either 1, if that voxel is inside or on the segmentation's boundary, or 0 otherwise. Throughout this paper, we will refer to these per organ label training images as the set I_{1-n}, where n is the number of images in this population.

2.2 Initialization

Our training process starts by hand manipulating a model to be modestly representative of the shape we are attempting to train. This typical shape is the presumptive geometric template for the shape space, R_0. R_0 is our reference shape, from which all other shapes will initially be derived and relative distances measured. As long as the initial model has sufficient sample resolution and correct topology (e.g., tubes for vessels, a bent blob for kidney, a multi-blob figure for liver), our

method converges for even arbitrarily poor starting models such as a generic slab. However, it is quite clear that the quality of the SDSMs we compute and the speed of convergence is directly related to the input quality at each step, making a good initial shape estimate particularly important.

Using R_0 as a coordinate system governing our region of interest, we can identify in R_0 a set of model-centric landmarks, *LM*, which we wish to keep explicitly in correspondence to within some tolerance throughout the training population. These *LM* correspond to a set of explicitly identified anatomic features, *LI*, landmarks noted by the raters on the segmentations they manually produce. For example, when training a prostate shape model, we use posterior and anterior poles and the urethral entrance and exit. Although we typically use only surface points as landmarks, our shape parameterization provides a volumetric model-centric coordinate system, which allows landmarking of positions both inside or on the object, such as the opening of the prostatic urethra, and outside the object, such as the nearest boney landmark to the prostate.
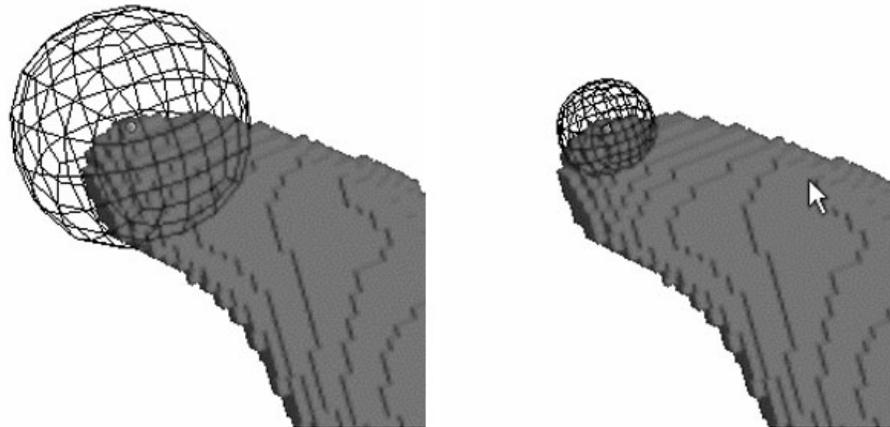
Fig. 4. An image landmark identified at the tip of a segmentation with (left) large tolerance and (right) a tighter tolerance.

The initial step in fitting R_0 to a particular label image is then to align R_0 to the image according to the landmark pairings LM to LI. This alignment can remove global transformations such as translation, scale, or rotation from the shape-space we are attempting to estimate. The more tightly aligned the training cases are to each other, the more tightly the geometric template will be able to fit the data, and the more representative the derived statistics of deformation will be. We have explored a variety of methods for designating this initial alignment: we frequently use a Procrustes similarity transform, although with some highly variable objects, we could use an affine transform.

And one could imagine integrating even more complex non-linear landmark based registrations such as [37,38], but then the probability distributions would not be describing these shape changes. In the absence of explicit landmarks, implicit landmarks from curvature or other derived properties may be computed. For simple synthetic objects, it is sometimes sufficient to align the centroid, volume, and orientation of R_0's surface to by similarity transform to the image moments of the binary training image.

2.3 Defining the Energy Function

The best deformation of R_0 into each training image is computed numerically, by optimizing model parameters according to a metric that measures the goodness of a model, *M*, fit to a given image, *I*. This energy metric normally consists of two terms, an "external" image match term and an "internal" geometric typicality term. This division can be theoretically founded in probabilistic terms. We seek $M = \text{Arg Max}_M \, p(M|I)$, i.e., the model with the greatest probability density given the image. This formulation is frequently called the posterior density, so the method is called one of posterior optimization [26]. An application of Bayes rule together with an application of the logarithm to both sides of the equation and removing terms that are constant in M yields $M = \text{Arg Max}_M \, [\log p(M) + \log p(I|M)]$. $\log p(M)$ then measures the *geometric typicality*, and $\log p(I|M)$ measures the *image match*. Other choices for these two terms are possible; but if one assumes Gaussian distributions, it is natural to use squared distances as proxies for log p since the log of a normal distribution is linear in a squared distance. The remainder of this section describes how we go about designing the distance terms that are minimized in our energy function.

2.4 Internal Geometric Energies

Without statistics, estimating shape likelihood requires approximations specific to the characteristics of the underlying parameterization. The essential problem is that the underlying parameter space from which the models are drawn is much larger than the shape space that we are trying to cover. Therefore, we restrict the models generated during training to a compact and legal

subspace where we are confident that features are in correspondence across the training population. To this end, we impose three purely geometric conditions on the posterior.

First, we impose a penalty against nonuniformity of samples, i.e., irregularity in the discrete medial grid. Second, we use a legality measurement based on differential geometry to identify and prevent local self-intersections. And third, we penalize shapes that are not near the presumptive mean. In addition, in recent trials a reward for model smoothness has been shown to be of use.

2.4.1 Sampling Control

Establishing feature correspondence across the training population implies that there is a unique set of parameters that best describes each training case. However, it is possible to find two quite different models that have nearly the same image match in *I*, so we differentiate them by establishing an additional geometric criterion. As an example, there are many possible cubic approximations to a given function, but by preferring certain end conditions, we can identify a unique "best" approximating function.
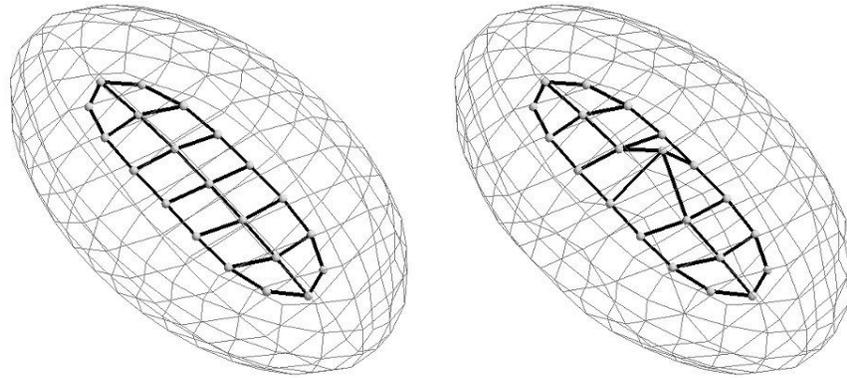
Fig. 5. A medial mesh (thick lines) and implied surface (thin lines) with a (left) high nonuniformity penalty and (right) low nonuniformity penalty. Meshes with high irregularity may imply similar surfaces as more regular meshes, but can result in qualitatively inferior results and break our volumetric correspondence assumptions.

In our case, we want a unique and legal "best" model to represent a given training image. That is, if shape models *M* and *O* both represent the same data, we need that $M \approx O$. Otherwise, the correspondence requirements will not be met. The geometric criterion that we use to distinguish medial sheets with similar implied boundaries is the uniformity of the medial grid. Medial sheets with samples bunched together form difficult to interpolate configurations that poorly capture the

object. Following the Markov assumption that the likelihood of a sample conditioned on the model is the same as the likelihood of the sample conditioned on its neighbors, we define nonuniformity as a function of the agreement between any sample and the expectation of its neighbors. The total nonuniformity of the model is the sum of such sample agreements. Since we are producing a dissimilarity term, Nonuniform(M) is actually defined as the distance $d^2(M, \overline{M})$, where $\overline{M} = M$ smoothed by a low-pass filter. As a side effect, this definition of smoothness also tends to act as an *ad hoc* enforcement for shape legality. Fig. 5 shows an example of uniform and nonuniform organizations of medial samples.

In particular, in order to define regularity of a sampled medial representation, we use the Riemannian distances and means as described in Appendix I. We define $\overline{M}$ as the model such that each medial sample $\overline{m}$ is the distance-minimizing Fréchet mean of its neighboring samples. Practically, using our parameterization, $\overline{M}$ is a curvature minimizing medial sheet with regularly spaced samples. Nonuniform(M), the distance between M and its smoothed $\overline{M}$, can be reduced and rewritten as (1).

$$Nonuniform(M) \sim \sum_{m_i \in M} \sum_{m_j \in \text{Neighbors}(m_i)} d^2(m_i, m_j) \tag{1}$$

This simple function, which essentially normalizes volume per respective medial sample point is suitable to our parameterization because medial representations inherently imply desirable types of correlated surface deformation such as bending, twisting, and magnification.

2.4.2 Legality

Correspondence assumptions can also be undermined by illegal shapes. As we perturb the model parameters to find the best fit to a segmented image, the model's surface may fold or develop self-intersections. Medial representations have very precise analytic forms for computing shape legality on the medial manifold [39]. The largest eigenvalue of the radial shape operator yields a legality measure at every sample point with a well defined threshold for identifying folded surfaces [40]. As we narrow down the shape space during the fitting procedure, we tend to relax Nonuniform() and rely more on the legality measure. Similar smoothness and legality functions can be mathematically derived for other shape representations specifying local orientations.

The penalty for illegality for any medial sample is set to zero when the largest eigenvalue of the radial shape operator is nonnegligibly less than its threshold. Then it rises first slowly and then sharply as that eigenvalue passes its threshold. The overall penalty is the sum of the penalties of the samples.

2.4.3 Reference Model

Given a space full of possible shapes, we desire to identify the mean shape and expand our shape model about it. Under Gaussian assumptions, given a mean and the variances of the principal directions of deformation, this method provides a shape-normalized Mahalanobis distance from any particular shape to the mean shape. Using this as the geometric prior in optimization tends to keep the candidate models clustered near the mean where our statistical shape model is most likely to be valid.

However, with limited *a priori* knowledge, we use our initializing model R_0 as a tentative reference point for the shape space. In the early iterations of model fitting we assume an isotropic shape probability distribution, i.e., that the difference between two shapes is measured only by a weighted Euclidean distance, we can estimate a geometric prior as the Riemannian model-to-model distance between a candidate model, M, and R_0, with distance defined as in the previous section.

$$\text{Reference}(M, R) \sim \sum_{m_i \in M} \text{d}^2(m_i, r_i) \tag{2}$$

As both of the geometric penalties, Nonuniform() and Reference(), are in the form of a distance, they can equally be thought of as log probabilities on Gaussian distributions in the feature space. Taken together with the illegality penalty, they comprise the internal geometric energy in our optimization metric.

2.5 External Image Energies

The external image terms of the energy function measure how well the model's surface fits the label image. This is measured in two terms, one that computes distances between specific landmark points in the model to corresponding landmarks identified in the image, and another that computes a

general surface-to-surface distance between the model surface points and the nearest boundary voxel in the label image.

2.5.1 Landmarks

The landmarks identified automatically or manually in section 2.1 require an explicit expression in the metric in order to prevent them from drifting out of alignment during optimization. We assume that the model coordinate landmarks should be normally distributed about the landmarks identified in the image, that is, the model landmarks have the distribution $N(LI,\rho^2)$, where ρ_i is the standard deviation, i.e., tolerance, for the ith landmark. Because the log probability of $N(LI,\rho^2)$ is simply ρ^{-2} $d^2(LI,LM)$ with d^2 the standard Euclidean square distance, we can express the error in landmark match as (3).

$$\text{Landmark}(LM,LI) \sim \sum_{lm_i \in LM} \frac{1}{\rho_i} d^2(lm_i, li_i) \tag{3}$$

2.5.2 Binary Image Match

The image match term is computed as the sum of squared distances between the boundary voxels, B, of the label image, found using a six-connected neighbor test, and the continuous boundary surface, Ω, implied by a candidate model M (see Appendix I). This is equivalent to thinking of each surface point or boundary voxel as an equally weighted landmark corresponding to its nearest point in the other set.

Ideally we desire to measure two terms, the distance from each member of Ω to the closest voxel in B, denoted $d^2(\Omega,B)$, and the distance from each member of B to the closest surface point on Ω, denoted $d^2(B,\Omega)$, which are not equivalent. The distance $d^2(\Omega,B)$ can be computed fairly quickly across many candidate models because the label boundary remains static over all the trials. We generate a single space filling lookup table for distance from the label boundary by a modified anisotropic version of Danielsson's algorithm [41]. Trilinear interpolation gives a very fast measure of the distance at any point in space to the closest boundary point on B. Discrete samples ω are taken arbitrarily densely from each candidate Ω and we then let $d(\omega,B)$ be the lookup of the position of ω in the distance map. Integrating over all surface samples gives $d^2(\Omega, B) = \sum_{\varpi \in \Omega} d^2(\varpi, B)$.

The inverse distance $d^2(B,\Omega)$, however, is computationally exorbitant given the finely sampled subdivision surfaces required for accurate matches and the large number of candidate surfaces generated for optimization. However, simply ignoring this part of the term can lead to undesirable results in areas of high curvature. In these areas, the $d^2(B,\Omega)$ distance-minimizing Ω tends to be more volume filling than the minimizer of the $d^2(\Omega,B)$ distance. Our solution is to compute the $d^2(B,\Omega)$ distance at only a minimal number of points where we would expect large distance asymmetries. We identify such points on Ω by computing the angle between the gradient of the distance map and the surface normal. If the angle is greater than a threshold, we compute a new distance along the surface normal at that point as can be seen in fig. 6. With $\theta^2(x,y)$ as this modified minimum distance function, then our data likelihood term can be expressed as in (4).

$$\text{Image}(M,I) \sim \sum_{\omega_i \in \Omega} \theta^2(\omega_i, B) \tag{4}$$
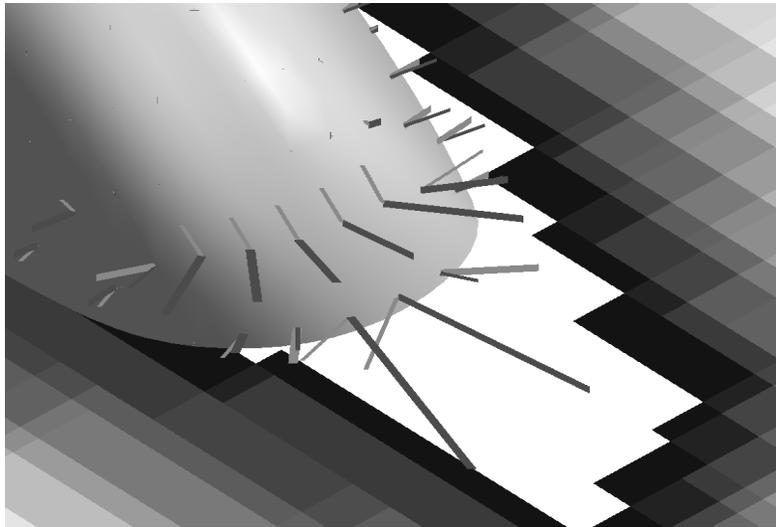
Fig. 6. A slice from a distance map and a suboptimally fit surface illustrating θ. The light gray lines show the distance map gradient direction; dark gray lines show the surface normal direction.

This modified distance map method is dramatically faster and produces superior results to the standard approach of understanding binary images using local edge detectors, such as the derivative of Gaussian filter. While edge detectors are a logical extension of the model-to-image match as it would be computed in a grayscale image, they suffer from significant problems with capture range and orientation.

An additional advantage to using a medial parameterization for M is that M and B are strongly related according to morphological erosion. To fit models to images with structures that are only a

few voxels in thickness, we can fit an initially dilated model to a dilation of the labeling, and then contract the model surface by the same amount by an inverse scaling of the thickness parameter. As seen in fig. 7, this morphologically closed model approximates the thin object much better than is possible otherwise.
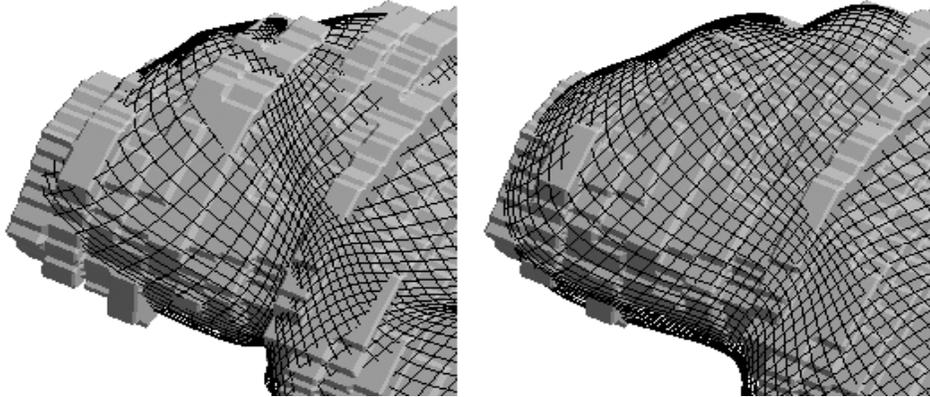
Fig. 7. Two candidate model meshes compared to tiled surface of a segmentation of the thin masseter muscle in the neck. The mesh on the right has been fit naïvely; the better fitting mesh on the left has been fit to a dilated image and then contracted.

In the Bayesian framework, the image term, P($I|M$) is the probability that the surface of the model is in alignment with the boundary voxels of the segmentation. The landmark and binary image match energies we describe can be thought of as the log probabilities of a joint distribution and added together into the optimization metric.

2.6 Optimization Framework

We now have an initial geometric template, R_0, that has been coarsely aligned to the image and a well defined non-statistical objective function which has as its minimum a model configuration which should be both well fit to the target image and qualitatively similar to other models fit to other images of the population. Our process is then to deformably fit each image in the training population about R_0. The mean of the population of the fit models, R_1, is computed along with an estimate of the principal modes of deformation. The population is then refit iteratively about R until R converges to a suitable mean of the training population when $R_{i-1} \approx R_i$.

Each step of the iteration is a search for the error minimizing *M*, with error, *E*, computed via our complete dissimilarity metric, that is, the sum of (1-4) and the Legal() term described in 2.4.2, using α, β, δ, and γ as relative weighting factors.

$$E = \alpha \text{ Landmark} + \beta \text{ ImageMatch} + \delta \text{ Nonuniform} + \text{Legal} + \gamma \text{ Reference} \qquad (5)$$

Our optimization engine uses a conjugate gradient descent [42]. Because conjugate gradient search performs best given a relatively isotropic global minimum, some experimentation is required to fine tune the weighting factors to new shape studies. These weighting factors are essentially scaling between the Euclidean voxel-space distances and the Riemannian sample-space distances.
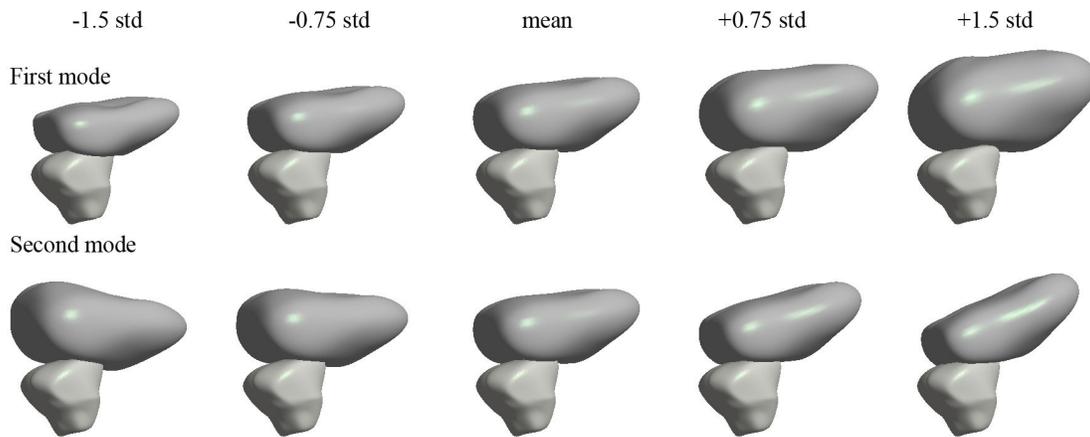
Fig. 8. A mean bladder (darker upper object) and its first two principal modes of deformation relative to the mean prostate (lighter lower object). These two modes of deformation together cover over 65% of the shape variability across the 18 images of this patient.

Before we have modes of shape change, we assume that samples can move about freely in their feature space, restricted only implicitly by the distance metrics in the geometric constraints. After the first round of fitting, we can compute intermediate shape statistics by looking at correlations in feature relationships across the training samples. For boundary models with implicitly Gaussian distributions, [12] proposes using principal component analysis, which relies on the eigenstructure of the feature covariance to describe correlated shape change. [43] gives a relevant example of applying PCA to anatomic shape variability. M-reps and other representations with explicit orientational components cannot use PCA because distances between features that encode orientation and scale are not Euclidean. These representations are governed by a generalized PCA known as principal geodesic analysis (PGA) [35]. As with PCA, an observation matrix relating measurements of corresponding features across the training population is formed and the covariance

of the observation matrix is computed and diagonalized. However, the non-linear orientation and size terms of our feature vectors are first mapped onto a tangent plane fit locally to the Fréchet mean of the shape space, *R*. The eigenvectors of the system orthogonal basis of correlated feature changes which are projected back into the non-linear feature space and imply correlated surface deformations such as those shown in fig. 8. This basis can be truncated according to eigenvalue thresholds, which describe the significance of each of direction of shape change. See Appendix I for details of to decompose a set of m-reps according to PGA and how to apply the resulting statistics to individual instances.

As we iterate and gain confidence in our intermediate statistics, we replace our sampled medial shape parameterization with a short vector of coefficients of each principal mode of deformation. This restricts the shape changes we optimize over and both obviates our smoothness requirement and allows us to compute the geometric prior by the correct shape-to-mean Mahalanobis distance. This covariance weighted distance is computed directly as the eigenvalue scaled PGA coefficients. That is, the Mahalanobis distance squared $= \sum \alpha_i^2 / \lambda_i^2$, where α_i are the coefficients of *M* expressed in PGA eigenvectors and λ_i are the corresponding eigenvalues.

Following an intermediate step of statistical shape optimization, local refinements may still be required for individual samples within the model. These refinements can be computed deterministically, by another optimization of our error metric (5) restricted to the parameters of a given sample, or they can be probabilistically estimated by PGA over the differences between the results of the statistical stage and the deterministic refinement. Tracking the residual optimization work on individual samples still required after an intermediate statistical fitting gives us a measurement of when the training is complete.

Extensions of the statistical model allow us to compute joint statistics of interdependent multiple fig. shapes [44] and multiple object shape models, which leads to the multi-scale methods for describing both global and local phenomena discussed in our image segmentation paper [12].

3. Results

Our method is routinely tested by computing a shape template and modes of deformation for a set of procedurally generated binary 3D images of bent, twisted, and tapered ellipsoids such as those in fig. 9. A set of images is generated by randomly sampling values for each of the three parameters, analytically generating the corresponding ellipsoid and scan converting it into voxels. Vertex

landmarks are computed for each image. The medial manifold of a standard ellipsoid is a simple primitive in our framework. This standard ellipsoid is designated R_0. R_0 is initially aligned by landmarks to each of the training cases, and then the parameters of each medial sample are optimized according to the error function described in our method section. When the best model for each image has been found, we compute the final average and maximum voxel distances from each model to its corresponding image as shown in fig. 10. Fig. 11 is a histogram of these distances over the standard twenty test cases. Shape statistics as in fig. 12 are computed and we refit the population by optimizing over the principal geodesics of shape change. Note that PGA has substantially reduced the complexity of the optimization. We input a three parameter distribution into the problem, our initial fitting optimizes on the order of one hundred parameters, and using PGA we can reconstruct over 90% of the shape variability using only five model parameters. The distance histogram of the statistically fit models is also given in fig. 11. Another encouraging result is that while the two training outliers we see in fig. 11 were thrown out of the statistical analysis, they have been adequately covered by the statistical fitting.
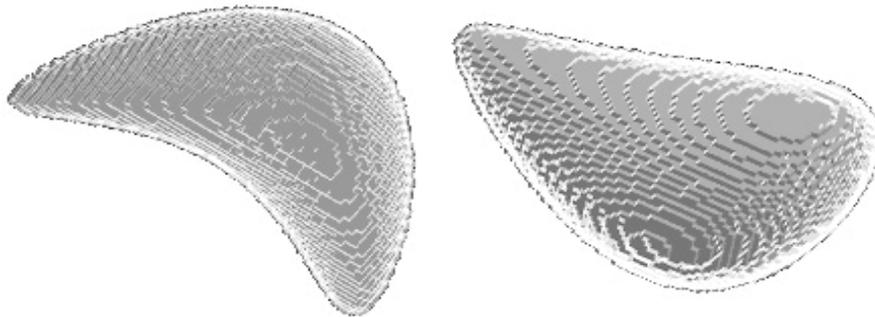
Fig. 9. Tiled surfaces from two procedurally generated warped ellipsoid test objects showing bending and tapering.

Fig. 10. Relationship between the boundary voxels of an ellipsoid binary training image (gray) and the fitted model's surface (black) through a transaxial slice.
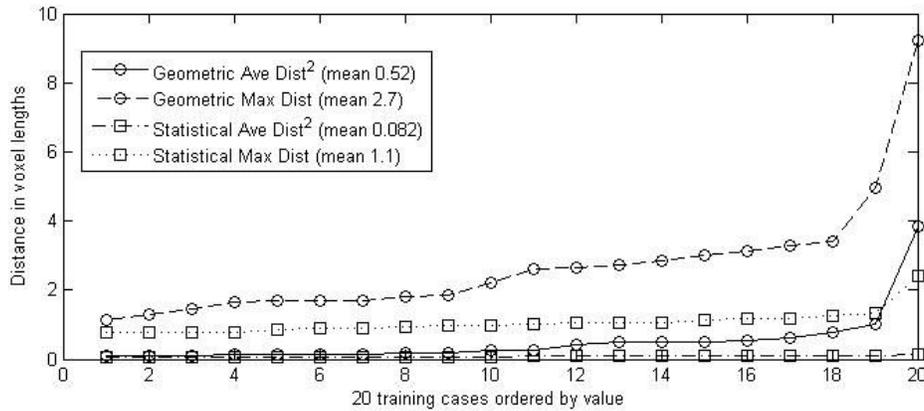
Fig. 11. Histogram of average and max distances for warped ellipsoid models over 20 training cases. The two outliers were excluded from the first round statistics, but were successfully fit in the next round using the recovered statistical modes of deformation.
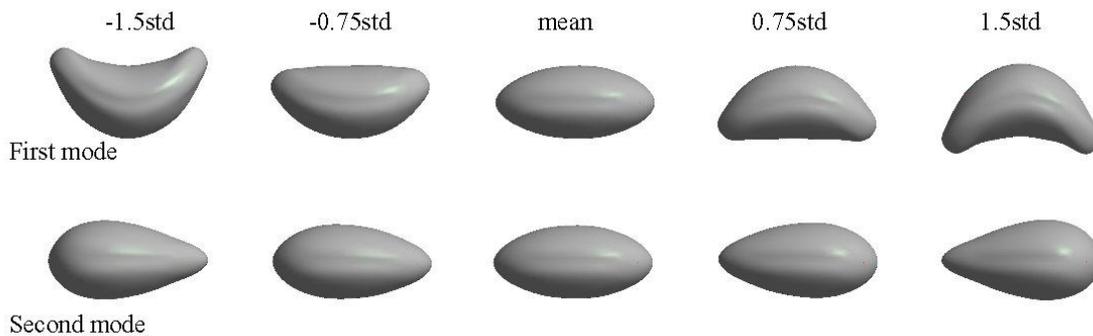
Fig. 12. As we would expect, the first two principal modes of deformation trained from 20 bent, twisted, and tapered ellipsoids reflect bending and tapering.

The ultimate indication of our methodology's effectiveness is its application to real medical and scientific problems, such as modeling shape variability in inter-patient and cross-patient images. In fig. 2, we show a mean bladder being initialized and deformably fit to a patient image, represented by a single sagittal slice of the label and corresponding gray image. In this case, we start with a model hand-fit to day 1 of each patient series. Subsequent daily images are aligned according to the position and orientation of the urethral landmarks. We then proceed with two iterations of the training as outlined in our method, first a non-statistical and then a coarse statistical stage. This procedure is repeated for the bladder and prostate. Fig. 13 shows a histogram of average and maximum distances over models fit to bladders, prostates, and rectums from sets of inter-fractional images of twenty-five different patients. The average error across all organs is less than a voxel and the large maximum distance seen in the rectum training is artifactual of segmentation variability.

The bladder is a particularly challenging object to model effectively because it exhibits large shape changes from day to day. The first two PGA modes of shape change for one patient's bladder and prostate are shown in fig. 8. These two modes account for over 65% of the patient's daily shape variability in those two organs, and as we would hope, could be seen as roughly accounting for the bladder filling and emptying. The statistical deformable shape models for each patient from such training, combined with a suitable CT image match term, allowed us to produce leave-one-out segmentations of actual images from four of the patients with mean volume overlaps (intersection/average) of 92.5%, 93.7%, 91.6%, and 94.7% respectively using only eight modes of deformation [45].
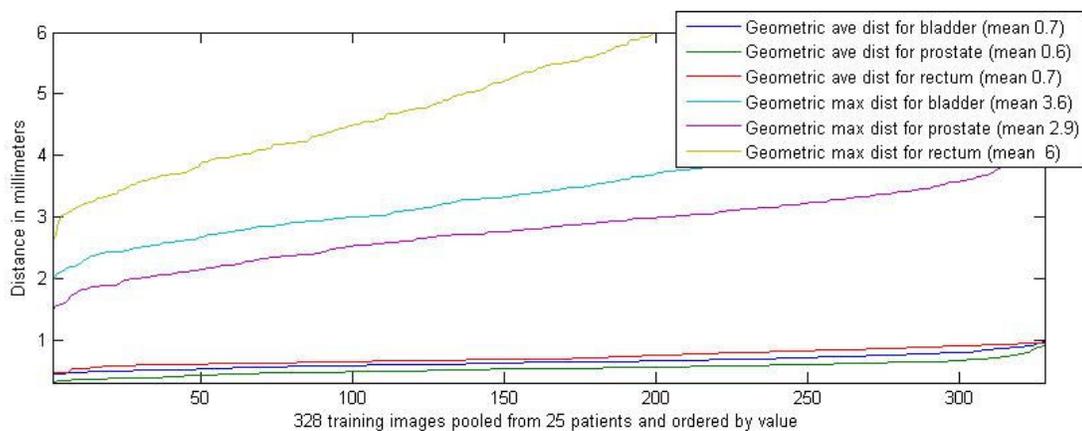
Fig. 13. Histogram of average and max distances for fit bladder models over 67 training cases pooled from 4 sets of same patient inter-fractional images.

We also applied out method to a set of 50 cross-patient kidney images. In this case, the SDSM must account for actual anatomic variability across patients, a much broader kind of shape change than that usually seen in sets of within-patient images. The eigvenvalues of the recovered principal modes of deformation show that we now need fifteen modes of variance cover 95% of the anatomic variability. Our training resulted in 95.3% volume overlaps (int/ave) of the models to the segmentations. The trained 15 parameter SDSM applied to leave one out segmentation of the actual patient images resulted in models with volume overlaps only slightly lower, on average, 91.1%. For details of our kidney segmentation results, see [46,47,48].

Our method has also been applied to several other target areas. Head and neck models suitable for radiotherapy planning are shown in fig. 14, left. The deep brain structures shown fig. 14, right, were taken from an autism shape study where the researcher used the SDSM based shape-to-shape metric to discriminate autism by shape characteristics [11]. We have also modeled

hippocampi for a large statistical shape study. Temporal extensions to the method have enabled studies of heart motion [49], which has obvious extensions to lung motion and 4D-ART. [50] uses an SDSM trained with our method to establish feature correspondences in dog hip joints for a longitudinal study characterizing canine hip dysplasia.
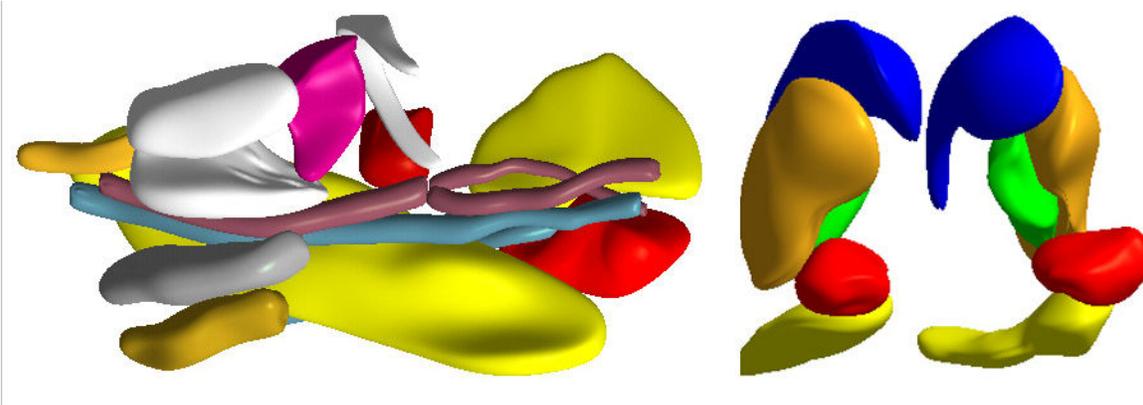
Fig. 14. Multi-object shape models. (Left) A fifteen object complex of structures from the head and neck. (Right) Deep brain structures from an autism study, left and right hippocampus, amygdala, putamen, caudate, and globus pallidus.

All the results discussed were produced using a C++ implementation of the algorithm. Using our implementation on a modern 2GHz desktop, a model can be fit to an image in less than two minutes. Batch fitting is trivially parallelizable onto any number of machines up to the number of images in the training population. Using such a parallel computing framework and given a set of segmented images and landmarks, a completely trained SDSM can be produced in less than an hour.

4. Discussion

Shape training is a key step in the application of deformable shape models to such problems as image registration required for IGRT or ART, shape classification, and longitudinal shape studies. Our methodology for training statistical deformable shape models is a special case of image segmentation, using estimates based on desiderata of our shape space and successive refinement to derive a principled statistical model. The internal geometric terms we use guarantee training shapes that are regularly sampled, legal, and compact about a mean, which together cause credible shapes to be reflected in the statistics. The method described is fast and routinely gives good meshes. The framework has been described in the context of the discrete medial representation, but should

extend to b-reps with orientation or other representations where measures for sampling regularity and non-self-interpenetration can be analytically computed.

Acknowledgments

Appendix I – The Discrete Medial Representation

Medial geometry [51] describes 3d objects in terms of a skeletal surface, a 2d curved sheet lying midway between opposing surfaces of the object, and a set of spokes extending to the object boundary from both sides of the skeletal surface. The medial manifold, **M**, of a three dimensional object has eight parameters at each point (u,v): **M**(u,v) = {position (3), spoke length (1), and two spoke directions (2x2)}. Some additional complexity is introduced along the crests at the edges.
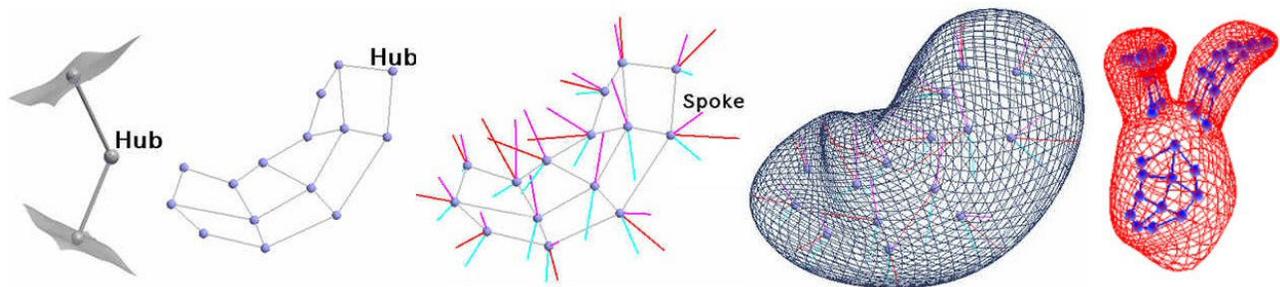
Fig. 15. Discrete medial representations. (Left) A medial sample with two equal length spokes that touch opposing surface patches. (Mid Left) A sampled skeletal sheet for a kidney with neighbor relations marked. (Middle) Spokes at each medial sample describe the orientation of the implied surface at that hub. (Mid Right) A densely sampled surface can be interpolated from the medial samples. (Right) A prostate model with sub-figures defined for the left and right seminal vesicles.

The discrete medial representation, m-reps, samples the continuous manifold on a grid, yielding a set of 8- dimensional medial samples which taken together act as control points for the object's

volume, as shown in fig. 15. Additional medial points can be interpolated according to [52], which in turn imply a denser surface sampling. Alternatively, additional surface points can be approximated directly using a modified Catmull-Clark subdivision algorithm [53] with additional normal constraints. **M** also implies a volume filling hexahedral mesh useful for computing mechanical deformations according to finite element methods [54]. An object made from a single grid of medial samples is called a *figure*. A single column grid implies a tube figure; a multi-column grid implies a slab figure. Indentations and protrusions are handled as attached subfigures [44]. A figure along with any associated sub-figures is called a *model*, shown in fig. 15 (right).

The discrete medial representation is well suited to Bayesian segmentation because it provides 1) a volumetric coordinate system in which to gather image intensity statistics, 2) a low dimensional parametric shape space amenable to numerical optimization, and 3) an analytic description of legality. Each figure can be examined sample-by-sample where we need only eight parameters to represent complex object-based deformations such as bending, twisting, and magnification. Parameter changes can be constrained to imply only legal models by computing a radial shape operator similar to the surface shape operator described in [55], but dependent both on local shape and on local orientation. Measuring the largest eigenvector of the spoke length radial shape operator at each grid position [39,40] gives an immediate indication of local self-interpenetration in the volume.

Distances and Means of Samples and Models

Every medial sample can be understood as a translation, magnification, and rotation of any other. Thus, we can define Riemannian distances between them for computing means and variabilities. The spoke directions describe a rotation, which along with the radius, are multiplicative terms which must be logarithmically mapped so that sample-to-sample distances can be computed in a Euclidean space [35]. The Fréchet mean, $\overline{m}$, of a set of samples $\{m_i\}$ can be computed as the point in the parameter space that minimizes its distance to every member of $\{m_i\}$. Samples on the grid are also given neighbor relationships, which along allows for Markov fashion predictions (see section 2.4.1). Sample-to-sample distances can be extended to a distance between two models, M and O with samples $\{m_1,...,m_n\}$ and $\{o_1,...,o_n\}$ respectively, by taking the sum of the Riemannian distances between corresponding samples as $d^2(M,O) = \sum_{m_i \in M} d^2(m_i, o_i)$. Distances taken from object-to-object and then refined sample-by-sample provide the basis for multi-scale shape analysis.

Using these definitions for distance, a typical shape and descriptions of shape variability can be computed from a set of models according to a generalizing of principal component analysis called principal geodesic analysis [35]. The mean for each atom over a training population is computed by the Fréchet approach of minimizing the sum of squared geodesic distances to the result point. The atoms in each training m-rep is projected (mathematically called the Log map) to a tangent space at the computed mean, where PCA is done on the collection of atoms. Mapping the resulting principal direction vectors back into an m-rep increment from the mean creates modes of global variability that encode not only position but orientation and width. The relative weight of the associated eigenvalues give a reciprocal standard deviation weighting to this space of deformations and provide the basis for a shape-to-shape Mahalanobis distance. The Mahalanobis distance from the mean gives the log likelihood of the shape, except for a linear multiplier and an additive constant,. and thus it serves as the basis for methods for Bayesian segmentation or statistical shape analysis.

See [12,48] for additional details of the discrete medial shape representation and its applications to medical image analysis.

References

1. Yan D, Lockman D, Brabbins D, Tyburski L, Martinez A. *An off-line strategy for constructing a patient-specific planning target volume in adaptive treatment process for prostate cancer*. Int J Radiat Oncol Biol 48(1):289-302, 2000.

2. Wong JR, Grimm L, Uematsu M, Oren R, Cheng CW, Merrick S, and Schiff P. *Image-guided radiotherapy for prostate cancer by CT-linear accelerator combination prostate movements and dosimetric considerations*. Int J Radiat Oncol Biol Phys 61: 561–569, 2005.

3. Lu W, Chen M-L, Olivera G, Ruchala K, and Mackie T. *Fast free-form deformable registration via calculus of variations*. Phys Med Biol 49:3067-3087, 2004.

4. Wang H, Dong l, Lii M, Lee A, De Crevoisier R, Mohan R, Cox J, Kuban D, and Cheung R. *Implementation and validation of a three-dimensional deformable registration algorithm for targeted prostate cancer radiotherapy*. Int J Rad Oncol Biol Phys 61:725-735, 2005.

5. Mohan R, Zhang X, Wang H, Kang Y, Wang X, Liu H, Ang K, Kuban D, and Dong L. *Use of deformed intensity distributions for on-line modification of image-guided IMRT to account for interfractional anatomic changes*. Int J Rad Oncol Biol Phys 61(4):1258-1266, 2005.

6. Foskey M, Davis B, Goyal L, Chang S, Chaney E, Strehl N, Tome S, Rosenman J, and Joshi S. *Large deformation 3D image registration in image-guided radiation therapy.* Phys Med Biol, 50: 5869-5892, 2005.

7. Freedman D, Radke R, Zhang T, Jeong Y, Lovelock M, and Chen G. *Model-based segmentation of medical imagery by matching distributions*, IEEE Transactions in Medical Imaging 24(3):281-292, 2005.

8. V. Pekar, T. McNutt, M. Kaus. *Automated model-based organ delineation for radiotherapy planning in prostatic region*. Int J Rad Oncol Biol Phys 60: 973-980, 2004.

9. Tobias Klinder, Cristian Lorenz, Jens von Berg, Sebastian P. M. Dries, Thomas Bülow, Jörn Ostermann. *Automated Model-Based Rib Cage Segmentation and Labeling in CT Images*. Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2007: 195-202

10. Grimson WE, Kikinis R, Jolesz FA, Black PM. *Image-guided surgery*. Sci Am. 1999 Jun;280(6):62-9.

11. M Styner, K Gorczowski, T Fletcher, JY Jeong, SM Pizer, G Gerig. *Statistics of Pose and Shape in Multiobject Complexes Using Principal Geodesic Analysis*, Medical Imaging and Augmented Reality 2006, LNCS 4091, pp 1-8.

12. Pizer SM, Fletcher PT, Joshi S, Gash G, Stough J, Thall A, Tracton G, and Chaney EL. *A Method and Software for Segmentation of Anatomic Object Ensembles by Deformable M-Reps*. Med Phys 32(5):1335-1345, 2005.

13. Timothy F. Cootes, A. Hill, Christopher J. Taylor, and J. Haslam. *The use of active shape models for locating structures in medical images*. Information Processing in Medical Imaging (IPMI), 1993, 33–47.

14. J. -P. Thirion, Image matching as a diffusion process: an analogy with Maxwell's demons, Medical Image AnalysisVolume 2, Issue 3, , September 1998, Pages 243-260.

15. Horn, B. K. and Schunck, B. G. 1980 *Determining Optical Flow*. Technical Report. UMI Order Number: AIM-572., Massachusetts Institute of Technology.

16. Gee JC, Reivich M, and Bajcsy R. Elastically deforming 3D atlas to match anatomical brain images. J Comput Assist Tomogr 17(2):225-236 (1993)

17. Christensen GE, Miller MI, and Vannier M. 3D *Brain mapping using a deformable neuroanatomy*. Phys Med Biol 39:609-618 (1994)

18. Joshi S, Davis B, Jomier M, Gerig G. *Unbiased diffeomorphic atlas construction for computational anatomy*. Neuroimage. 2004;23 Suppl 1:S151-60.

19. Brock KK, McShan DL, Ten Haken RK, Hollister SJ, Dawson LA, Balter JM. *Inclusion of organ deformation in dose calculations*, Med Phys. 2003 Mar;30(3):290-5.

20. Brock KK, Hollister SJ, Dawson LA, Balter JM. *Technical note: creating a four-dimensional model of the liver using finite element analysis*. Med Phys. 2002 Jul;29(7):1403-5.

21. Keith E. Muller, Yueh-Yun Chi, Jeongyoun Ahn, and J. S. Marron, *Limitations of High Dimension, Low Sample Size Principal Components for Gaussian Data*, under revision for resubmission.

22. IEEE Trans Med Imaging, Special Issue on Model-Based Analysis of Medical Images,18(10), 1999.

23. Haris, K., S.N. Efstratiadis, N. Maglaveras, Costas Pappas, J. Gourassas, and G. Louridas, *Model-Based Morphological Segmentation and Labeling of Coronary Angiograms*, IEEE Trans Med Imaging, Special Issue on Model-Based Analysis of Medical Images,18(10), 1999. pp. 1003-1015

24. McInerney, T. and D. Terzopoulos, *Topology Adaptive Deformable Surfaces for Medical Image Volume Segmentation*, IEEE Trans Med Imaging, Special Issue on Model-Based Analysis of Medical Images,18(10), 1999. pp. 840-850.

25. Hawkes DJ, Barratt D, Blackall JM, Chan C, Edwards PJ, Rhode K, Penney GP, McClelland J, Hill DL. *Tissue deformation and shape models in image-guided interventions: a discussion paper*. Med Image Anal. 2005 Apr;9(2):163-75.

26. Duda, R. O., Hart, P. E., and Stork, D. G. 2000 *Pattern Classification (2nd Edition)*. Wiley-Interscience.

27. Marsland S, Twining CJ. *Constructing diffeomorphic representations for the groupwise analysis of nonrigid registrations of medical images*. IEEE Trans Med Imaging. 2004 Aug;23(8):1006-20.

28. Cates J, Fletcher PT, Styner M, Shenton M, Whitaker R., *Shape modeling and analysis with entropy-based particle systems*, Inf Process Med Imaging. 2007;20:333-45.

29. Cootes, T., Taylor, C., Cooper, D., Graham, J.: *Active shape models – their training and application*. Computer Vision, Graphics, Image Processing: Image Understanding 1, 1994, 38-59

30. Brechbuhler, C., Gerig, G., Kubler, O.: *Parameterization of closed surfaces for 3-D shape description*. Computer Vision, Graphics, and Image Processing: Image Understanding 61, 1995, 195-170

31. Kelemen, A.; Szekely, G.; Gerig, G., "Elastic model-based segmentation of 3-D neuroradiological data sets ," *Medical Imaging, IEEE Transactions on* , vol.18, no.10, pp.828-839, Oct 1999

32. Yan D, Jaffray DA, Wong JW. *A model to accumulate fractionated dose in a deforming organ*. Int J Radiat Oncol Biol Phys. 1999 Jun 1;44(3):665-75.

33. Delingette H., Montagnat J., *Shape and Topology Constraints on Parametric Active Contours*, Computer Vision and Image Understanding, Volume 83, Number 2, August 2001 , pp. 140-171(32)

34. Kass M, Witkin A, Terzopoulos D. *Snakes: Active Contour Models*. Int J Comp Vision 1:321-331 (1987)

35. Fletcher, P.T., Conglin Lu, Pizer, S.M., Sarang Joshi, *Principal geodesic analysis for the study of nonlinear statistics of shape*. IEEE Transactions on Medical Imaging 23 (2004) 995-1005

36 Warfield, S.K., Zou, K.H., Wells, W.M., *Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation*, IEEE Trans on Med Imaging, 2004, 23:7, 903- 921.

37. F.L. Bookstein, *Principal warps: Thin-plate splines and the decomposition of deformations*, IEEE Transactions on Pattern Analysis and Machine Intelligence 1, 1989, no. 6.

38. Bookstein, F.L.: *Morphometric Tools for Landmark Data*. Cambridge University Press, 1991.

39. Damon, J., *Geometry and Medial Structure* in Medial Representations: Mathematics, Algorithms and Applications, Siddiqi, K., Pizer, S., eds., 2007.

40. Q. Han, D. Merck, J. Levy, C. Villarruel, E. Chaney, and S. M. Pizer, *Proper Training*, Presented at Information Processing in Medical Imaging (IPMI), Nico Karssemeijer and Boudewijn Lelieveldt, eds., 2007.

41. Danielsson, P.E.: *Euclidean distance mapping*. Computer Graphics and Image Processing, 1980, 227-248.

42. William H. Press, William T. Vetterling, Saul A. Teukolsky, and Brian P. Flannery, *Numerical recipes in c++: The art of scientific computing*, 2002.

43. M Söhn, M Birkner, D Yan and M Alber. *Modelling individual geometric variation based on dominant eigenmodes of organ deformation: implementation and evaluation*. Physics in Medicine and Biology, Vol. 50, 2005.

44. Q Han, SM Pizer, D Merck, S Joshi, and J-Y Jeong, *Multi-figure anatomical objects for shape statistics*, Presented at Information Processing in Medical Imaging (IPMI), 2005.

45. Chaney, E., Pizer, S., Joshi, S., Broadhurst, R., Fletcher, T., Gash, G., Han, Q., Jeong, J., Lu, C., Merck, D., Stough, J., Tracton, G., J. Bechtel, M., Rosenman, J., Chi, Y., Muller, K.: *Automatic male pelvis segmentation from CT images via statistically trained multi-object deformable m-rep models*. Presented at American Society for Therapeutic Radiology and Oncology (ASTRO), 2004.

46. M. Rao, J. Stough, Y. Chi, K. Muller, G. Tracton, S. Pizer, E. Chaney, *Comparison of human and automatic segmentations of kidneys from CT images*. Int J Radiat Oncol Biol Phys, 61:3, 954-960, 2005.

47. Broadhurst, R., Stough, J., Pizer, S., Chaney, E.: *A statistical appearance model based on intensity quantile histograms*. Proceedings of the 2006 IEEE International Symposium on Biomedical Imaging.

48. Pizer, S., Han, Q., Joshi, S., Fletcher, P.T., Yushkevish, P., Thall, A., *Synthesis, Deformation, and Statistics of 3D Objects via M-reps* in Medial Representations: Mathematics, Algorithms and Applications, Siddiqi, K., Pizer, S., eds., 2007.

49. R. Pilgram, P. T. Fletcher, S. M. Pizer, O. Pachinger, and R. Schubert. *Common shape model and inter-individual variations of the heart using medial representation: a pilot study*. Technical report, Institute for Medical Knowledge Representation and Visualization, University for Health Informatics and Technology, Tyrol, Austria, 2003.

50. Wendy S. Vanden Berg-Foels, Steven J. Schwager, Rory J. Todhunter, Anthony P. Reeves. *Femoral head secondary center of ossification shape development in a canine model of hip dysplasia*, in submission.

51. Blum H., *A transformation for extracting new descriptors of shape*. In W. Wathen-Dunn (Ed.) Models for the Perception of Speech and Visual Form. MIT Press, Cambridge MA: 363-380. (1967)

52. Han, Q., Damon, J.: *Medial interpolation using s-rad*. Mathematical Methods in Biomedical Image Analysis (MMBIA), 2006.

53. E. Catmull, J. Clark, *Recursively generated b-spline surfaces on arbitrary topological meshes*, Computer Aided Design 10, 1978, 183–188.

54. Crouch JR, Pizer SM, Chaney EL, Zaider M. Medially Based Meshing with Finite Element Analysis of Prostate Deformation.: Medical Image Computing and Computer-Assisted Intervention, Proceedings, Part I, Randy E. Ellis, Terry M. Peters (Eds.) 108-115, Springer, NY. (2003)

55. Koenderink, J. J. 1990 *Solid Shape*. MIT Press.