Computer Vision Methods for Image Guided Surgery

Eric Grimson, Michael Leventon,
Olivier Faugeras, William Wells
Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge MA, USA 02139

Abstract

Image guided surgery (IGS) systems assist a surgeon in executing procedures with minimal time, invasiveness and collateral damage. A key element in IGS is providing the surgeon with a detailed model of the patient's anatomy, which he can visualize for planning, navigation and evaluation of delivery. This requires computer vision algorithms to extract geometric models from medical imagery, to register those models to one another and to the position of the patient in the operating room, and to adapt those models to reflect intra-operative changes in anatomy. Thus, a key aspect of IGS is creating accurate, detailed, patient-specific models from medical imagery. In this paper, we briefly outline an approach to image guided surgery, and present an example of current methods for model building.

1 Introduction

Recent developments in computer vision and robotics are changing the practice of many surgical procedures. Image guided surgical methods are providing a surgeon with the ability to visualize internal structures and their geometric relationships, often in alignment with live imagery or direct views of the patient. Such visualization methods allow a surgeon to plan minimally invasive procedures. Tracking methods further allow a surgeon to see the actual position of his instruments, and their physical relationship to critical or hidden structures during a procedure, thus leading to faster, safer and more effective minimally invasive procedures.

Our approach to Image Guided Surgery consists of the following stages:

- A set of medical images (such as MRI or CT) is acquired of the relevant portion of the patient's anatomy (Figure 1).
- These images are segmented into distinct anatomical structures, yielding a 3D patient-specific model (Figure 1).
- The model is then registered to the actual position of the patient on the operating table. Surgical instruments are tracked relative to the patient and the model, allowing the surgeon to effectively execute procedures while avoiding hidden, critical structures (Figure 1).

BMVC 2000 doi:10.5244/C.14.1

In this paper, we briefly outline current approaches to each of these stages, then examine in more detail the first stage -- segmentation of medical scans into patient-specific models that can be used for surgical planning, guidance, and navigation.

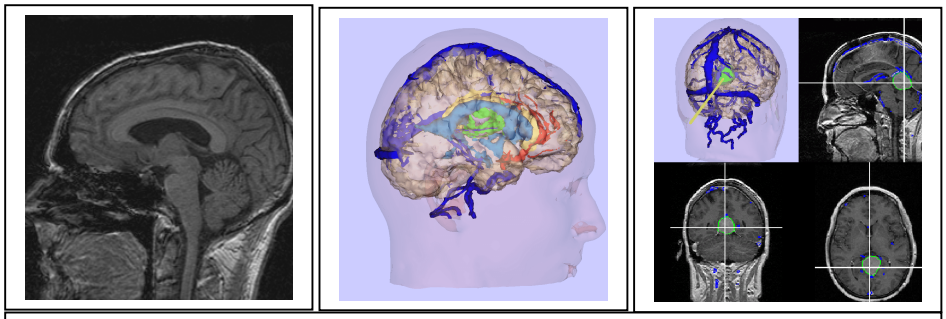

Figure 1: Standard imagery, such as MRI (l, highlights different tissue responses. These images are converted to patient-specific modesl of anatomy. They can be used both for planning (center) and to support navigation by tracking surgical instruments (right).

1.1 Segmentation

The first stage converts standard medical imagery into information that more directly reflects patient anatomy. As detailed in section 2, we use several algorithmic methods to label individual elements of a volumetric scan by tissue type, and to collect those labelled voxels into connected structures.

1.2 Registration

Given segmented models of the patient (see section 2), we want to use that information to assist the surgeon. One aspect of this is to allow the surgeon to visualize structures and their relationships, in order to support plan reliable procedures. To this end, it is often useful to augment the structure models created by segmenting MRI data, with functional and other information from other imagery (e.g., fMRI, PET, SPECT). This requires registration of two or more different modalities. Since there is often no set of corresponding landmarks between the scans, we have relied on a registration method that finds the best alignment between image volumes, based on the alignment of all the data in the volumes. In particular, our group has developed registration methods that maximize the Mutual Information between data sets, and have demonstrated that such methods can robustly align a wide range of medical imagery [24].

Once such augmented models are created, we also need to register them to the actual position of the patient. To do this, we use a surface-matching algorithm. 3d data points from the skin surface of the patient are acquired using a tracked probe (typical methods involve locating synchronized LED's on the probe in a set of three cameras and using triangulation to locate the probe, or using electromagnetic coils tracking small metal balls attached to the probe, and similarly using triangulation to localize the probe). These data points are then matched to the skin surface of the reconstructed patient model to find the best fit, and thus to align the model with the actual patient position [11, 12].

1.3 Visualization

Once the model and patient coordinate frames have been aligned, we can support visualization of the model. In particular by calibrating a camera looking at the patient over the surgeon's shoulder we can acquire live imagery of the surgical site. Since we also know the position of the model with respect to the camera, we can render an image of the internal structures of the model from the perspective of that camera. By combining the two through a video mixer, we can generate augmented reality visualizations, which enable the surgeon to see the positions of nearby structures that are otherwise hidden from view below the currently exposed surgical surface.

1.4 Navigation

Finally, by tracking the position of any surgical instrument in the operating site, one can relate the position of the tip of the instrument to the corresponding point on the segmented model. By displaying this information on a monitor, the surgeon can rapidly navigate to desired target sites while avoid critical structures.

2 Segmentation

The first stage of this pipeline is key. One must extract relevant information from the medical imagery, in order to most effectively present it to the surgeon. This becomes a problem of segmentation -- labelling each voxel of the image with the associated tissue type, and then agglomerating adjacent voxels with common labels into connected structures. In this section, we describe several current techniques.

2.1 Statistical labelling of images

Our base method uses statistical properties of tissue response in MRI imagery to classify tissue type associated with each voxel [14,15,25]. The basis of the method is straightforward. Suppose one could identify a small set of voxels of each desired tissue type in the imagery, for example by using known anatomical landmarks. Record the intensity in the MRI data at such points. This gives us a set of samples of the intensity associated with each tissue type. In principle, a straightforward application of nearest neighbor labelling can then be used to identify the tissue type associated with all other points in the imagery.

Unfortunately, MRI imagery contains significant gain artefacts, which act as a nonlinear multiplier on the observed signal. This causes simple nearest-neighbor schemes to fail. As a consequence, our group has developed [14,15,25] methods that automatically account for the gain artifact, while correctly identifying the tissue type, using the Expectation/Maximization algorithm to iteratively solve for the gain field and the tissue labelling. The method proceeds by finding the best labelling of the voxels, given the current estimate of the gain field, then uses those labels to find the gain field estimate that maximizes the expected value of the field. This cycle iterates to convergence and provides accurate estimates of both the gain field and the tissue labels.

The E/M method assumes a stationary prior on the intensities associated with different tissue types (which can be found by measuring relative volumes of labelled voxels in training images). Better use of anatomical knowledge can be made, however. One can use Markov Random Field methods to impose local continuity on the labelling, and one can use rough knowledge of the distribution of tissue types as a function of distance from known landmarks to create non-stationary prior probabilities [14,15]. Such methods essentially incorporate coarse atlas information to guide segmentation.

2.2 Using anatomical knowledge

Because segmentation often requires distinguishing ambiguous information, having prior information about the expected shape of a structure can significantly aid in the segmentation process. For example, one can find corresponding points across training images, and use that to construct a statistical model of shape variation from the point positions. In this case, the best match of the model to the image is found by searching over the model parameters [8]. A second method [23] adds global shape information into the segmentation task by using an elliptic Fourier decomposition of the boundary and placing a Gaussian prior on the Fourier coefficients.

Our approach to object segmentation extends geodesic active contours [4,17] by incorporating shape information into the evolution process. We first compute a statistical shape model over a training set of curves. To segment a structure from an image, we evolve an active contour both locally, based on image gradients and curvature, and globally to a maximum likelihood estimate of shape and pose.

2.3 Probability Distribution on Shapes

To incorporate shape information into the process of segmenting an object in an image, we consider a probabilistic approach, and compute a prior on shape variation given a set of training instances. To build the shape model, we choose a representation of curves, and then define a probability density function over the parameters of the representation.

2.3.1 Curve Representation

Each curve in the training dataset is embedded as the zero level set of a higher dimensional surface, u, whose height is sampled at regular intervals (say N^d samples, where d is the number of dimensions). The embedding function chosen is the signed distance function [22], where each sample encodes the distance to the nearest point on the curve, with negative values inside the curve (Figure 2). Each such surface (distance map) can be considered a point in a high dimensional space ($u \in \Re^{N^d}$). The training set consists of a set of surfaces $\{u_1, u_2, ..., u_n\}$. Our goal is to build a shape model over this distribution of surfaces.

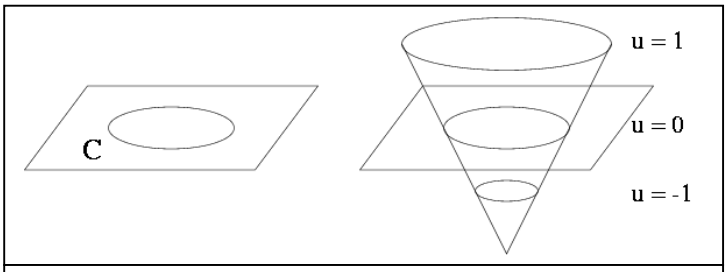

Figure 2: Level sets of an embedding function u for a closed curve C in the plane.

The mean surface, μ, is computed by taking the mean of the signed distance functions, $\mu = \sum_{1}^{n} u_i$. The variance in shape is computed using Principal Component Analysis (PCA). The mean shape is subtracted from each sample surface to create an mean-offset map, each of which is placed as a column vector in an $N^d xn$ dimensional matrix M. Using Singular Value Decomposition (SVD), the covariance matrix $M^T M$ is decomposed as:

$$U\Sigma^T U = M^T M$$

where U is a matrix whose column vectors represent the set of orthogonal modes of shape variation and Σ is a diagonal matrix of corresponding singular values. An estimate of a novel shape, u, of the same class of object can be represented by k principal components in a k-dimensional vector of coefficients, α:

$$\alpha = U_k^T (u - \mu)$$

where U_k is a matrix consisting of the first k columns of U that is used to project a surface into the eigen-space. Given the coefficients α, an estimate of the shape is reconstructed from U_k and μ:

$$u' = U_k \alpha + \mu$$

Note that in general u' will not be a true distance function, since convex linear combinations of distance maps do not produce distance maps. However, the surfaces generally still have advantageous properties of smoothness, local dependence, and zero level sets consistent with the combination of original curves.

Under the assumption of a Gaussian distribution of shape represented by α, we can compute the probability of a certain curve as:

$$P(\alpha) = \frac{1}{\sqrt{(2\pi)^k \,|\Sigma_k|}} \exp\left(-\frac{1}{2}\alpha^T \Sigma_k^{-1} \alpha\right)$$

where Σ_k contains the first k rows and columns of Σ.

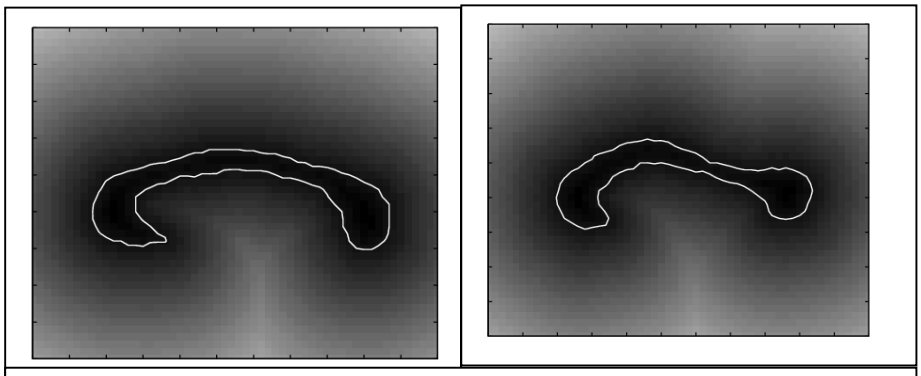

Figure 3: Outlines of the corpus callosum of 2 out of 51 patients in the training set embedded as the zero level set of a higher dimensional surface, the signed distance function.

Figure 3 shows two of the 51 training curves used to define the shape models of the corpus callosum. The original segmentations of the images are overlaid as curves on the signed-distance map. Before computing and combining the distance maps of these training shapes, the curves were aligned using centroids and second moments to approximate the correspondence. Figure 4 illustrates zero level sets corresponding to the means and two primary modes of variance of the shape distribution of the corpus callosum. Note that the mean shape and primary modes appear to be reasonable representative shapes of the class of objects being learned. The first mode caputres size, while the second mode captures the degree of curvature of the corpus. The third mode appears to represent the shifting of the bulk of the corpus from front to back.

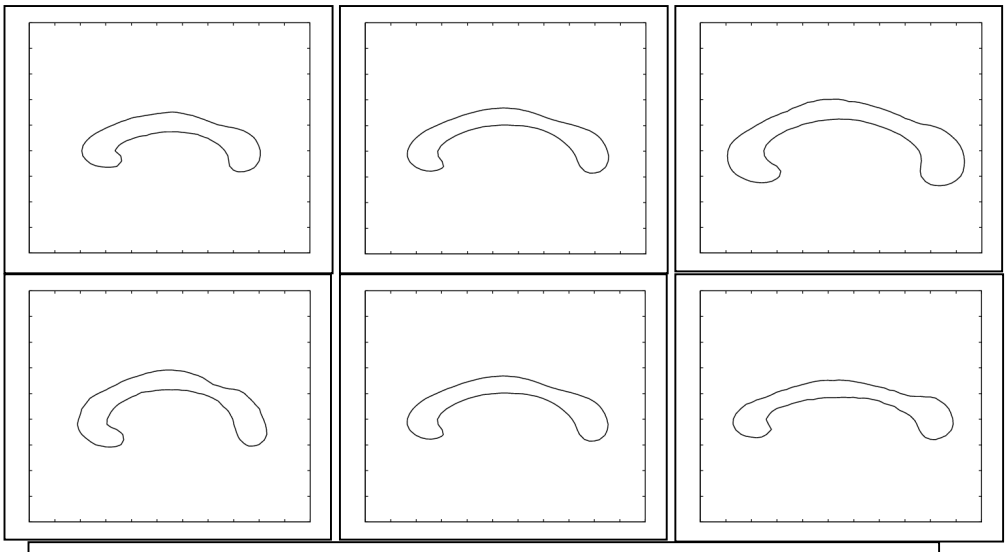

Figure 4: The first two primary modes of variance of the corpus callosum training dataset.

2.3.2 The Correspondence Problem

When measuring shape variance of a certain part of an object across a population, it is important to compare like parts of the object. For example, when looking at variances in the shape of the vertebrae, if two training examples are misaligned and a process of one is overlapping a notch of the other, then the model will not be capturing the appropriate anatomical shape variance seen across vertebrae. One solution to the correspondence problem is to explicitly generate all point-wise correspondences to ensure that comparisons are done consistently, although this is difficult to automate and is manually tedious (see [8]). Another approach to correspondence is to roughly align the training data before performing the comparison and variance calculation. A rough alignment will not match every part of each training instance perfectly, so one must consider the robustness of the representation to misalignment.

Using the signed distance map as the representation of shape provides tolerance to slight misalignment of object features, in the attempt to avoid having to solve the general correspondence problem. In the examples presented here, the rough rigid alignment of the training instances resulted in the model capturing the shape variances inherent in the population due to the dependence of nearby pixels in the shape representation.

2.4 Shape Priors and Geodesic Active Contours

Given a curve representation (the k-vector α) and a probability distribution on α, prior shape information can be folded into the segmentation process. This section describes adding a term to the level set evolution equation to pull the surface in the direction of the maximum likelihood shape and position of the final segmentation.

2.4.1 Geodesic Active Contours for Object Segmentation

The snake methodology defines an energy function $E(C)$ over a curve C as the sum of an internal and external energy of the curve, and evolves the curve to minimize the energy [16].

$$E(C) = \beta \int |C'(q)|^2 \, dq - \int |\nabla I(C(q))| \, dq$$

In [4], Caselles, *et al.* derive the equivalence of geodesic active contours to the traditional energy-based active contours (snakes) framework by first reducing the minimization problem to the following form:

$$\min_{C(q)} \int g(|\nabla I(C(q))|) \, |C'(q)| \, dq$$

where g is a function of the image gradient (usually of the form $1/(1+|\nabla I|^2)$. Using Euler-Lagrange, the following curve evolution equation is derived [4]

$$\frac{\partial C(t)}{\partial t} = g\kappa N - (\nabla g \cdot N)N$$

where κ is the curvature and N is the unit normal. By defining an embedding function u of the curve C, the update equation for a higher dimensional surface is [4]:

$$\frac{\partial u}{\partial t} = g(c + \kappa)|\nabla u| + \nabla u \cdot \nabla g$$

where c is an image-dependent balloon force added to force the contour to flow outward [9, 19]. In this level set framework, the surface, u, evolves at every point perpendicular to the level sets as a function of the curvature at that point and the image gradient.

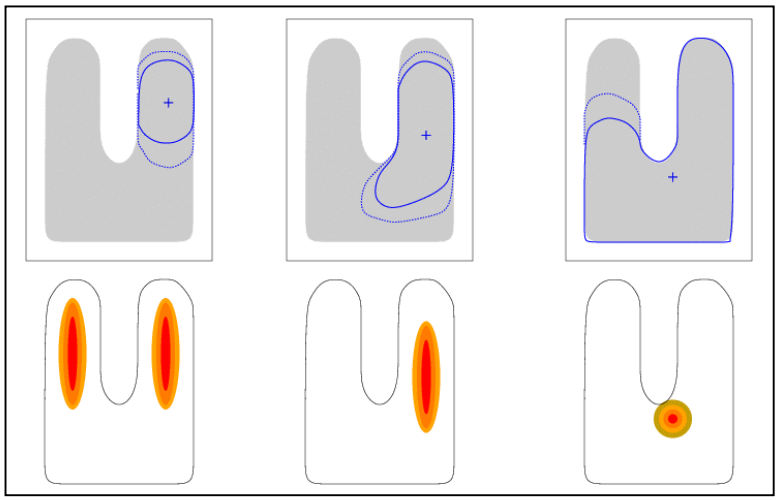

Figure 5: Three steps in the evolution process. The evolving curve is shown in solid blue superimposed on the image (top row). The curve is matched to the expected curve to obtain a PDF over pose (bottom row). The next evolution step (based on pose and shape) is shown as the dotted blue line.

2.4.2 Estimation of Pose and Shape

In addition to evolving the level set based on the curvature and the image term, we include a term that incorporates information about the shape of the object being segmented. To add such a global shape force to the evolution, the pose of the evolving curve with respect to the shape model must be known (see Figure 5). Without an estimate of the pose, the shape model cannot adequately constrain or direct the evolution. Therefore, at each step of the curve evolution, we seek to estimate the shape parameters, α, and the rigid pose parameters, p, of the *final* curve using a maximum likelihood approach.

$$\langle \alpha_{ML}, p_{ML} \rangle = \arg\max_{\alpha,p} P(\alpha, p \mid u, \nabla I)$$

In this equation, u is the evolving surface at some point in time, whose zero level set is the curve that is segmenting the object. The term ∇I is the gradient of the image containing the object to be segmented. By our definition of shape and pose, the final segmentation curve is completely determined by α and p. Let u^* be the estimated final curve, which can be computed from α and p. Therefore, we also have

$$u_{ML}^* = \arg\max_{u^*} P(u^* \mid u, \nabla I)$$

To compute the maximum likelihood final curve, we expand the terms using Bayes' Rule.

$$P(\alpha, p \mid u, \nabla I) = \frac{P(u \mid \alpha, p)P(\nabla I \mid \alpha, p, u)P(\alpha)P(p)}{P(u, \nabla I)}$$

We proceed by defining each term of this equation in turn. We discard the normalization term in the denominator as it does not depend on shape or pose. More details are available in [18].

The first term computes the probability of a certain evolving curve, u, given the shape and pose of the final curve, u^* (or $\langle \alpha, p \rangle$). Notice that this term does not include any image information whatsoever. We model this term as a Laplacian density function over $V_{outside}$, the volume of the curve u that lies outside the curve u^*.

$$P(u \mid \alpha, p) = \exp(-V_{outside})$$

This term assumes that any curve u lying inside u^* is equally likely. Since the initial curve can be located at any point inside the object and the curve can evolve along any path, we do not favor any such curve.

The second term computes the probability of seeing certain image gradients given the current and final curves. Let $h(u^*)$ be the best fit Gaussian to the samples $(u^*, |\nabla I|)$. We model the gradient probability term as a Laplacian of the goodness of fit of the Gaussian.

$$P(\nabla I \mid u^*, u) = \exp(-\frac{1}{2}\alpha^T \Sigma_k{}^{-1} \alpha)$$

The last two terms are based on our prior models, as described earlier. Our shape prior is a Gaussian model over the shape parameters, α, with shape variance Σ_k.

$$P(\alpha) = \frac{1}{\sqrt{(2\pi)^k \mid \Sigma_k \mid}} \exp(-\frac{1}{2}\alpha^T \Sigma_k{}^{-1} \alpha)$$

In our current framework, we seek to segment one object from an image, and do not retain prior information on the likelihood of the object appearing in a certain location. Thus, we simply assume a uniform distribution over pose parameters, which can include any type of transformation, depending on application.

$$P(p) = U(-\infty, \infty).$$

These terms define the maximum likelihood estimator of shape and pose, which estimates the final curve or segmentation of the object. For efficiency, these quantities are computed only in a narrow band around the zero level set of the evolving surface, and the ML pose and shape are re-estimated at each evolution step using simple gradient ascent on the log likelihood function in the Bayes rule expansion. While each ascent may yield a local maximum, the continuous re-estimation of these parameters as the surface evolves generally results in convergence on the desired maximum. Next, we incorporate this information into the update equation commonly used in level set segmentation.

2.4.3 Evolving the Surface

Initially, the surface, u, is assumed to be defined by at least one point that lies inside the object to be segmented. Given the surface at time t, we seek to compute an evolution step that brings the curve closer to the correct final segmentation based on local gradient and global shape information.

The level set update expression provides a means of evolving the surface u over time towards the solution to the original curve-minimization problem. Therefore, the shape of the surface at time $t+1$ can be computed from $u(t)$ by:

$$u(t+1) = u(t) + \lambda_1(g(c+\kappa)\,|\nabla u(t)\,| + \nabla u(t)\cdot\nabla g)$$

where λ_1 is a parameter defining the update step size.

By estimating the final surface u^* at a given time t, we can also evolve the surface in the direction of the maximum likelihood final surface:

$$u(t+1) = u(t) + \lambda_2(u^*(t) - u(t))$$

where $\lambda_2 \in [0,1]$ is the linear coefficient that determines how much to trust the maximum likelihood estimate. Combining these equations yields the final expression for computing the surface at the next step.

$$u(t+1) = u(t) + \lambda_1(g(c+\kappa)\,|\nabla u(t)\,| + \nabla u(t)\cdot\nabla g) + \lambda_2(u^*(t) - u(t))$$

The two parameters λ and λ_2 are used to balance the influence of the shape model and the gradient-curvature model. The parameters also determine the overall step size of the evolution. The trade-off between shape and image depends on how much faith one has in the shape model and the imagery for a given application

2.5 Results

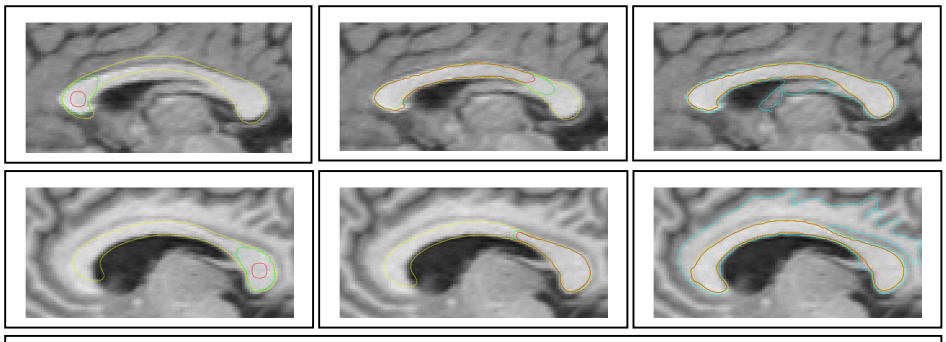

Figure 6: Three steps in the segmentation of two different corpora callosa. The last image in each case shows the final segmentation in red. The cyan contour is the result of the standard evolution without the shape influence.

Segmentation experiments were performed on 2D slices of MR images of the corpus callosum (Figure 6). The corpus callosum training set consisted of 49 examples like

those in Figure 3. The segmentations of two corpora callosa are shown in Figure 6. Notice that while the ML shape estimator is initially incorrect, as the curve evolves, the pose and shape parameters converge on the boundary. The segmentations of the femur slices and the corpora all converged in under a minute on a 550 MHz Pentium III.

K	Corpus 1	Corpus 2
95%	1.3 mm	1.5mm
99%	1.6 mm	2.0 mm

Table 1: Partial Hausdorff distance between our segmentation and the manually segmented ground truth.

To validate the segmentation results, we compute the undirected partial Hausdorff distance [13] between the boundary of the computed segmentation and the boundary of the manually segmented ground truth. The directed partial Hausdorff distance over two point sets A and B is defined as

$$h_K(A,B) = K^{th} \min_{a\in A, b\in B} \| a-b \|$$

where K is a quantile of the maximum distance. The undirected partial Hausdorff distance is defined as $H_K)A,B) = \max(h_K(A,B), h_K(B,A))$. The results for the corpora shown in Table 1 indicate that virtually all the boundary points lie within one or two voxels of the manual segmentation.

Acknowledgements
Support for our research was provided in part by the NSF under grant no. IIS-9610249, and in part by NSF ERC (Johns Hopkins University agreement) 8810-274.

References

[1] L. Ambrosio, and H.M. Soner, Level set approach to mean curvature flow in arbitrary codimension. Journal of Differential Geometry **43** (1996): 693—737.
[2] V. Caselles, F. Catte, T. Coll, and F. Dibos, A geometric model for active contours. Numerische Mathematik **66** (1993): 1—31.
[3] V. Caselles, J.M. Morel, G. Sapiro, and A.Tannenbaum, Introduction to special issue on partial differential equations and geometry-driven diffusion in image processing and analysis. IEEE Transactions on Image Processing **7(3)** (1998): 269—273.
[4] V. Caselles, R. Kimmel, and G. Sapiro: Geodesic active contours. Int'l Journal of Computer Vision **22(1)** (1997): 61—79.
[5] Y.G. Chen, Y. Giga, and S. Goto, Uniqueness and existence of viscosity solutions of generalized mean curvature flow equations. Journal of Differential Geometry **33** (1991): 749—786.
[6] G. Christensen, R. Rabbitt, and M. Miller. Deformable templates using large deformation kinematics. IEEE Trans. Image Processing **5(10)** (1996): 1435--1447.
[7] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models - their training and application. Computer Vision and Image Understanding, 1995.
[8] T. Cootes, C. Beeston, G. Edwards, and C. Taylor Unified Framework for Atlas Matching Using Active Appearance Models, Information Processing in Medical Imaging, 1999.

[9] L. Cohen, On active contour models and balloons. CVGIP: Image Understanding **53(2)** (1991): 211--218.
[10] L.C. Evans and J. Spruck: Motion of level sets by mean curvature: I. Journal of Differential Geometry **33** (1991): 635—681.
[11] W.E.L. Grimson, G.J. Ettinger, T. Kapur, M.E. Leventon, W.M. Wells III, and R. Kikinis, Utilizing Segmented MRI Data in Image-Guided Surgery. Int'l Journal Pattern Recognition and Artificial Intelligence , (1996)
[12] W.E.L. Grimson, T. Lozano-Perez, W.M. Wells III, G.J. Ettinger, S.J. White and R. Kikinis, An Automatic Registration Method for Frameless Stereotaxy, Image Guided Surgery, and Enhanced Reality Visualization, IEEE Trans. Medical Imaging, **15(2),** April 1996, 129--140.
[13] D. Huttenlocher, G. Klanderman, and W. Rucklidge, Comparing images using the Hausdorff distance, IEEE Trans PAMI, **15** (1993):850-863.
[14]T. Kapur, E. Grimson, R. Kikinis, W. Wells, Enhanced spatial priors for segmentation of magnetic resonance imagery', Medical Image Computation and Computer Assisted Interventions, Boston, October 1998.
[15] T. Kapur, E. Grimson, W. Wells, R. Kikinis, Segmentation of Brain Tissue from Magnetic Resonance Images', Medical Image Analysis **1(2)** 1997:109--128.
[16] M. Kass, A. Witkin, and D. Terzopoulos: Snakes: active contour models. Int'l Journal Computer Vision **1(4)** (1988): 321—331.
[17] A. Kichenassamy, A. Kumar, P. Olver, A,.Tannenbaum,. and A. Yezzi: Gradient flows and geometric active contour models. Proc. IEEE Int'l Conf. Computer Vision (1995) 810—815.
[18] M. Leventon, E. Grimson, O. Faugeras. Statistical Shape Influence in Geodesic Active Contours. CVPR 2000.
[19] L. Lorigo, O. Faugeras, W.E.L. Grimson, R. Keriven, R. Kikinis. Segmentation of Bone in Clinical Knee MRI Using Texture-Based Geodesic Active Contours. MICCAI, 1998.
[20] L. Lorigo, O. Faugeras, E. Grimson, R. Keriven, R. Kikinis, and C.-F. Westin, Co-dimension 2 geodesic active contours for MRA segmentation. Int'l Conf. Information Procession Medical Imaging. Lect. Notes Comp. Sci. **1613** (1999) 126—139.
[21] S.J. Ruuth, B. Merriman, J. Xin, J., and S. Osher, Diffusion-generated motion by mean curvature for filaments. UCLA Computational and Applied Mathematics Report 98--47 (1998).
[22] J. Sethian. Level Set Methods. Cambridge University Press (1996).
[23] L. Staib and J. Duncan. Boundary Finding with Parametrically Deformable Models PAMI **14(11)** (1992): 1061--1075.
[24] P. Viola and W. Wells. Alignment by maximization of mutual information. IJCV **24(2)** (1997):137—154.
[25] W.M. Wells, III, W.E.L. Grimson, R. Kikinis, and F.A. Jolesz, Adaptive Segmentation of MRI data, IEEE Trans. Medical Imaging **15(4)** (1996): 429—442.
[26] X. Zeng, L. Staib, R. Schultz and J. Duncan, Segmentation and measurement of the cortex from 3D MR images. Proc. Int'l Conf. Medical Image Computing Computer-Assisted Intervention. Lect. Notes Comp. Sci. 1496 (1998) 519—530.