IPMI 1997 - LNCS 1230: 127-140

Localization and Segmentation of Medical Image Objects using Deformable Shape Loci

Daniel S. Fritsch[1,2], Stephen M. Pizer[1,2,3,4], Liyun Yu[3], Valen Johnson[5], and Edward L. Chaney[1]

Medical Image Display and Analysis Group
The University of North Carolina at Chapel Hill

Departments of [1]Radiation Oncology, [2]Biomedical Engineering, [3]Computer Science, and [4]Radiology
The University of North Carolina at Chapel Hill

[5]Institute of Statistics and Decision Sciences
Duke University

ABSTRACT

Robust localization and segmentation of normal anatomical objects in medical images require (1) methods for creating descriptive object models that adequately capture object shape and expected shape variation across a population, (2) methods for combining such shape models with unclassified image data, and (3) means for localizing and extracting corresponding objects from the image data using the model. A Bayesian approach is well suited as a general analytic framework for such a process; object shape models are associated with the prior, image information is associated with the likelihood function, and the posterior provides a means for combining models with images in a way that makes it possible to localize and segment normal figural shapes. Prior models that include multiscale medial and boundary analysis are well matched with a Bayesian approach since such models directly capture i) *figural shape,* ii) *inter-figural shape* relationships, and iii) *boundary shape* and location relative to figural shape. Described in this paper is such an approach to model-based segmentation, called *deformable shape loci* (DSLs), that has been successfully applied to 2D MR slices of the brain ventricle and CT slices of abdominal organs. The method combines the model and image data by warping the model to optimize an objective function measuring both the conformation of the warped model to the image data and the preservation of local neighbor relationships in the model. Large scale medial information stabilizes the localization of the object boundary region in the presence of image disturbances such as noise, blurring, and poor contrast resolution. Methods for forming the model on which the prior is based and for optimizing the posterior are described.

Keywords: Deformable contours, multiscale analysis, medial analysis, segmentation

Correspondence to:
Daniel S. Fritsch, Ph.D.
Assistant Professor
Departments of Radiation Oncology and Biomedical Engineering
The University of North Carolina at Chapel Hill
Chapel Hill, NC 27599-7512
Phone: (919) 966-7704
E-mail: fritsch@radonc.unc.edu

1. INTRODUCTION

A variety of model-directed methods have been developed to address the increasing clinical need for more reliable and automatic means for segmenting objects in medical images. One popular class of models incorporates some form of *a priori* information about characteristics of the object boundary, requiring (for example) that the extracted boundary be sufficiently smooth [e.g., Kass et al., 1987]. Another class of techniques attempts to model *a priori* object information in the form of some kind of local consistency constraint(s) on the image intensities, requiring (for example) that the local intensities within a segmented object vary smoothly from pixel to pixel [Leahy et al., 1989]. More powerful methods exhibit some combination of the two approaches [e.g., Cootes et al., 1993; Chakraborty et al., 1995], wherein both boundary-based and regional intensity properties of the object are taken into account during the modeling process. In all cases, the model serves to constrain the final segmentation, which is driven by certain measurements obtained from the unclassified image data.

A shortcoming, we believe, of many such methods is their lack of use of important, global properties of object shape and their failure to recognize the importance of measurement scale during model construction and during the segmentation process wherein the model is combined with image information. For example, models based on only local boundary information are unable to take advantage of higher order figural information that manifests itself at multiple, object-relevant scales. Without such figural information, such methods often require that the model be initialized very close to the actual object boundary to avoid convergence to undesired local extrema and are notoriously sensitive to image disturbances, such as intensity noise and blurring.

As we demonstrate in this paper, a more complete representation of shape — one that recognizes the object and its component figures as a unit — can serve as a powerful basis for automatic *localization*[1] and *segmentation* of anatomical objects in medical images. In building shape-based models, which will ultimately be used to segment similar objects, we have considered three aspects of object shape we believe are necessary and sufficient for these two processes: a) figural shape, b) inter-figural shape relations (the relative position, size, and orientation of figures that comprise an object), and c) boundary shape. Measures of medial strength called medialness [Fritsch, 1993; Morse, 1993; Pizer, 1996] capture figural shape directly from the image intensities, and do so in a way that exhibits great insensitivity to image disturbances such as intensity and boundary noise, and blurring [Morse, 1996]. In addition, medial loci extracted using medialness permit the specification of inter-figural relations and stably allow for determination of figural boundaries [McAuliffe, 1996].

In this paper we report on a method of modeling such aspects of object shape in order to generate prior shape probability distributions which can then be applied, using a formal Bayesian framework, to an unclassified image to robustly locate and segment similar objects. This method has in common many of the attractive features of other deformable loci (active contour) methods, but introduces and integrates several particular strengths. Some of these strengths, which are elaborated upon in later sections, include the following:

1. A rich prior model of object shape based on the formality of Markov random fields and related statistical methods.
2. A means for incorporating more stable, object-related information into the model fitting process.
3. Computational efficiency via use of local neighbor relations.
4. Means for recognizing local object shape abnormalities.

[1] By the term localization, we mean first finding the approximate position, orientation, and size of a modeled object (and its component figures) in an image and then using the model shape prior to approximate its shape properties.

1.1 Bayesian Approaches to Model-Based Image Segmentation

Many of the model-based approaches to image segmentation may be cast in a Bayesian framework, wherein the model prescribes a prior distribution on allowable shape deformations and wherein the likelihood function describes how well measurements derived from the image accord with a given geometric state of the model. Let Θ be a parameter vector describing a particular object model configuration and let I be an image known to contain an instance of that object. Bayes rule allows us, given a particular image I, to relate the prior probability (a model of what we expect to see) to the *a posteriori* probability (what the image data supports given the model configuration) through the equation

$$p(\Theta|I) = \frac{p(I|\Theta)p(\Theta)}{p(I)} \tag{1}$$

Where $p(\Theta|I)$ is the *a posteriori* probability (posterior), $p(I|\Theta)$ is the state-conditional probability density function for I (the likelihood function), $p(\Theta)$ is the *a priori* probability (prior), and $p(I)$ is a constant with respect to Θ. In Bayesian, model-based image analysis the prior term assigns a probability to a given configuration of the model and, when combined with the likelihood function, permits one to perform recognition and/or segmentation through maximization of the posterior. Taking the logarithm on both sides of equation (1) yields

$$\log p(\Theta|I) = Z + \log p(I|\Theta) + \log p(\Theta) \tag{2}$$

where the constant $Z = -\log p(I)$. Hence, maximizing the posterior probability may be accomplished through maximization of the sum of the log likelihood and log prior terms. In such a Bayesian framework, the problem becomes one of how to (a) model and encode object shape information in the prior term in a way that adequately captures normal object shape and its variabilities and that (b) permits the model to be combined with the image (via the likelihood term) through appropriate choice of image information consistent with the model representation.

1.2 Placing Active Contours within a Bayesian Framework

In the classical papers on deformable contours, or active contours (*snakes*), an initial contour is deformed to optimize an objective function that is the sum of an image energy term, measuring the degree to which the contour matches the image data, and an internal energy term, measuring the degree to which the contour is smooth.[2] This objective function, except for an additive constant, can be interpreted as the logarithm of the posterior probability $p(\Theta|I)$ of the location of the set of boundary sites Θ, given the image data I; the image energy can be interpreted as proportional to the log of the likelihood $p(I|\Theta)$; and the internal energy can be interpreted as the log of the prior $p(\Theta)$, a model of expected shape properties. Summarizing the above,

$$\begin{aligned} \text{Objective function} &= \text{log likelihood} + \text{log prior} \\ &= k_{image} \times \textit{Image Energy} + k_{internal} \times \textit{Internal Energy} \\ &= k_{image} \times \textit{Boundariness Summed Over Boundary Sites} \\ &\quad + \; - k_{internal} \times \textit{Boundary Curvature Summed Over Boundary Sites}, \end{aligned} \tag{3}$$

where the boundariness at a location is the degree to which image intensities around that location in the image behave like a boundary at that point (for example, the gradient magnitude of intensity at the point is a commonly used measure of boundariness) and k_{image} and $k_{internal}$ are constants which weigh the relative importance of the two energy terms.

[2] For a thorough review of deformable contour techniques and their medical applications, see McInerney and Terzopoulos [1996].

The prior is intended to measure the degree to which the contour has the location, size, orientation, and shape of the family of objects that are to be extracted, e.g., of a particular organ. Wilson [1995], Johnson [1996], and Pizer [1996] have argued that such object properties are captured more richly by a linked locus of medial and boundary sites than by a locus of only linked boundary sites. Medial sites have not only a position, but also a width and yield two directions normal to the boundary at a scale proportional to the width. With such a locus, the agreement of the locus with the image, i.e., the image energy, involves not only the boundariness at the boundary sites but also the medialness at the medial sites, where medialness measures the degree to which the medial site, with its width and link directions, behaves as being medial between a pair of boundaries. Also, the log prior measures the variation of the inter-site links from a model. Therefore, the relations defining the objective function change as follows:

$$\begin{aligned} \text{Objective function} &= \text{log likelihood} + \text{log prior} \\ &= k_{image} \times \textit{Image Energy} + k_{internal} \times \textit{Internal Energy} \\ &= k_{bound} \times \textit{Boundariness Summed Over Boundary Sites} \\ &\quad + k_{med} \times \textit{Medialness Summed Over Medial Sites} \\ &\quad + -k_{internal} \times \textit{Sum Of Variation Inter-Site Links From Model} \end{aligned} \tag{4}$$

The method which we have implemented involves creating such a model from a training image, and deforming the model to optimize the objective function.

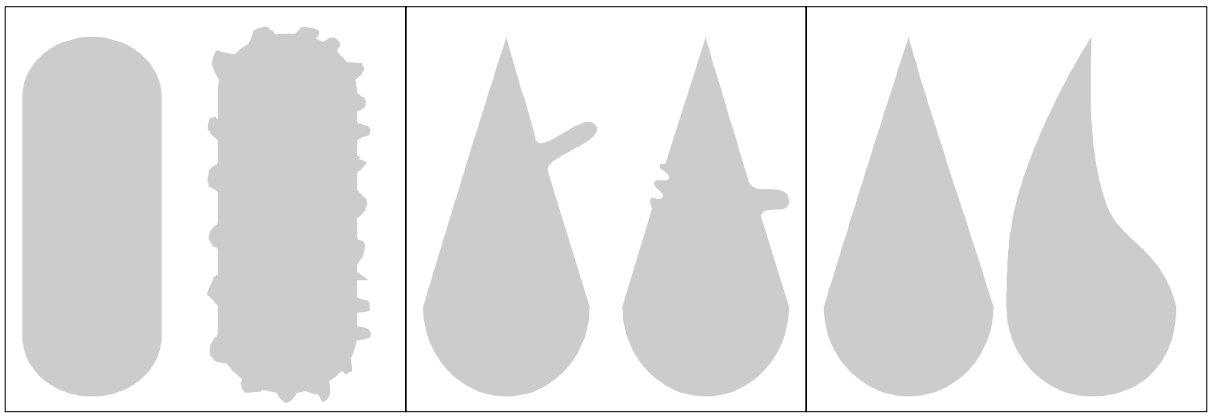
2. THEORY OF DEFORMABLE SHAPE LOCI

In this section, we lay out the basic principles for deformable shape loci (DSLs) and describe the construction of the prior (or external energy term) and the likelihood function (or image energy term) and the means for combining each to permit segmentation of an object via optimization of the posterior (the objective function). To motivate the discussion to follow, consider the three pairs of similar objects in Figure 1. We wish to capture, in a model, a means for describing how such shapes are similar and how they differ.

In the first case (left panel in Figure 1), note that both shapes appear tube-like and hence appear similar at the global, or *figural*, level. Yet, one of the shapes has a smooth boundary while the other has a wiggly boundary. We argue, and will make or case in later sections, that a model-driven approach to segmentation should allow for recognition (at the figural level) of both of these instances of a similar object, while also providing for determination of differences in their local boundary shapes.

In the second case (middle panel in Figure 1), we again have the situation where the overall shapes are similar and the objects may be described as being teardrop shaped. That is, they have a straight figural axis and a narrow top and increase in width from top to bottom. The overall boundary shape is also quite similar (i.e., the boundaries are relatively smooth), but these objects differ in that the position, size, and orientation of the protrusion relative to the tear-drop is different and in the appearance of a wiggle on a small portion of the boundary. If one considers the protrusion as an attached "subfigure" on the larger teardrop, then the natural way to describe the shape difference is to characterize how the shape of the subfigure varies in relation to the parent figure. We call this kind of shape relationship inter-figural shape and will show that it is possible to directly capture such aspects of object shape in a probabilistic model.

Finally (right panel in Figure 1) we have similar objects in the sense that they both has similar (smooth) boundaries, yet differ in the sense that the one of the objects appears straight while the other is curved at the figural level.



Similar figural shape
Different boundary shape

Similar figural shape and boundary shape
Different inter-figural shape

Similar boundary shape
Different figural shape

Figure 1. Three pairs of objects with similar shapes.

In the sections to follow, we first describe how we encode such aspects of object shape in a prior model called a shape template (section 2.1). We then discuss the construction of the likelihood function (section 2.2), which uses image measurements as to both figural (medial) and boundary features to allow for the stable extraction of objects from unclassified images via optimization of the posterior probability (section 2.3).

2.1 The Prior (Shape Templates): Modeling Normal Object Shape

Shape may generally be defined as the set of measurable properties of an object that remain unchanged, i.e., that are invariant to, a given set of geometric operations; for example, translation, rotation, and magnification. As discussed previously, object shape, as we consider it in this paper, has three important aspects:

a) Figural shape: includes a description of the track of the middle of a figure and the variation in width of the object along that track (medial axis).
b) Inter-figural shape: relative position, size, and orientation of component figures.
c) Boundary shape and boundary location relative to figural location and size.

Each of these aspects of object shape are represented, in a model, by relationships between coarsely samples medial sites and boundary sites, as illustrated in Figure 2.

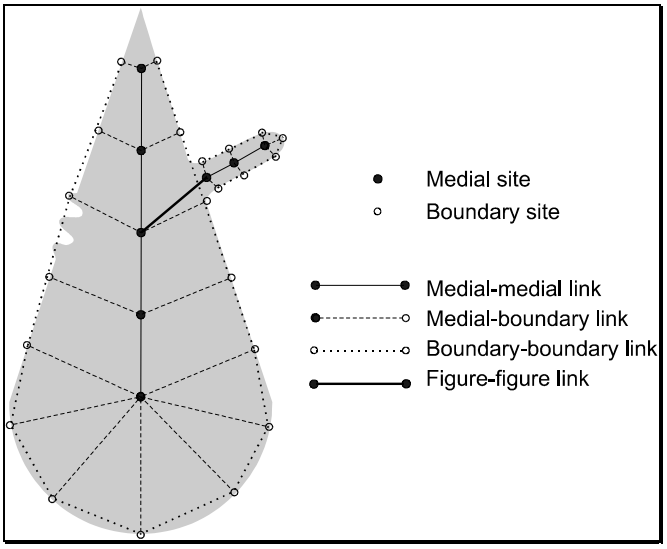


Figure 2. An object shape template for a teardrop figure with a protrusion subfigure.

In the following subsections, we describe a means for integrating such global and local aspects of object shape into a rich shape model, or template. There are two essential aspects to these models: (1) The discrete, sparse set of points, the *template sites*, which serve as landmarks at which certain image features are to be measured, and (2) the local relations between these sites, the *template links*, which serve to constrain the possible deformations that the template can undergo when it is applied to an image (see Figure 2).

2.1.1 Specifying template sites in the prior

The template sites for our object shape models consist of an orderly arrangement of landmarks at which certain important image features are to be measured when the template is applied to an unclassified image. In general, these image features may be considered to be either multilocal (figural) or local (boundary) features, which are captured by medial measurements or boundary-based measurements, respectively. Each figure is represented by a set of linked medial sites, sampled from the core of that figure. The medial sites establish both the position and local width of the figure and specify a set of scaled *medialness* measurement apertures, each of which responds most strongly to centers of objects of similar width. In turn, each medial site is associated with a pair of boundary sites to which we associate a scale value proportional to the scale of the linked medial site. The scale value serves to establish the width of *boundariness* measurement apertures.

2.1.2 Specifying and using template links in the prior

Our objective function has a log prior term that measures the difference in location, orientation, size (relative to a parent figure), and shape between a deformed model and the model itself. This measure is expressed in terms of differences in the links illustrated in Figure 2. The measure has a basis in the theory of Markov random fields and Gibbs priors [Wilson, 1995], but for the purposes of this paper, we will take the pragmatic point of view that, except for the multiplicative constant ($k_{internal}$ in Equation 4), the log prior term is simply the sum over the links of the norm-squared of the difference of that link in the deformed model from the same link in the training image:

$$\text{log prior} = -k_{internal}\left(\sum_{l_k \in MM\,links} \left\| l_k - l_k^{\text{model}} \right\|^2 + \sum_{l_k \in MB\,links} \left\| l_k - l_k^{\text{model}} \right\|^2 + \sum_{l_k \in BB\,links} \left\| l_k - l_k^{\text{model}} \right\|^2 + \sum_{l_k \in FF\,links} \left\| l_k - l_k^{\text{model}} \right\|^2 \right) \quad (5)$$

Using Figure 2 as an example, the links labeled *MM* are the medial-medial links; those labeled *MB* are the medial-boundary links; those labeled *BB* are the boundary-boundary links; and those labeled *FF* are the figure-figure links.

The remainder of this section covers two issues. The first is the process by which one specifies the links in the model, whose deformation will form the links in the deformed locus. The second is the measures of link difference which we use in the above formula to calculate the prior term in the objective function.

The k^{th} link connects two neighboring template sites, i and j. The model loci, and thus the links, are established through interaction by a user who understands the shape in question and thus is able to specify which figures are of importance in the model (see section 2.1.3). These figures might include not only portions of the anatomical object to be extracted but also those containing the object of interest (e.g., the whole abdomen when the object of interest is the liver), those in objects neighboring the object of interest (e.g., the gall bladder when the object of interest is the liver), or those made from the gap between the object of interest and a boundary of a neighboring object (e.g., the chest wall when the object of interest is the liver). For each figure that is to appear in the model, a medial locus must be extracted, and we provide tools for aiding the

model definer to do this. We frequently use stimulated core extraction in this process, which is described in more detail in section 2.1.3. We also provide tools that, given a figure's medial locus, calculate associated boundary points in the training image. These tools also are also described further in section 2.1.3. Finally, the figure-figure links are determined by the model definer by defining which figure is a child of which parent figure.

In the program we have implemented to deform a model locus by optimization of the objective function, the norm of a link difference that we use is the same for each link type. Each link l_k is a vector in scale space specifying both

1) the vector $\vec{v}_k$ in image space (2D in this paper) that indicates the change of location between the i^{th} and j^{th} positions $\vec{x}_i$ and $\vec{x}_j$, and

2) a pair of scales, σ_i and σ_j, indicating the change in widths between the i^{th} and j^{th} positions.

The difference between a link's position and width of changes in the model and the corresponding pair of changes in the deformed model, is measured by the magnitude of the difference, in scale space geometry, between the two scale space vectors. This can be computed [Wilson, 1995] from a scale space distance function d_{SS} on two scale space points and a scale space cosine function $\cos_{SS}$ on two scale space vectors, both of which are defined in [Eberly, 1994]:

$$\begin{aligned} \left\| l_k - l_k^{\text{model}} \right\|^2 = {} & d_{SS}{}^2\left[(\vec{x}_i,\sigma_i),(\vec{x}_j,\sigma_j)\right] + d_{SS}{}^2\left[(\vec{x}_i^{\text{model}},\sigma_i^{\text{model}}),(\vec{x}_j^{\text{model}},\sigma_j^{\text{model}})\right] \\ & - 2d_{SS}\left[(\vec{x}_i,\sigma_i),(\vec{x}_j,\sigma_j)\right] \times d_{SS}\left[(\vec{x}_i^{\text{model}},\sigma_i^{\text{model}}),(\vec{x}_j^{\text{model}},\sigma_j^{\text{model}})\right] \\ & \times \cos_{SS}\left[(\vec{x}_i,\sigma_i)-(\vec{x}_j,\sigma_j),(\vec{x}_i^{\text{model}},\sigma_i^{\text{model}})-(\vec{x}_j^{\text{model}},\sigma_j^{\text{model}})\right] \end{aligned} \tag{6}$$

This formulation provides invariance under the operations of translation and zoom (magnification) of the template, but relaxes the requirement for rotation invariance due to the angle-preserving term in the $\cos_{SS}$ function. We are investigating means, using 3-neighbor, triangular cliques, of restoring rotational invariance.

2.1.3 Automating the formation of and editing of the prior shape template

In this section, we briefly describe how we make use of existing algorithms and programs for extraction of the medial loci called cores to permit efficient construction of object shape templates on a training image (see Figure 3). As described in Pizer [1996], core extraction can be initiated, under user control, by the selection of an approximate middle and width of an object in question. From such a "stimulus" point, a nearby point on track in scale space that is the core is first found, and then the core itself is tracked. Using such an approach, we have been able to compute cores in a variety of imaging modalities and for a variety of objects in a few to tens of seconds, with the user having control of the endpoint positions.

Given a set of extracted cores, one for each of the figures which comprise the object in question, we have developed and implemented automated methods for extracting associated, approximate object boundaries. For each figural core, we compute the boundary at the scale of the core (BASOC), defined as the boundary of an object whose medial axis is the core (see Blum, 1978). The BASOC, in many cases, closely conforms to the actual figural boundary, or at the very least yields a good starting locus from which we subsequently apply an active contours approach [McAuliffe, 1996].

To further automate the model building process, a figural core is first coarsely sampled to yield the set of linked medial sites. We currently use a user-specified, equidistant spatial sampling of the core, but we are investigating other means of sampling

that take into account the shape of the figure as represented by the core (e.g., its width, curvature, etc.). The linked medial sites serve to establish a model for the global figural shape, and thus capture figural width and positional changes.

For each medial site, there exist a pair of opposing boundary points (called medial involutes) which lie on the BASOC and to which we establish links to that medial site. These boundary points become the boundary sites in the model. To provide for zoom-invariance in both the model and in the likelihood function, each boundary site is assigned a scale value (i.e., a Gaussian measurement aperture) proportional by a factor less than one to the scale of the medial site. These medial-to-boundary site links model the expected location of the boundary relative to the stable figure, and thus provide for stable boundary localization in the model fitting process.

Finally, the adjacent figural boundary sites are linked. These boundary-boundary site links serve to establish a model for expected boundary shape. These connections also serve to define the coarsely sampled contour that will be used to produce the segmentation of a figure after posterior optimization.

At any point in the automated model building stage, the user may wish to edit the site configurations. To facilitate user interaction, we have developed a model editing tool which incorporates the automatic model generation discussed above and which allows for insertion, deletion, and movement of sites and links. The example models of structures in the brain in shown in Figure 3 were generated in a few minutes using this tool. The model building tool communicates directly with a relational database stored on a centralized server for convenient access, storage, and management of model data.

Figure 3. A core-based tool for the efficient construction of shape templates.

2.2 The Likelihood Function: Capturing Image Information on Shape

Just as the measurement of shape change of the deformed template from the model template, producing the internal energy (prior), is intended to have built-in invariances to translation, rotation, and magnification, so should the image energy (likelihood) function, so that the segmentation can be unaffected by overall translations, rotations, and magnifications of objects in the image space. That is, for a given mapping of the model template to a given image, producing a posterior value, that posterior value should not change if the same transformation is applied to both the template and the image. Such measurement invariances can be obtained if (1) the measurement at a given site is based on a differential invariant (e.g., gradient magnitude), and (2) such a measurement is obtained with a Gaussian aperture that scales in size in proportion to object zoom [Pizer, 1996] (see also Lindeberg [1994a, 1994b] for detailed discussions of Gaussian scale space and differential invariants).

In the construction of the likelihood function we make use of two types of features, boundaries and middles, each of which is measured using scale-space differential invariants (as described in sections 2.2.1 and 2.2.2). The likelihood function in our method is a weighted sum of boundariness measurements at the N_b boundary sites, $(\vec{b}_i, \sigma_{b_i})$, and the medialness measurements at the N_m medial sites, $(\vec{m}_i, \sigma_{m_i})$:

$$\text{log likelihood} = k_{bound} \sum_{i=1}^{N_b} B(\vec{b}_i, \sigma_{b_i}) + k_{med} \sum_{i=1}^{N_m} M(\vec{m}_i, \sigma_{m_i}), \tag{7}$$

where the weights, k_{bound} and k_{med}, on the boundariness and medialness values, $B(\vec{b}_i, \sigma_{b_i})$ and $M(\vec{m}_i, \sigma_{m_i})$, respectively, are to-be-determined hyperparameters.

2.2.1 Medialness at points on the medial locus

The medialness at a point and width is intended to accumulate boundariness contributions at a distance from the point equal to the width and in normal directions associated with rays emanating from the point. Magnification invariance requires that the scale of measurement of these boundariness contributions are proportional to the width; this leads to medialness values that are produced from boundariness measurements at a significant scale when the object has significant width. This large scale, together with the accumulation of multilocal boundariness contributions, has the property of producing a measurement that is much more stable against a variety of image disturbances than boundariness at the somewhat smaller scale at which it is typically measured. Figure 4 illustrates this effect, which is proven at some length in Morse [1996], for a simple object in a noise-free image and in an image with a significant amount of additive noise. The effect is that an objective function optimization that includes medialness optimization stabilizes the computation as compared to deformable contours methods that only use boundariness information at a small, object-width-independent scale. In effect, when we perform optimization of the posterior in the model fitting process, the boundary locus localization is guided by the more stable medial locus localization.

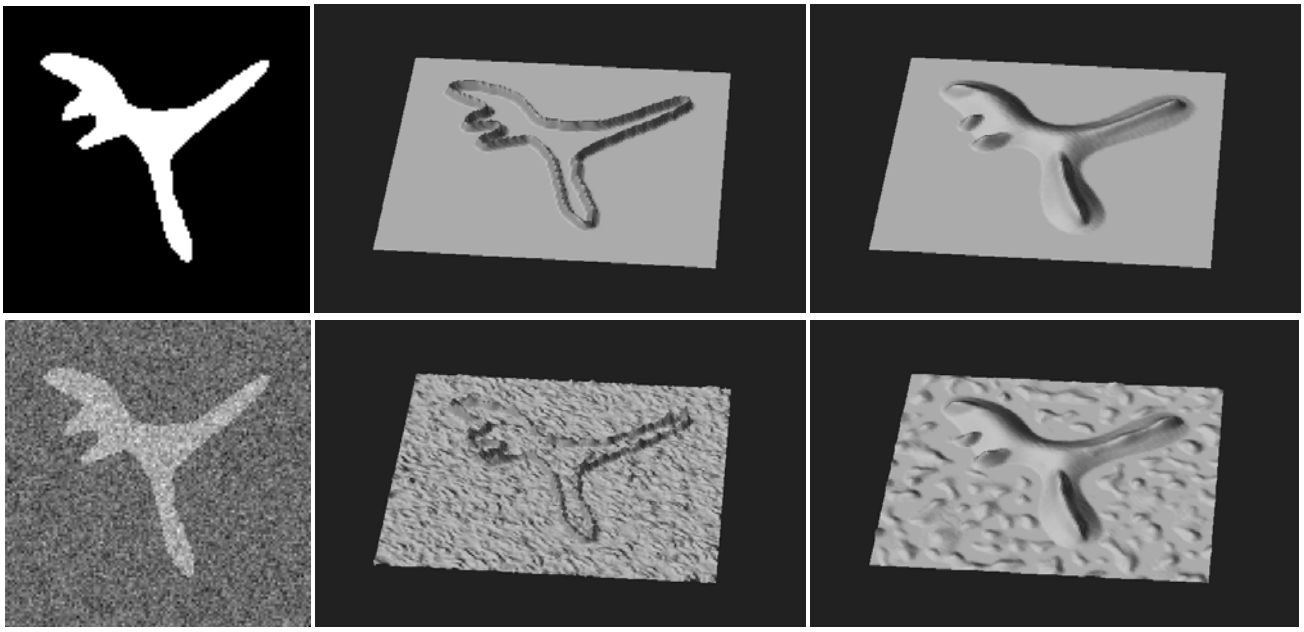


Figure 4. Effects of image disturbances on boundariness and medialness. Top row, left to right: Original image and graphs of boundariness (gradient magnitude) and medialness (at optimal scale). Bottom row, left to right: Image with additive noise and graphs of boundariness and medialness. Note the greater prominence of the ridge of medialness than the corresponding ridge of boundariness, for the noisy image.

The medialness function that we have used in our programs, which has been shown to work well in many image analysis tasks, is the maximum over orientation of the second derivative of the image at medial position $\vec{m}$ and scale σ_m,

$$M(\vec{m},\sigma_m) = -{\sigma_m}^2 L_{\vec{p}\vec{p}} = -{\sigma_m}^2 \vec{p}^{\,t} D^2 L \vec{p} = -{\sigma_m}^2 \lambda\,, \tag{8}$$

where L the scale-space representation of the original image,

$$L(\vec{m},\sigma_m) = I(\vec{m}) \otimes G(\vec{m},\sigma_m)\,, \tag{9}$$

where G is a unit-normalized Gaussian kernel of standard deviation σ_m. In equation 8, $D^2 L$ is the 2x2 matrix of second-order spatial derivatives of L, $D^2 L\vec{p} = \lambda\vec{p}$, and λ is the largest magnitude eigenvalue of $D^2 L$. In essence, such a medialness operator will give an optimal response (with respect to nearby positions and scales) at the center of a bar at a scale proportional (by a factor of 1.0) to the width of the bar, with the orientation determined by the orientation of the long axis of the bar.

2.2.2 Boundariness at points on the boundary locus

Boundariness, at a much smaller scale than that typically used for medialness measurement, is needed to guide each boundary point from its position relative to the medial locus in the model to the position which the image information indicates. This allows the boundary locus to reflect details while limiting its ability to deviate too much, relative to the figural width, from the location predicted by the model.

In our method boundariness measurements are computed at an image location $\vec{b}$ that is linked to a single medial point $\vec{m}$. The boundariness is computed at a scale σ_b that is a fixed fraction of the scale σ_m of the medial point to which it is linked. The boundariness $B(\vec{b},\sigma_b)$ is computed as a directional derivative of the image intensity in the direction of the link $\vec{u} = \vec{b} - \vec{m}$. That is,

$$B(\vec{b},\sigma_b) = \vec{u} \bullet \nabla L(\vec{b},\sigma_b), \tag{10}$$

where $\vec{u}$ is a unit vector and $\nabla L(\vec{b},\sigma_b)$ is the image space gradient of the intensity at position $\vec{b}$ and scale σ_b.

2.3 Optimizing the Posterior: Localization and Segmentation via Deformable Shape Loci

This section describes the process whereby the model (the prior or internal energy component) is combined with the image information (the likelihood or image energy component) such that the deformation of the model template results in a segmentation of a modeled object in an unclassified image (a relative maximum in the posterior or objective function). Specifically, we describe how our method uses large-scale, figural information captured by the medial locus to allow for the stable localization of the model in the image (section 2.3.1) and for the subsequent local optimization of the posterior to allow for segmentation (section 2.3.2).

To summarize the discussion in the previous sections, to perform a segmentation we wish to optimize an objective function, the log posterior, that consists of an image energy term, the log likelihood, and a model or internal energy term, the log prior:

$$\begin{aligned}
&\text{log posterior} = \text{ log likelihood } + \text{ log prior} \\
&= k_{bound} \sum_{i=1}^{N_b} B(\vec{b}_i,\sigma_{bi}) + k_{med} \sum_{i=1}^{N_m} M(\vec{m}_i,\sigma_{mi}) \\
&- k_{internal} \left(\sum_{l_k \in MM\ links} \left\| l_k - l_k^{\text{model}} \right\|^2 + \sum_{l_k \in MB\ links} \left\| l_k - l_k^{\text{model}} \right\|^2 + \sum_{l_k \in BB\ links} \left\| l_k - l_k^{\text{model}} \right\|^2 + \sum_{l_k \in FF\ links} \left\| l_k - l_k^{\text{model}} \right\|^2 \right),
\end{aligned} \tag{11}$$

where the hyper-parameters, k_{bound}, k_{med}, and $k_{internal}$, can be estimated from training data obtained from human observers [Wilson, 1996].

2.3.1 Initialization of the shape template for object localization

Like most deformable contour methods, we require that the model be initially positioned "close" to the object to be segmented in order to avoid convergence to an incorrect local maximum during the optimization process (e.g., a model for the kidney will not segment the desired object in a candidate image if it is initialized in the brain). To avoid such convergence problems in other deformable contour methods, which can occur with even modestly poor initializations, several authors have chosen an implementation using coarse-to-fine schemes in which either the most "salient" features are first considered [e.g., Radeva, 1995] or using multi-resolution images (pyramids) in a top-down fashion where the output at a lower resolution image is used as the input to the next highest-resolution image [e.g., Leymarie and Levine, 1993]. Because we use a prior that reflects shape richly and explicitly recognizes figural width as an important descriptor of shape, we believe our method has advantages, both computationally and theoretically, over other approaches. That is, we are able to perform initialization via saliency of figures, captured by the statistics of the object or figural hierarchy (see section 3.3) and

via use of multiscale and multilocal information, in the form of the medial sites which carry with them figural width information.

In our current implementation, initialization of a model to a candidate image containing an object to be segmented occurs in three stages (Figure 5). In the first stage, the model template is mapped to the image space of the candidate image. We currently perform a coarse registration of the training image and the candidate image to establish the approximate location of the object in question. In the second stage, an optimization procedure is initiated which attempts to fit the collection of medial sites to their respective positions in the candidate image via maximization of the sum of the medialness in the candidate image at the template medial sites as a function of a global translation, rotation, and scaling applied to the template. This process is exactly the same as the registration method proposed and implemented by Fritsch [1995] and, due to the invariance properties built into the prior, does not change the value of the prior (internal energy) from its mean. Due to the use of relatively large scale medial information, this stage is remarkably robust in the presence of small-scale image disturbances (e.g., boundary and/or intensity noise) and exhibits a region of convergence for translation on the order of the average scale (half-width) of the figures comprising the object. In the third stage of initialization, each of the individual figures is allowed to deform rigidly (with magnification) in order to optimize the same functional as with global initialization. This step allows for any small adjustments in the position, orientation, and size of the figures relative to the global object shape. In the case where the figures comprising an object are linked, this stage can produce small changes in the prior term as figure-subfigure links (if present) are deformed from the mean. We have found that initialization, using an optimization strategy called simplex optimization [Press et al., 1995], takes on the order of a few seconds per figure.

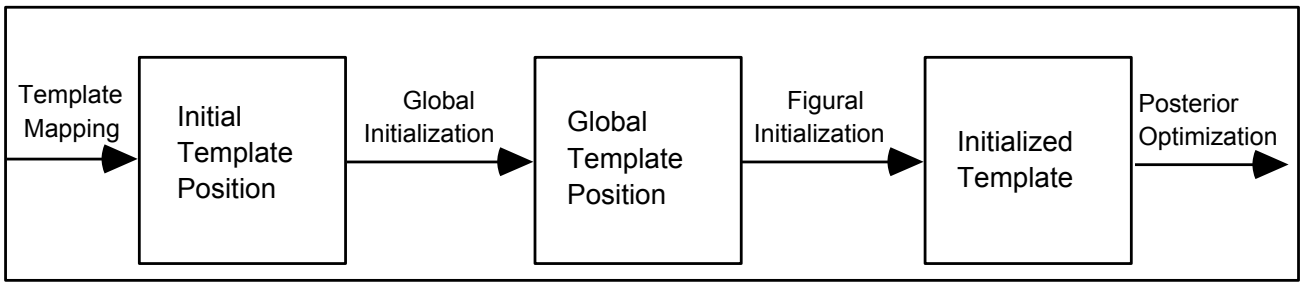
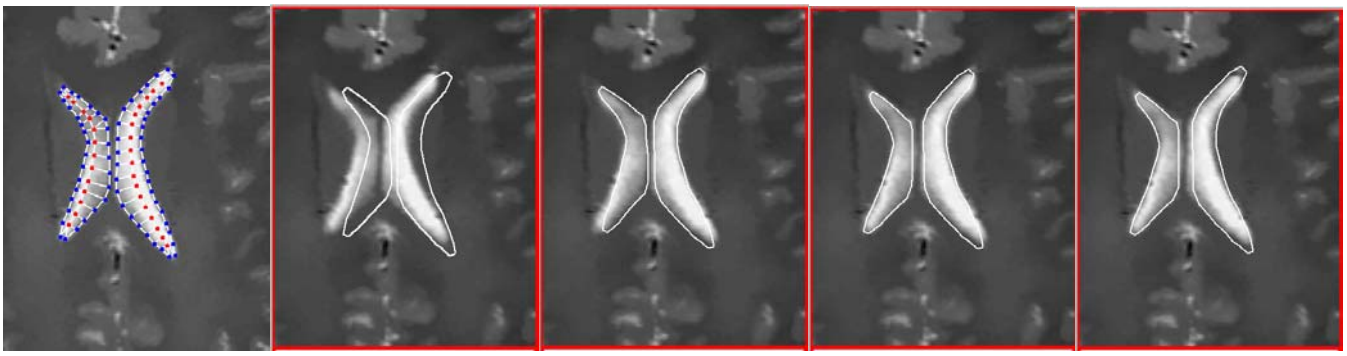


Figure 5. Schematic of the template initialization process.

Figure 6 illustrates the initialization procedure on an axial MR image of a brain ventricle. Figure 6a shows a sample model for a ventricle consisting of two unlinked figures (for simplicity, no figural hierarchy nor cross-figural links have been defined). In Figures 6b-e, a warped image with the model template (shown in the form of the segments connecting the boundary sites) in various stages of initialization is shown. This image was generated by warping the image from which the model was generated (see section 3.5.2 for a discussion of the warping procedure). In Figure 6b, the model template is displayed in its initial configuration in the warped image. Figure 6c shows the result of optimizing the integrated medialness in the warped image at all of the core sites in the model as a function of a global rotation, translation, and scaling of the template. In Figures 6d and 6e, each of the individual figures have been initialized in the warped image using the same criterion of maximizing integrated medialness.



(a) (b) (c) (d) (e)

Figure 6. Initialization of the model's global parameters on a warped image. (a) Original model superimposed on the training image. (b-e) Various states of model initialization on a warped image of the ventricle. (b) Initial mapping of the model to an image. (c) Position of the model after global initialization. (d) Position of the model after the left figural initialization. (e) Position of the model after right figural initialization.

2.3.2 Optimization of the shape template for object segmentation

Once an object in localized and the template is initialized, the template links are allowed to deform locally, via movement of the individual template sites, in order to optimize the objective function (the log of the posterior probability). Again, we take advantage of the stability properties of the relatively large-scale medial sites by first performing sequential, local optimization over all medial sites with associated boundary sites moving in concert so as to maintain the deformation energy (link difference) in the medial-boundary links at its post-initialization level. This allows the template links for each individual figure to deform based on the most stable features, the medial points, and thus to adjust to slight variations in figural shape from the prior mean.

Following this step, all of the template sites are allowed to move, in a sequential fashion, in an attempt to optimize the posterior probability. During this stage the boundary sites, which have been approximately aligned with the image features by the initialization process and by the previous step, move toward regions of high boundariness subject to the stabilizing constraints imposed by the prior links in the model. Fig. 7 shows a plot of the posterior probability versus iteration (over all sites) and the final configuration of the linked boundary sites after the last iteration. Typically, convergence toward a maximum in the posterior occurs in a number of iterations on the order of 10-20, and on an ordinary workstation the entire optimization process takes on the order of tens of seconds to several minutes (depending on the total number of sites in the model). Again, we use simplex optimization [Press, 1995] to optimize each point as a function of its scale space position.

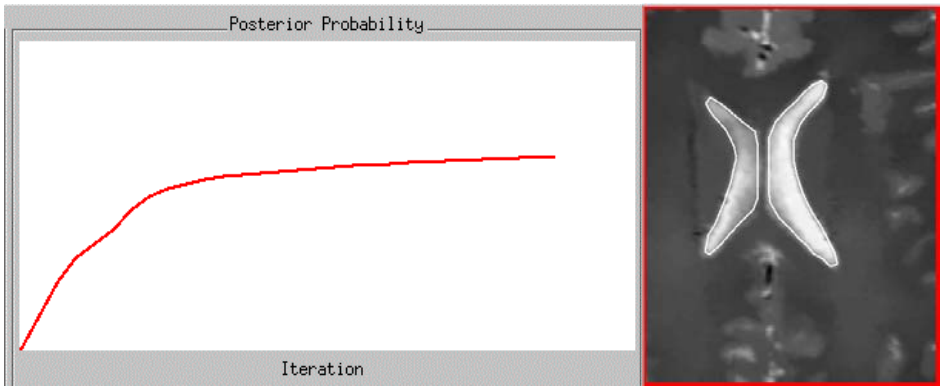


Figure 7. Left: Plot of the posterior probability versus iteration (over all sites) for the model in figure 6a applied to the image in figure 6e. Right: Final configuration of the boundary after the last iteration.

3. RESULTS

This section presents results of Monte Carlo studies undertaken to investigate the effectiveness of the segmentation method and to test its robustness in the presence of image disturbances (section 3.1). Also presented are recent results demonstrating the applicability of this method to clinical MR and CT images (sections 3.2-3.4), including a discussion of how object hierarchies may be used to facilitate object localization (section 3.4).

3.1 Monte Carlo Investigations

To test the effectiveness of our proposed segmentation methods, we have performed a pair of pilot studies to examine the robustness of the method under increasing object deformation and under increasing levels of intensity noise. The methodologies and results of these studies are described in sections 3.1.1 and 3.1.2.

3.1.1 Deformation

To be clinically useful, a model-based segmentation method must be able to handle the normal variability seen in object shape across a population of patients. To test the robustness of our method in the presence of such variability, we have devised a method for deforming, in a random way, a test object in an image on which we have defined a model template. In the procedure, we take the model image and define a warping on that image which displaces a set of fixed reference points by a random distance drawn from a uniform bivariate normal distribution with standard deviation σ_{warp}. Given the reference points and their respective randomly perturbed positions, we apply Bookstein's [1989] thin-plate-spline warping algorithm to produce a new image containing a warped version of the test object. The same warping function is applied to the model template to generate "truth" in the warped images.

Following application of the initialization and posterior optimization method to each of the warped images (see Figure 8 for representative examples), the final template is compared to the warped model to give a measure of segmentation quality. We currently use the mean distance between the polygonalized model template and warped template boundary points as our error measure. Results of this study are shown in the plot in Figure 9. Note that rather "large" warps still lead to subpixel mean errors. Moreover, the outliers with the higher errors for any level of warp have most of their error at figural ends, where we have not yet implemented measures of figural endness to attract the boundary.

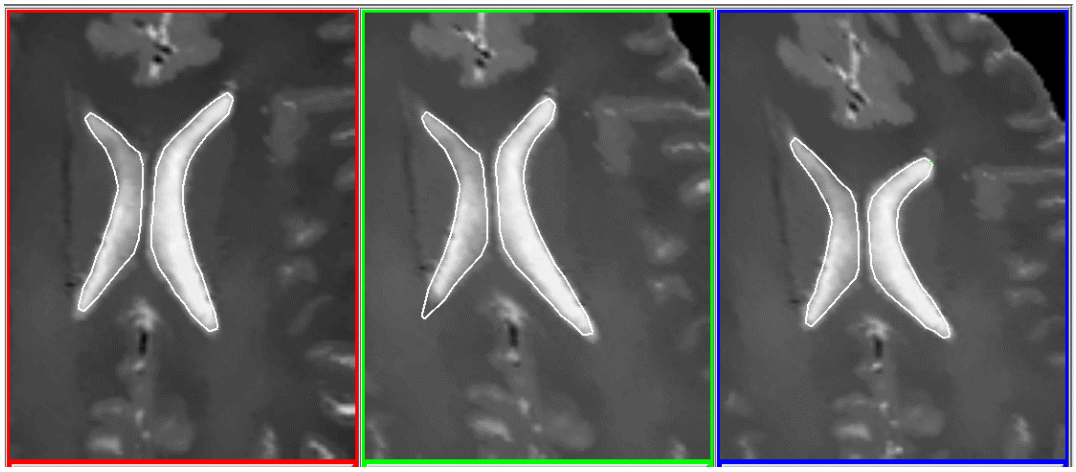


Figure 8. Boundaries resulting from optimization of the posterior probability function in warped images. The standard deviations for the warps, σ_{warp}, were 2, 4, and 8 pixels for the left, center, and right images, respectively. Note the inability of the method to conform to the lower end of the left-side of the ventricle in the center image due to lack of use of measurements capturing figural endness.

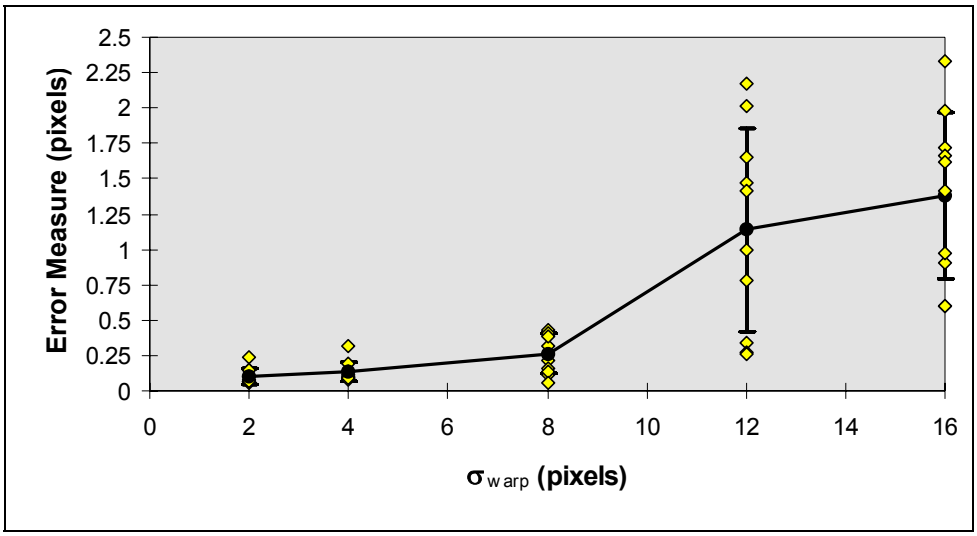


Figure 9. Mean segmentation error (in pixels) plotted as a function of warp.

3.1.2 Noise

One of our major claims is that our methods are able to robustly handle image disturbances because of the stabilizing effects of large-scale, medial measurements. To illustrate this robustness, we have repeated the experiment above using increasing amounts of additive, Gaussian-distributed white noise. In this study, a sample warped image near the mean in the middle group in figure 9 (with $\sigma_{warp} = 8$) was used as the unclassified test image. Shown in figure 10 are the final optimized boundaries for representative samples from the Monte Carlo runs with signal-to-noise levels of approximately 6, 3, and 1.5.

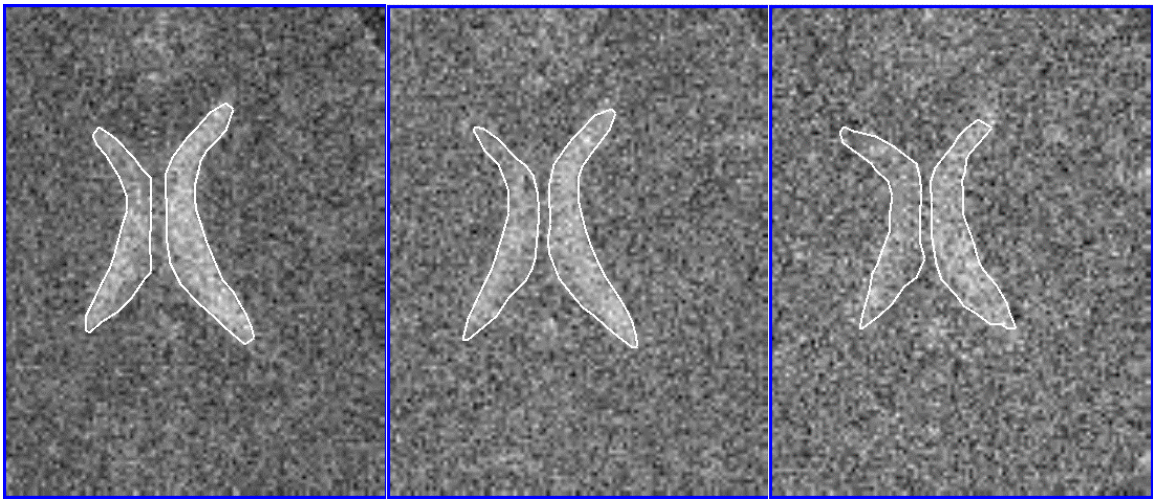


Figure 10. Boundaries resulting from optimization of the posterior function for increasing levels of noise (SNR ~ 6, 3, and 1.5 from left to right)

The plot of the results in figure 11 shows that even with noise levels much greater than one would expect to find in clinical images, the method is able to yield segmentations of the ventricle with sub-pixel errors.

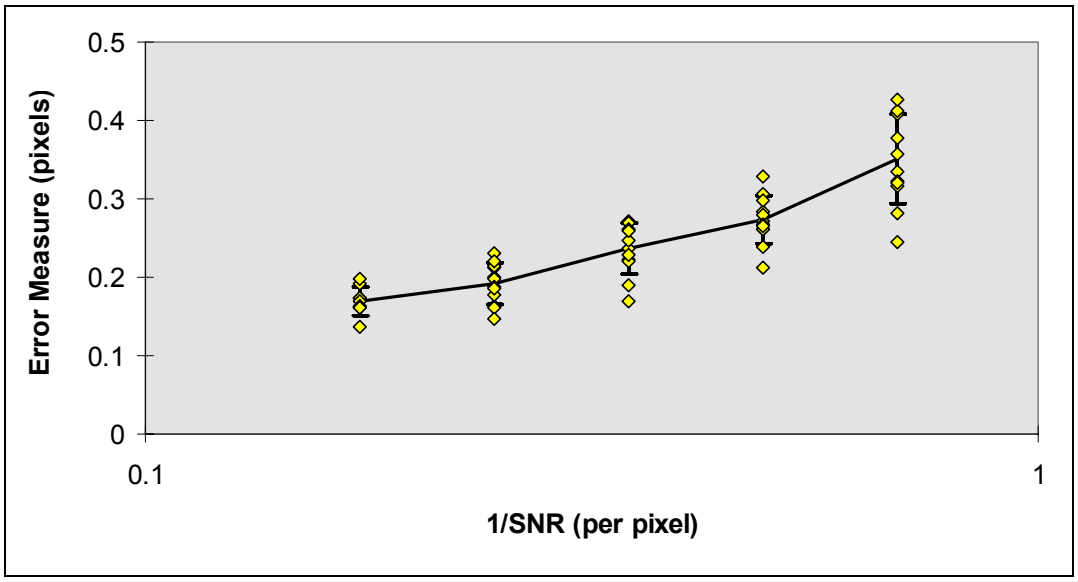


Figure 11. Mean segmentation error (in pixels) plotted as a function of image noise for a warp of 8 pixels.

3.2 Use of a Model with Inter-Figural Links on Clinically-Variable Ventricles

The simple, two-figure model of the ventricle used to produce the results in the preceding section made no use of inter-figural links. Hence, except during the global initialization stage, each of the figures was allowed to deform independently during the optimization process. We have recently begun to investigate the use of models that include such inter-figural links, and to take advantage of figural hierarchies wherein the most "stable" figures are used to establish the approximate location of subfigures.

Figure 12 shows a five-figure model for a ventricle superimposed on its training image. The model template, generated using the model building tool described in section 2.1.3, represents the ventricle by a parent figure comprised of a linked medial and boundary locus running down the relatively large-scale central portion, and by four subfigures with linked medial and boundary loci corresponding to each of the horns of the ventricle. In addition, we have defined a set of inter-figural links connecting neighboring boundary sites on the subfigures and the parent figure.

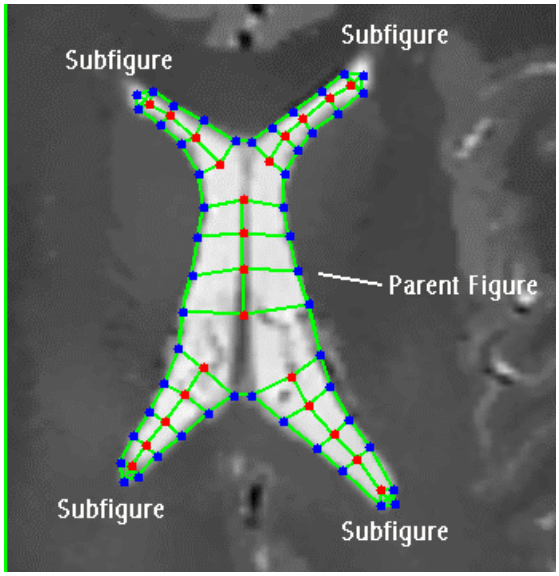


Figure 12. A five-figure model of a ventricle, with inter-figural links, superimposed on its training image.

This model was applied to a series of axial MR images, taken at approximately the same slice depth in the head, from a population of patients using a *hierarchical* localization scheme. First, the template sites for only the parent figure were matched to the images in order to optimize the sum of the boundariness and medialness at these sites as a function of global translation, rotation, and scaling. Following this initial mapping, each of the subfigures was placed relative to the parent figure such that the inter-figural link differences were zero, and each of the sites for the subfigures were then optimized over the individual figural parameters of translation, rotation, and scaling. Finally, sequential optimization over all template sites was performed to yield the results shown in Figure 13.

Such a hierarchical approach provides for more flexible localization, as each of the subfigures is allowed to adjust for slight differences in size, position, and orientation relative to the larger scale, more stable parent figure. As Figure 13 illustrates, a model generated from a single training image is capable of yielding segmentations of objects with a variety of subtle shape differences. Experience has shown that removal of the medial sites results in very few successful segmentations, as the medial sites are needed to help guide the boundary sites to their approximate locations in the initialization stage and provide stability during the local optimization stage.

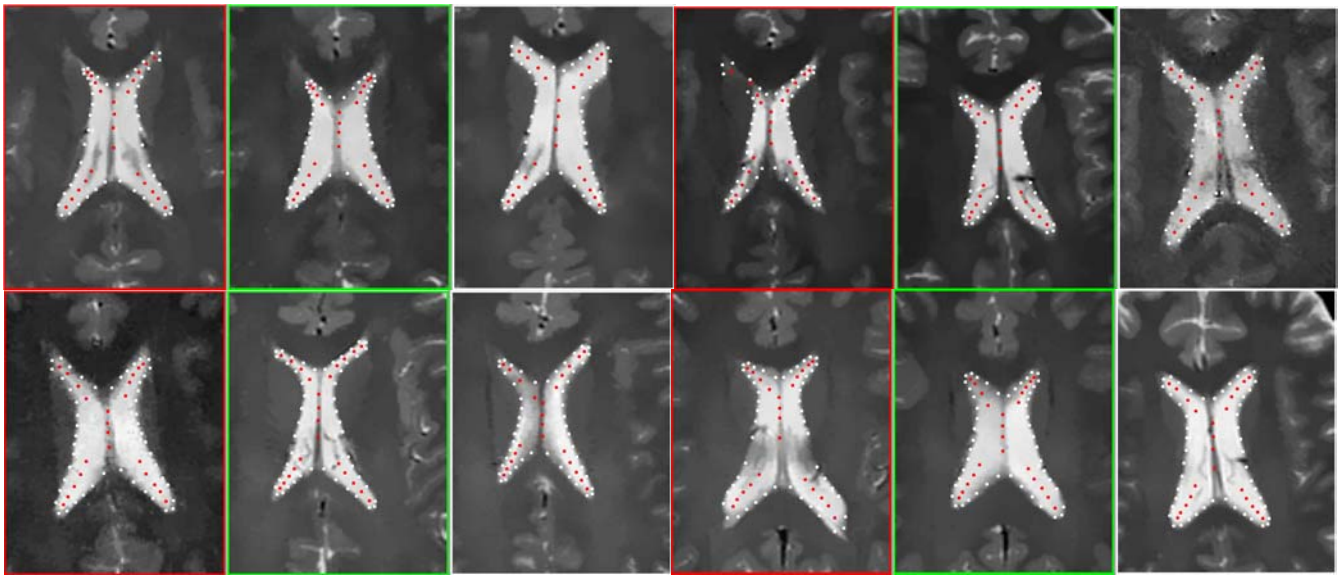


Figure 13. Final position of the deformed template sites on ventricles taken from a population of images from many patients. The model template used in the optimization is shown in Figure 12.

3.3 External Models

While the ventricle examples shown thus far have made use of only the figural information in the object (internal medial sites), we are investigating the possibility of using external structures, such as indentations, neighboring objects, and the spaces between objects, as means for defining models. When gaps between the anatomical object to be segmented and a neighboring object are narrow, the use of the gaps to guide the segmentation by the correspondingly narrow medialness apertures can avoid problems of inter-object interference. A recent example of such an approach applied to a liver in an axial CT slice through the abdomen is demonstrated in Figure 14. In this study, an external model was generated, using the interactive model building tool, on a characteristic slice. The model consists of 4 figures corresponding roughly to the gap between the liver and the chest wall, the gap between the liver and the kidney, the anterior tip of the liver (the only internal structure used), and a structure identified as the gall bladder. A boundary was defined by choosing a set of connections between certain boundary sites in the model and is shown superimposed on the adjacent CT slices from the same scan, above and below the model slice, in the left and right image in the top row of Figure 14. The optimization process was applied to the full model on all slices and the resulting boundary is shown for each of the three images in the bottom row of Figure 14.

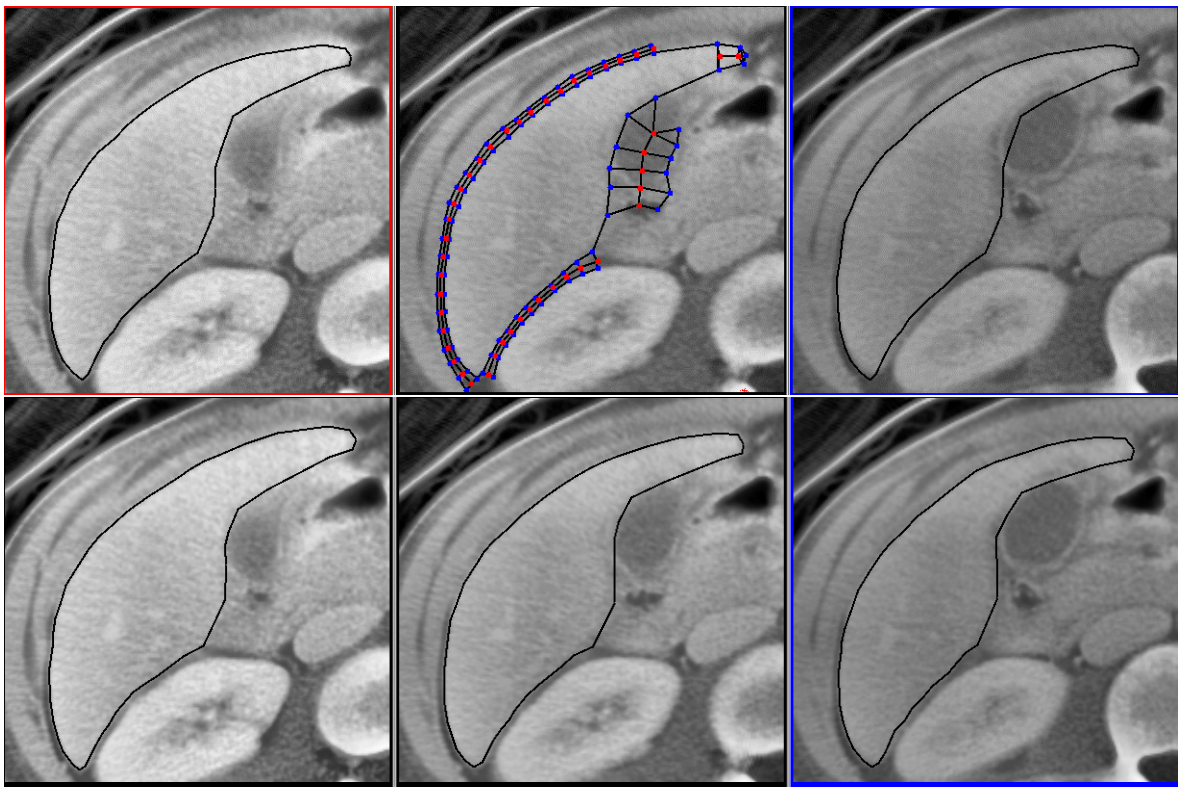


Figure 14. Segmentation results from a 2D external model of the liver. Top row shows 3 adjacent CT slices through the liver with the defined boundary in the initialized model superimposed. The full model is shown on the middle training image. Bottom row shows the same slices with the optimized model boundary superimposed.

3.4 Use of an Object Hierarchy

As a final example, we present an application of a DSL approach utilizing multiple objects organized in an object hierarchy. Similar to the use of figural hierarchies (for multi-figure objects), where subfigures are located relative to more stable parent figures, this approach seeks to utilize the most stable objects to establish the relative location of other objects. In this example, a model was constructed on an axial CT slice through the abdomen on five objects: the entire abdomen cross section, each of the kidneys, the vertebra, and the spinal cord where the linked boundary loci are shown in Figure 15. Using the strategy outlined below, the hierarchy of object models was fit to another slice (from the same CT scan) approximately 1.6 cm away from the model slice.

Because the model of the abdomen contains the largest-scale information in its medial sites, and hence is little-affected by the presence of smaller-scale information in the image, the localization approach is first applied to the template points in only the abdomen template. The shift, rotation, and scaling needed to match the abdomen template is then applied to all of the models such that the initial position of each of the sub-objects (kidneys, vertebra, and spinal cord) is defined relative to the abdomen. Next, the localization process is applied to all of the sites in the kidneys, vertebra, and spinal cord and they move as a collection to optimize the image energy as a function of their overall position, orientation, and scaling. This establishes the approximate location of the group of sub-objects. Finally, the localization process is applied to each of the individual sub-objects, after which the local optimization process commences. In the current implementation, we have defined no inter-object links which could potentially constrain the shape of a sub-object conditional upon the shape and position of a parent

object. How to introduce such inter-object links and how to most effectively make use of hierarchies of object models is a subject of ongoing work.

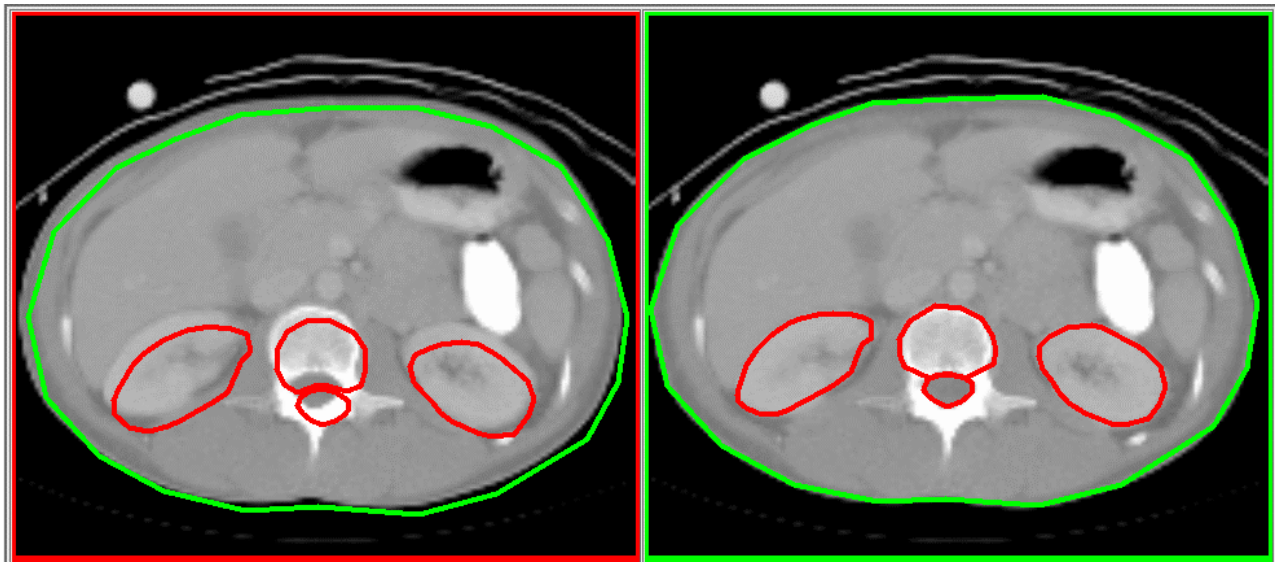


Figure 15. Use of an object hierarchy to segment structures in an axial CT slice through the abdomen. (a) Original configuration of object models (on an unclassified image. (b) Final template configurations following hierarchical object localization and local object posterior optimization.

4. DISCUSSION AND CONCLUSION

We have described a method which takes a deformable loci approach to segmentation but uses a prior richly describing shape in terms of both figural and boundary aspects, and an image energy that stabilizes by its use of medialness that has a scale significant relative to figural width and yet sensitively locates the boundary using smaller-scale measurement of boundariness. The model is formed through a user's appreciation of what are the stable figures of the shape of a target object, so the model formation involves interaction, aided by core-extracting methods. On the other hand, the segmentation of the object for a new patient is intended to be fully automatic in most cases. The method, as initially programmed, requires only from several seconds to a few minutes to segment a multifigure object on an ordinary workstation. The fact that for N boundary sites there are only O(N) model sites and only O(N) links means that the method only requires O(N) time. When applied to relatively high contrast objects such as the ventricle in MR brain images, the method has been shown to produce subpixel accuracy in the presence of deformations of the model and large amounts of intensity noise. It has also been shown able to follow quite a range of normal clinical variation in ventricle shape.

We believe that the method has some additional advantages to be explored in future research:

1. The locality of the links allows it to localize shape anomalies or locations in which the intensity relations in clinical image deviates from what is expected.
2. The availability of links allows the method to imply directions in which intensity derivatives reflecting boundariness should be taken.
3. The locality of the model allows one to designate the intensity properties reflecting boundariness as a function of boundary site. Thus luminance change polarity, intensity range, contrast range, and the choice of type of edge: e.g., outline vs. luminance edge vs. texture edge, can be designated with boundary site.

Additional features that we expect to be added in the future include the following:

1. Testing of the method in more challenging clinical situations, with lower signal-to-noise, lengthy regions of boundary dropout, complex backgrounds, etc.
2. Including measurements of figural endness at figural ends in the image energy.
3. Using 3-site in place of 2-site cliques (links), leading to rotationally invariance, unlike the present situation.
4. Measuring variabilities in the model over various training images, and reflecting these variabilities in the hyperparameters (weights) of the objective function.
5. Extending the method to 3D, in which the medial and boundary loci are 2-manifolds and 5-cliques are used. The method remains O(N) in the number of boundary sites. We already have developed a method for creating such 3D models from a training image and will report this development in another paper.
6. Using the advantages just previously described.

5. ACKNOWLEDGMENTS

The authors of this paper gratefully acknowledge the contributions to this research by David Eberly, who has formulated algorithms and mathematics for core extraction, by Allyson Wilson, who developed much of the statistical framework on which this work is based, and by KC Low and Matthew McAuliffe, who have provided much assistance in the development of the segmentation software. We also acknowledge financial support from the National Cancer Institute Program Project Grant P01 CA47982, Medical Image Presentation, and from funds provided by the National Library of Medicine grant 5 RO1 LM05508, HPCC Technology for Real-time Medical Decision Support.

6. REFERENCES

Blum H and Nagel RN: Shape description using weighted symmetric axis features. *Pattern Recognition* **10,** 167-180 (1978).

Bookstein FL: Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans Pattern Anal and Machine Intell*, **11**:567-585 (1989).

Chakraborty A, Staib L, and Duncan JS: Deformable boundary finding influenced by region homogeneity. *Proc Conf Computer Vision and Pattern Recognition* (CVPR '94), 624-627 (1994).

Cootes TF, Hill A, Taylor CJ, and Haslam J: The use of active shape models for locating structures in medical images. *Information Processing in Medical Imaging* (IPMI '93), HH Barrett, AF Gmitro eds., Lecture Notes in Computer Science, **687**:33-47, Springer-Verlag, Berlin (1993).

Eberly D, Gardner RB, Morse BS, Pizer SM, Scharlach C: Ridges for image analysis. *Journal of Mathematical Imaging and Vision*, **4**:351-371 (1994a).

Eberly D: A differential geometric approach to anisotropic diffusion. In *Geometry-Driven Diffusion in Computer Vision*, BM ter Haar Romeny (ed.), Kluwer Academic Press: 371-392 (1994b).

Fritsch DS: Registration of radiotherapy images using multiscale medial descriptions of image structure. Ph.D. dissertation, Department of Biomedical Engineering, University of North Carolina at Chapel Hill (1993).

Fritsch DS, Chaney EL, Boxwala A, McAuliffe MJ, Raghavan S, Thall A, and Earnhart J: Core-based portal image registration for automatic radiotherapy treatment verification, *Int J Radiat Oncol Biol Phys* **33**(5): 1287-1300 (1995).

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

Leahy R, Herbet T, and Lee R: Application of Markov random fields in medical imaging. *Information Processing in Medical Imaging* (IPMI '89), 1-14 (1989).

Leymarie F and Levine MD: Tracking deformable objects in the plane using an active contour model. *IEEE Trans PAMI* **15** (1993).

Lindeberg T: *Scale-Space Theory in Computer Vision*, Kluwer Academic Publishers, Boston (1994a).

Lindeberg T and ter Haar Romeny BM: Chapter 1: Linear scale space I: Basic Theory, and Chapter 2: Linear scale-space II: Early visual operations, *Geometry-Driven Diffusion in Computer Vision*, Kluwer Academic Publishers, Dordrecht, 1-72 (1994b).

McAuliffe MJ, Eberly D, Fritsch DS, Chaney EL, and Pizer SM: Scale-space boundary evolution initialized by cores. *Proceedings of the Fourth International Conference on Visualization in Biomedical Computing*, Hamburg, IEEE Computer Society Press, Los Alamitos, CA (1996).

McInerney T and Terzopoulos D: Deformable models in medical image analysis. *Proc Workshop on Math Methods in Biomed Img Anal*, IEEE Cat. # 96TB100056: 171-180 (1996).

Morse BS, Pizer SM, and Liu A: Robust object representation through object-relevant use of scale. *Information Processing in Medical Imaging* (IPMI '93), HH Barrett, AF Gmitro eds., Lecture Notes in Computer Science, **687**:112-131, Springer-Verlag, Berlin (1993).

Morse BS, Pizer SM, Puff DT, and Gu C: Zoom-invariant vision of figural shape: Effects on cores of image disturbances. University of North Carolina technical report TR96-005. To appear in *Computer Vision and Understanding* (1996).

Pizer SM, Eberly D, Morse BS, and Fritsch DS: Zoom-invariant vision of figural shape: The mathematics of cores. University of North Carolina Computer Science Technical Report, TRN96-004. To appear in *Computer Vision and. Image Understanding* (1996).

Pizer SM, Fritsch DS, Johnson VE, and Chaney EL: Segmentation, registration, and measurement of shape variation via image object shape. University of North Carolina Computer Science Technical Report, TRN96-031 (1996).

Press WH, Teukolsky SA, Vetterling WT, and Flannery BP: Numerical Recipes in C: The Art of Scientific Computing, Second Edition. Cambridge University Press, Cambridge (1995).

Radeva P, Serrat J, and Marti E: A snake for model-based segmentation. *IEEE Trans PAMI* **8**: 816-821 (1995).

Staib L and Duncan J: Boundary finding with parametrically deformable models. IEEE Transactions on Pattern Analysis and Machine Intelligence 14: 161-175 (1992).

Wilson A: Statistical models for shapes and deformations. Ph.D. dissertation, Institute of Statistics and Decision Sciences, Duke University, Durham, NC (1995).

Wilson A, Johnson V, Pizer SM, Fritsch DS, Yu L, and Chaney E: Towards a framework for automated image analysis. *Image Fusion And Shape Variability Techniques*, *Proc. 16th Leeds Annual Statistical Research Workshop*, University of Leeds Press (1996) (In press).