Medical Physics Journal
May 2005, **32**(5): 1335-1345

A Method & Software for Segmentation of Anatomic Object Ensembles by Deformable M-Reps

Stephen M. Pizer, P. Thomas Fletcher, Sarang Joshi,
A. Graham Gash, Joshua Stough, Andrew Thall, Gregg Tracton, Edward L. Chaney
Medical Image Display & Analysis Group, University of North Carolina, Chapel Hill

Abstract

Deformable shape models (DSMs) comprise a general approach that shows great promise for automatic image segmentation. Published studies by others and our own research results strongly suggest that segmentation of a normal or near-normal object from 3D medical images will be most successful when the DSM approach uses 1) knowledge of the geometry of not only the target anatomic object but also the ensemble of objects providing context for the target object and 2) knowledge of the image intensities to be expected relative to the geometry of the target and contextual objects. The segmentation will be most efficient when the deformation operates at multiple object-related scales and uses deformations that include not just local translations but the biologically important transformations of bending and twisting, i.e., local rotation, and local magnification. In computer vision an important class of DSM methods uses explicit geometric models in a Bayesian statistical framework to provide *a priori* information used in posterior optimization to match the DSM against a target image. In this approach a DSM of the object to be segmented is placed in the target image data and undergoes a series of rigid and non-rigid transformations that deform the model to closely match the target object. The deformation process is driven by optimizing an objective function that has terms for the geometric typicality and model-to-image match for each instance of the deformed model. The success of this approach depends strongly on the object representation, i.e., the structural details and parameter set for the DSM, which in turn determines the analytic form of the objective function. This paper describes a form of DSM called *m-reps* that has or allows these properties, and a method of segmentation consisting of large to small scale posterior optimization of m-reps. Segmentation by deformable m-reps, together with the appropriate data representations, visualizations, and user interface, has been implemented in software that accomplishes 3D segmentations in a few minutes. Software for building and training models has also been developed. The methods underlying this software and its abilities are the subject of this paper.

1. Introduction

As humans who do manual segmentation of anatomic objects know, the extraction of one or more objects from a 3D medical image typically requires the knowledge of the conformation that normal anatomy takes. This conformation involves multiple objects: the target object or objects and neighboring objects. This paper describes a method that can represent this knowledge and use it for segmentation. It describes a method especially suited for representing models of object ensembles, as well as single objects, and a method for segmentation via deformation of a model of this form. Our objective is a segmentation method that is largely automatic.

We use the general segmentation approach already shown by others to lead to success [Cootes 1993, Staib 1996, Delingette 1999, among others; also see McInerny, 1996 for a survey of active surfaces methods], namely deforming a geometric model by optimizing an objective function that is the sum of a geometric typicality term and a geometry-to-image match term. We start from a statement of the segmentation objective as finding the

most probable conformation of the target object, given the image. Then the geometric typicality term ideally measures the logarithm of the probability that the candidate geometric entity exists in the population of objects [Pizer 2003]. And the geometry-to-image match term ideally measures the logarithm of the probability that the target image values, relative to the candidate geometry, would arise in the population of images from that modality. As a fundamental means of obtaining efficiency, we optimize such an objective function successively to make the deformation refine for successively smaller spatial tolerances (spatial scales), where each of the spatial scale levels are object-relevant: the object ensemble, the object, the slab figure, the figural section, and the object boundary detail.

The most common geometric representation in the literature of segmentation by deformable models is made up of directly recorded boundary locations, sometimes called *b-reps* [Cootes 1993, Kelemen 1999, papers surveyed by McInerny 1996]. Our *m-reps* representation and segmentation method uses a medial representation intended to produce improved segmentations because 1) it inherently allows the deformation of an object to be decomposed into local translations, local twistings and bendings (rotations), and local magnifications; 2) it is especially designed to deal with both objects and the surrounding ensembles of objects; 3) it provides an anatomic object and object ensemble based coordinate system in terms of which to deal with the geometry-to-image match; 4) it directly supports a series of object-based representation at successively smaller spatial scales, allowing the posterior probability optimization to proceed in an inherently efficient coarse-to-fine manner.

M-reps provide these important advantages over other deformable object representations at the expense of a level of complexity that required the development of special theoretical underpinnings, software, and validations. While m-reps do not share with more traditional medial methods their oversensitivity to boundary details, small indentations and protrusions of anatomic objects are impractical to model medially. Our approach to solving this problem is to implement a non-medial boundary stage. M-reps do share other fundamental properties of medial models that can be taken as disadvantages in certain special situations. These are discussed in section 7.

In this paper section 2 describes the general segmentation approach, including the means for measuring both geometric typicality and geometry-to-image match. Section 3 describes m-reps, the multi-scale representation of object ensembles that we use, illustrating with a number of anatomic models. It also shows how we use m-reps both to compute the boundaries of objects and to provide the positional correspondences needed in computing both the geometric typicality and the geometry-to-image match. Section 4 gives some of the details of the actual operation of our segmentation program. Section 5 characterizes and illustrates some segmentation results of our method. A companion paper [Rao 2004] describes details of a measured comparison with manual segmentation of kidneys that showed excellent agreement between m-reps and humans. In that section the speed of segmentation is also discussed. Section 6 describes how both geometric and intensity models are built. The final section includes a discussion of the method and the software and indicates future directions in which our both will be developed.

Our earlier paper on this subject, published in *Int J Comp Vis* [Pizer 2003], has image analysis experts as its target audience and focuses on the m-rep representation and its deformations. Readers of this paper who wish a full mathematical description of m-reps and of segmentation by deformable m-reps are referred to that paper. This paper has users of deformable m-reps based segmentation as its target audience and focuses on the segmentation method and the statistics used. Those readers who are interested in a general impression of our

method and its use may wish only to scan sections 3 and 6, focusing mainly on the figures there, as these sections are focused on technical matters of image analysis and object representation.

2. Segmentation by Multiscale Deformation of Figurally Based Models

2.1. Segmentation method and detailed geometric representation

Our method for deforming a model into image data begins with either a manually chosen initial positioning of the model (Fig. 1) or an automatic positioning achieved by gross fitting of the model to a large cross-section of the body. The segmentation process then follows a number of stages of segmentation at successively smaller levels of scale. The spatial tolerance of the resulting segmentation can be large at the largest scale level but decreases as the scale gets smaller.

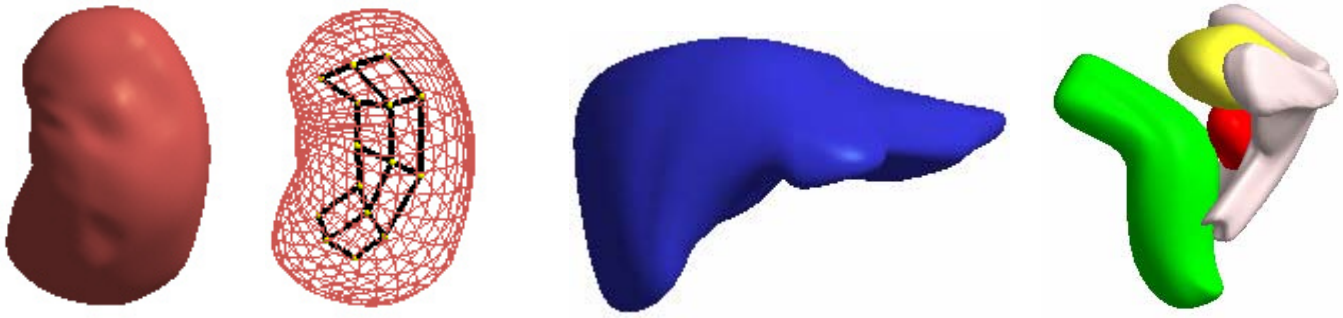

Fig. 1. M-rep modeled kidney with its medial mesh, liver, and male pelvis (two bones, bladder, rectum, prostate). The kidney model also shows the underlying representation of a sampled medial surface and a tiled boundary.

At each scale level the same objective function is optimized by geometrically transforming the entities at that scale level, using a transformation global to the respective entity. Thus, at the largest scale level, the object ensemble stage, the whole object ensemble undergoes a global transformation. At the next smaller scale level, each object making up the object ensemble separately undergoes a transformation global to it. And as the computation moves to successively smaller scale stages, successively smaller entities making up the entities at the next larger scale level are optimized with a transformation global to each of them. The series of optimizations concludes with the vertices of the tiles representing the object boundary each being optimized.

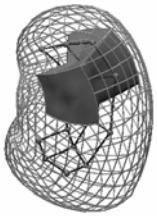

Fig. 2. An interior through section of a single-figure kidney

The scale levels smaller than the object stage are first the figure stage, where *figures*, of which an object is composed, are optimized. These figures are simple slabs, i.e., smooth dilations of smooth curved sheets. Each figure forms either the main slab of the object or a subfigure forming a protrusion or indentation. Each figure is composed of a 2D array of interior sections through the slab (Fig. 2), and working on individual interior sections in sequence forms the figural section stage. Each interior section corresponds to the basic primitive of the m-rep, the medial atom, which is described in section 3. Together the medial atoms forming the figures and thence the objects made from the figures imply a boundary made from a set of quadrilateral or triangular tiles, and the vertices of these tiles form the representation that is optimized at the smallest scale level, moving these vertices from the positions implied by the medial portion of the representation.

At each scale level larger than the boundary scale level, the geometric transformation of the entity is made up of a similarity transformation, i.e., a local translation, rotation, and uniform magnification, together with a few modes of warp specialized to that particular entity. At the boundary scale level, the optimization is over only a single parameter per vertex, namely offset normal to the medially implied boundary. This choice to optimize at most 10 parameters per entity provides efficiency and convergence of the segmentation at that scale level. Typically the warp term is based on principal warp modes in the statistics, over a training population, of the medial atoms making up the entity being optimized at that scale level. We refer the reader interested in the mathematical detailing of the transformations and their statistics to [Pizer 2003] and [Fletcher 2004], respectively.

At each scale level we use an adaptation of the conjugate gradient method to optimize an objective function of the same form: the sum of a geometric typicality metric and a geometry-to-image match metric. Both metrics are detailed further in this section. The geometric typicality metric at any scale level measures any candidate geometric change from the model provided by the result at the next larger scale level in terms of how likely that change is in the population of normal cases. The geometry-to-image match metric measures, in reference to the object boundaries of the candidate, how likely the target image is. This is achieved by providing a boundary-relative template image (Fig. 3), together with measures of the deviation of the target image from that template image.

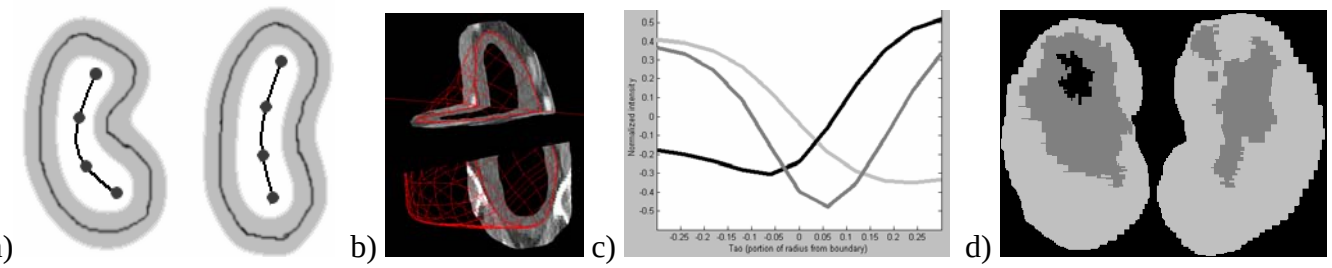

Fig. 3. The boundary-relative stencil region and the intensity profiles that make up the template stencil for measuring geometry-to-image match a) in 2D, both before and after deformation. b) in 3D, the mesh showing the object boundary and the greyscale in the collar showing the image intensities there in two orthogonal slices. c) intensity profiles for the kidney (trained from kidney images). d) locations where each profile is dominant. In Figs. c and d light grey indicates a light to dark profile, between the kidney and surrounding general tissue; black indicates a dark to light profile, between the kidney and an adjacent organ, e.g., the liver or spleen; medium grey indicates a light to dark to light notch profile, between the kidney, through adjacent general tissue, and then into an adjacent organ.

2.2. Geometric typicality metric

In the following discussion “geometric typicality” refers to a curvilinear generalization of the Mahalanobis distance. Details of characterizing the statistical distribution of target object shape in curvilinear space, called principal geodesic analysis, and computing the geometric typicality term are beyond the scope of this article. Interested readers are referred to [Fletcher 2004].

At all scale levels the geometric typicality metric of the relevant geometric entity reflects its location, orientation, and magnification properties but also its relation to immediately neighboring peer entities. That is, objects need to be understood in relation to neighboring objects, figures need to be understood in relation to neighboring figures, interior figural sections need to be understood in relation to neighboring sections, and boundary tile vertices need to be understood in relation to neighboring vertices.

Moreover, the neighbor-irrespective aspects of location, orientation, and magnification of an entity at any scale level can be most effectively measured relative to its values predicted by the previous, next larger, scale level. Thus, the typicality of an object is measured relative to the form it takes as part of the object ensemble of which it is a part, as well as relative to neighboring objects. The typicality of a figure is measured relative to the form it takes as part of the object of which it is a part, as well as relative to neighboring figures. The typicality of an interior section through a figure is measured relative to the form it takes as part of the figure of which it is a part, as well as relative to neighboring interior sections of that figure. The typicality of a boundary vertex is measured relative to the form it takes as part of the figural interior section of which it is a part, as well as relative to neighboring vertices.

Two special neighbor relations deserve comment. One is the relation among nearby (possibly abutting) objects (see male pelvis in Fig. 1). Not only the correct relative position, orientation and size need to be reflected in the geometric typicality, but also an interpenetration of the figures needs to result in a low geometric typicality. The second neighbor relation of note is that between a protrusion or indentation *subfigure* to the "host" figure on or into which it sits (see liver in Figs. 1 & 5). A useful idea is that the subfigure should ride on the boundary implied by the host's representation and be known in the figure-relative coordinates of the host. Thereby we can make measurements of typicality in terms the position of the subfigure relative to the host, the orientation of the subfigure relative to the host, and the size of the subfigure relative to the host.

2.3. Geometry-to-image match metric

The geometry-to-image match must relate the target image intensities to the image intensities in a set of training images, with all images given relative to where the target object and the neighboring object ensemble are in those respective images. We call these object-ensemble relative image values a *stencil*. We measure stencils (Fig. 3a,b) in terms of intensity profiles (Fig. 3c) normal to the target object and spatially scaled by the local object width. As described in [Stough 2003], we find those profiles in the training set that appear in approximately the same form in many cases and at many object boundary positions – we give the name *clusters* to these commonly appearing collections of profiles. In the kidney there are three clusters (Fig. 3d), roughly light inside to dark outside, dark inside to light outside, and light inside to a dark notch outside to light further outside. We compute the mean of each cluster and form a template stencil by placing along each profile at each target object boundary position the cluster mean most popular there in the training set, where location in the training set is given in object-relative coordinates $\underline{u}$ that are detailed in section 3.2.

The match of the template stencil to the stencil made from the object or object ensemble relative target image is today computed as a normalized correlation of these intensity arrays (the normalization involves being offset by the mean intensity of the respective profile and being divided by the rms value in the region of match).

2.4. The user's view of our segmentation method

Using the mechanisms described in the previous sections, our software system for segmentation, "Pablo" carries out successive optimizations to segment an object ensemble (possibly containing only 1 object) (see Fig.1). The user first chooses the model and places it in an initial position relative to the 3D image viewed in interactively controllable tri-orthogonal slices (for example, see Figs. 3b, 4-bottom row, and 9-bottom left). This view is maintained on the screen during the whole segmentation session. In addition, a sequence of display boxes appears, one after the other for each scale level. The box names the scale level, tells which entity is being

optimized, and gives objective function component values for each iteration of the optimization. It also gives the values of the two weights being used in the objective function; these are described in section 4.1. Weights that work best for segmenting this target object from this type of image are experimentally found . Users can change these weights but are encouraged not to.

At all of the scale levels, we follow the strategy of iterative conditional modes, so the algorithm cycles among the component entities in random order until the group converges[1]. For example at the figural atom stage, the algorithm cycles through the atoms in random order. The user sees the object continually change relative to the image slice presently appearing. An unlimited undo and redo feature allows the user to back the optimization up to a particular stage, at which he can edit the model, or to view the sequence of changes that have occurred during the successive optimizations.

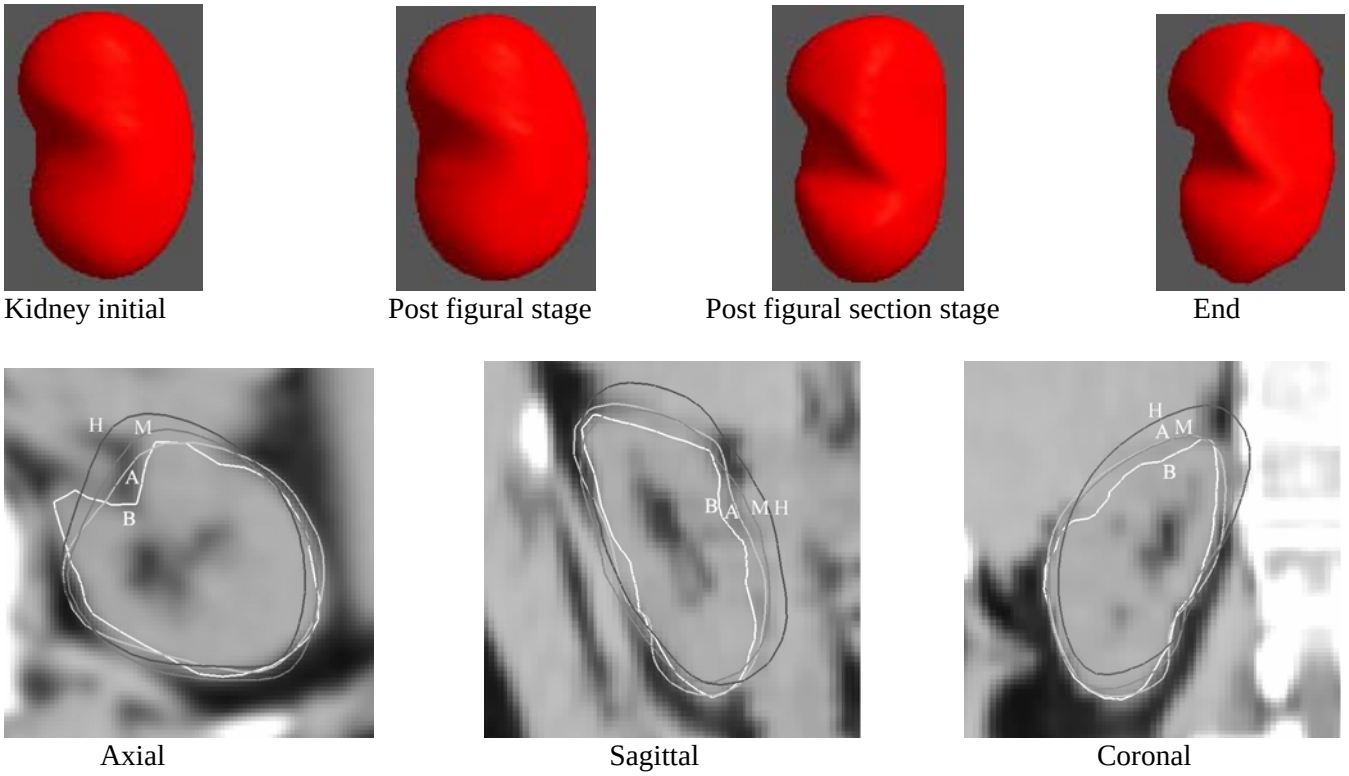

Fig 4. Stage by stage progress of deformable m-rep segmentation of the kidney. Top: rendered 3D view, after the stage indicated. . Bottom: results on axial, sagittal and coronal CT slices. Each image compares progress through consecutive stages via overlaid curves: H - initial (hand-placed) position; M – post figural model stage; A – post figural section (atom) stage; B – at end, i.e, post boundary displacement stage.

[1] The convergence properties are shared with all iterative conditional modes methods and are based on the underlying Markov random field. In practice, convergence always occurs, but sometimes the convergence is to a local maximum of the objective function rather than the desired global maximum.

2.4.1. Viewing an m-rep

M-rep-defined objects can be viewed as a boundary mesh (at any of a number of vertex spacing levels), a rendered surface, a collection of points at the aforementioned boundary vertices, or a medial atom mesh. Most users find the first two of these the most useful. Images are normally viewed in a tri-orthogonal display, with the three possible slice directions fixed to the cardinal within-image and cross-image slice directions given by the stored target image. Any subset of the slice directions are selectable, and the user can control which slice is shown in each of the selected directions, as well as an intensity window. The intensity window selected is applied to the image before segmentation takes place.

The displayed object can be presented together with the intensity display (see Fig. 4). In this case the intensity slice is taken to be opaque, typically occluding part of the object. However, the user can control his viewpoint on this whole complex, so the slices presented, as well as the object can appear to be seen from the front or back or at any angle to the viewer. Moreover, we also provide a boundary display mode on the displayed slice, in which the 3D object does not appear but the curve of its intersection with the displayed slice(s) is displayed on that slice (those slices).

2.4.2. Initially Locating and Editing an M-rep

Using the same controls as for choosing the viewing pose but with an additional key depressed, the user can manually translate, rotate, or magnify the m-rep within the image space. This capability, together with the triorthogonal intensity display just described, allows easy location of an m-rep into the initial position from which it should deform.

When a base m-rep for a particular anatomic object is built, certain points in its space, typically on its implied boundary, can be specified as landmarks. These landmarks appear as colored spots in the display space. One way to place an m-rep in its initial position is to approximately locate these points within the slices of the image, using the tri-orthogonal display. A Procrustes algorithm is then used to translate, rotate, and magnify the m-rep within the image space such that the landmarks are as close as possible to the selected positions.

These landmarks can also be used for editing an m-rep in the middle of the optimization process. Editing is also possible medial atom by medial atom, but this operation is used only for research and training purposes.

3. Representation of Object Ensembles by M-reps: Sampled Medial Representations Augmented by Boundary Displacements

In this section we briefly describe m-reps, the geometric representation scheme that has the properties listed in section 1 as necessary to support an efficient, effective segmentation method for anatomic objects based on posterior optimization. The reader interested in this segmentation method principally as a user can read this section lightly, if at all.

3.1. M-reps and their geometry

M-reps leverage the strengths of medial representations identified by many authors, in image analysis, geometry, human vision, computer graphics, and mechanical modeling, starting from Blum [1967]. It allows one to distinguish object deformations into along-object deviations, namely elongations and bendings, and

across-object deviations, namely bulgings and attachment of protrusions or indentations. An additional advantage is that distances, and thus spatial tolerances, can be expressed as a fraction of medial width. These properties allow positions and orientations to be followed through deformations of whichever level of locality (scale). Because geometric typicality requires comparison of corresponding positions of an object before and after deformation and because geometry-to-image match require comparison of intensities at corresponding positions, this ability to provide what we call a *figural coordinate system* is advantageous in segmentation by deformable models. Among the strengths of this coordinate system is that the test for whether a point is interior or exterior to a figure is computationally trivial, so the need for testing for and penalizing interpenetration of the boundary of one object into a nearby object is easily met.

An additional useful property of medial representations is that they divide an object ensemble into objects and objects into *figures*, i.e., slabs with an unbranching medial locus (see Fig. 5). This allows the modeler to subdivide anatomic objects into the very units of medical interest, units of relatively uniform tissue properties. It also provides the opportunity for segmentation at multiple levels of scale, from large scale to small.

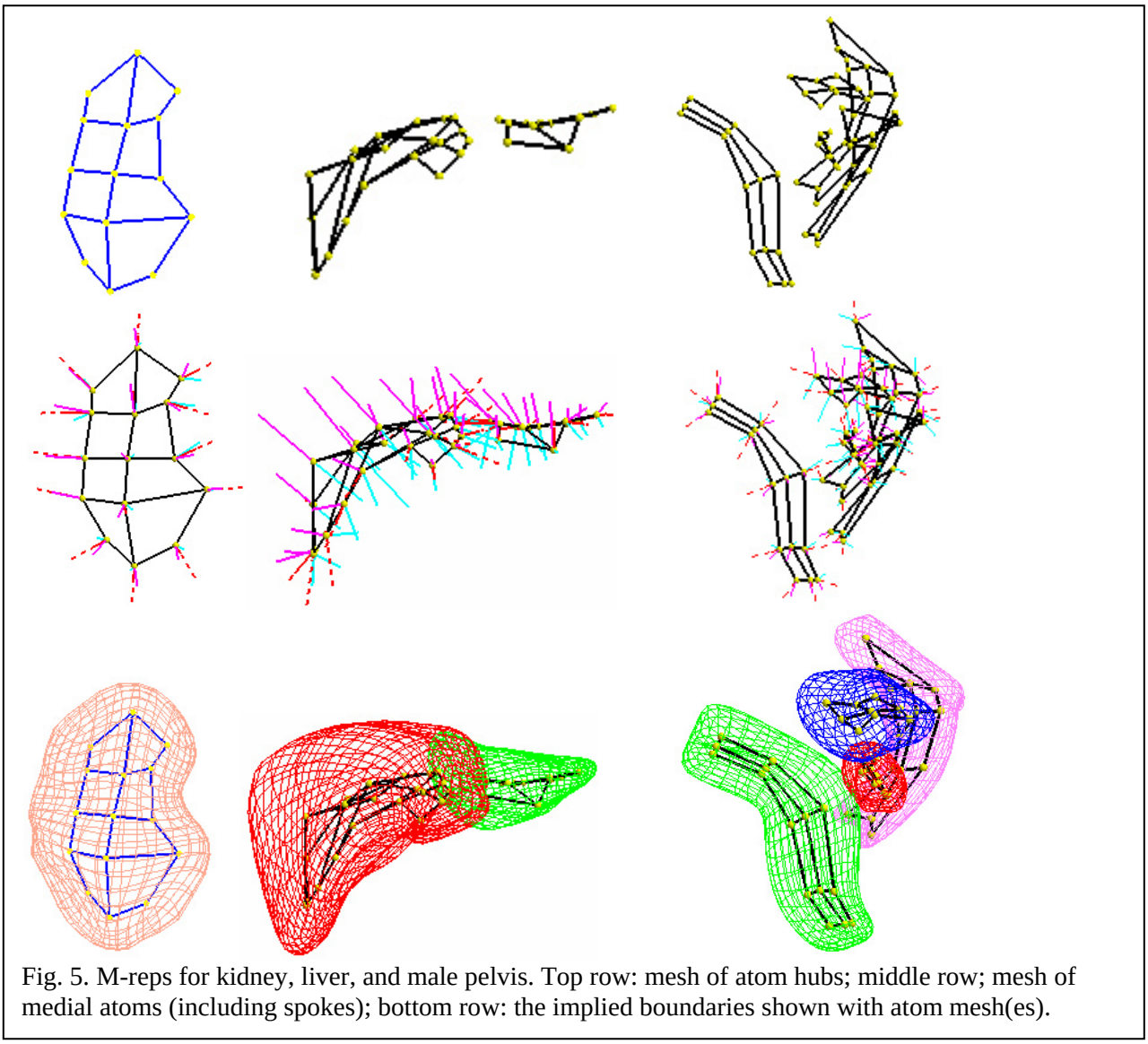

Fig. 5. M-reps for kidney, liver, and male pelvis. Top row: mesh of atom hubs; middle row; mesh of medial atoms (including spokes); bottom row: the implied boundaries shown with atom mesh(es).

In our m-reps representation described in [Pizer 1999, 2003, Joshi 2001] one or more meshes of medial atoms implies a boundary of an object or an ensemble of objects, where each boundary is tiled at a spacing that is fine compared to the mesh of medial atoms. Moreover, at each boundary vertex a displacement along the boundary normal is provided.

Each medial atom should be thought of as a hub and two equal length spokes (Fig. 6). The atom forms an infinitesimal through-section of the interior of the object. The two opposing sections of boundary implied by the atom pass incident to and orthogonal to the two spokes. The medial atom captures translation of the local interior section by translation of the hub, bending or twisting of the local interior section by rotation of the atom, magnification of the local interior section by increase of the length of the spokes, and change of the angulation of the two opposing boundary sides by the rotation of the spokes towards or away from each other in their common plane.

For slabs a sequence of end atoms forms a crest region closing the slab. As illustrated in Fig. 6, these segment closed ends may be rounded with any level of crest sharpness η: a new spoke is placed along the bisector of the two ordinary atom spokes, with length ηr. We use this formulation for ends in order to stabilize the image match at ends as compared to the possibility where the medial end atoms determine a single crest point.

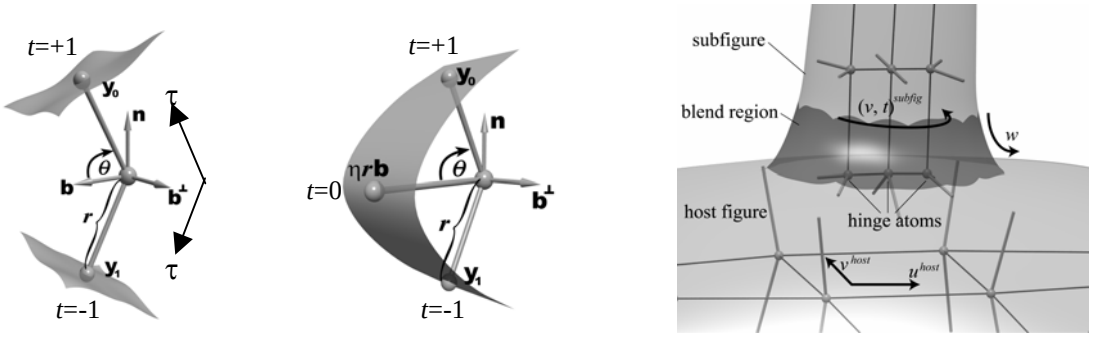

Fig. 6. Left two: medial atoms, made from a hub position ***x*** and two spokes of equal length *r*. The atom on the left is for an internal mesh position, implying two boundary sections. The atom on the right is for a mesh edge position, implying a section of boundary crest. An atom is represented by the medial hub position ***x***; the length *r* of the spokes; a frame made from the unit-length bisector $\vec{b}$ of the spokes, the $\vec{b}$ -orthogonal unit vector $\vec{n}$ in the plane of the spokes, and the complementary unit vector $\vec{b}^{\perp}$; and the "object angle" θ between $\vec{b}$ and each spoke. The implied figure section is slab-like, and the local implied boundary is incident to and orthogonal to the spoke ends. Right: the figure / subfigure arrangement, with the subfigure (6 medial atoms appearing, 3 of which form the "hinge") on top, the host figure (4 medial atoms showing) below, and the blend region shown darker.

3.2. Figural coordinates

Both the geometric typicality and the geometry-to-image match are typically measured in reference to the object boundaries. As a result, we describe locations in object relative terms via the location on the medial sheet to which they correspond, the side (*t*) of the medial locus at which they fall, and the medial radius (spoke length *r*) proportional distance (τ) they are from the implied boundary (details can be found in [Pizer, 2003]). Together these give the object-relative coordinates $\underline{u}$ for that figure. Segmentation methods that access the image data via an object relative parametrization must be able quickly to compute the boundary positions and normals as the candidate object ensemble changes under the successive geometric transformations applied during the optimization. In our framework this transformation between the image's $\underline{x}$ coordinates and the

object-relative $\underline{u}$ coordinates is done through the mechanism of subdivision surface methods [Thall 2002], which are discussed further in section 3.3.

We call the object or object part corresponding to a single mesh of medial atoms a *figure*. Each object has a main figure, but protrusions from or indentations into any other figure can be used to form more complicated objects, such as the two-figure liver shown in Fig. 5. Also shown in Fig. 6 and in the liver example in Fig. 5, the hinge medial atoms of the subfigure ride on the implied boundary of the host figure, and the object boundary is formed from three pieces that fit smoothly together: the host figure with a hole taken out for the subfigure, the subfigure cut off before it intersects with the host figure boundary, and a smooth blend placed between the cut-open boundaries of the two figures. The amount of dilation of the subfigure used to make the hole in the main figure and the amount by which the subfigure is cut off are two parameters of the object. The object-relative coordinates $\underline{u}$ of the blend region are $\underline{u} = (v,w,t,\tau)$, where v and t are the cross-figure figural coordinates of the subfigure and $w \in [-1,1]$ moves along the blend from the curve on the subfigure terminating the blend (w=-1) to the curve on the host figure terminating the blend (w=+1).

Sometimes one wishes to represent and then segment multiple disconnected objects at the same time. An example, shown in Fig. 5, is the prostate, rectum, bladder, and pubic bones in the male pelvis in which the structures are related but one is not a protrusion or and indentation on another. Another example are the pair of kidneys and the liver. In our system these can be related by one or more connections between the representations of the respective objects, allowing the position of one figure to predict boundary positions of the other. The details are left to a paper covering the segmentation of object ensembles [Fletcher 2002].

3.3. Figural boundary computation

An m-rep mesh of medial atoms for a single figure should be thought of as a representation for a continuous sheet of atoms and a corresponding continuous implied boundary. The boundary and the sheet of atoms are interpolated from the medial atom samples in the mesh. Since the geometry to image match term in the objective function requires boundary-relative image positions to be computed in large number, this interpolation must be very efficient, so we have taken great pains to achieve such efficiency. In Pablo at present, we interpolate the boundary first and consequently can interpolate medial atoms at any position on the sheet of atoms. The implied boundary is computed from the set of atom spokes connected into quadrilateral and triangular tiles (Fig. 7). Outside the blend regions between a subfigure and its host figure quadrilateral boundary tiles are formed from the spoke ends of the medial atoms in those figures. The blend region is made of triangular and quadrilateral tiles. The boundary interpolation is accomplished by a variation of the very efficient Catmull-Clark subdivision [Catmull 1978] of the mesh of polygonal tiles. Thall's variation [2002] of Catmull-Clark subdivision produces a limit surface that iteratively approaches a surface interpolating in position to spoke ends and with a normal interpolating the respective spokes. That surface is a B-spline at all but finitely many points on the surface. The program gives control of a tolerance on the normal and on the closeness of the interpolations.

The result of such an interpolation is that with each boundary position we can associate a *boundary figural coordinate* $\underline{u}$, which for non-blend regions is (u, v, t) [τ=0 at the boundary, by definition] and for blend regions is (v, w, t). Moreover, for any boundary point given by $\underline{u}$ a medial atom interpolated from the integer (u,v) values of the surrounding input points can be computed, with its normal and associated r value. Also, an arbitrary fine triangular tiling of the medially implied boundary can be computed. Rendering is based on these triangular tiles.

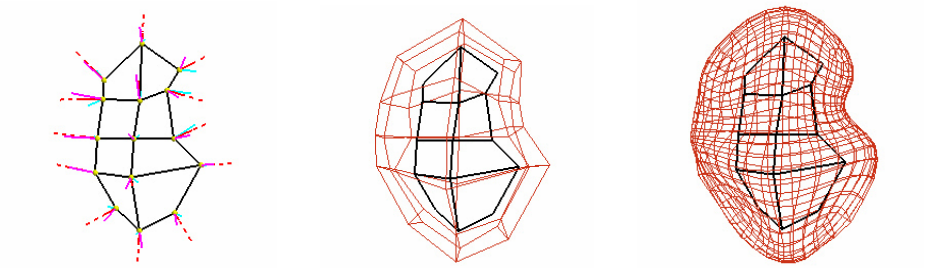

Fig. 7. A single-figure m-rep for a kidney (left) and the boundary points implied by its spoke ends (middle). Right: interpolated boundary mesh, from which surface rendering can be done.

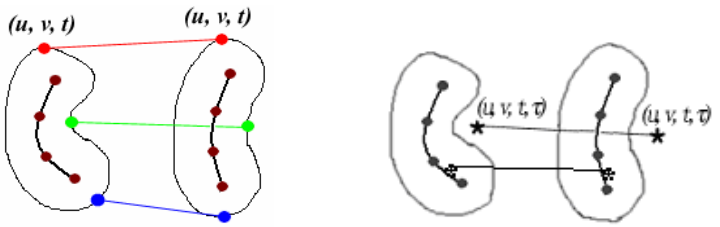

Fig. 8. Correspondence over deformation via figural correspondence. In each pair of marked points connected by lines, the two points have the same value of the figural coordinate $\underline{u}$.

Points (x,y,z) in space can also be given a figural coordinate by adding an r-proportional distance to the figural coordinates of the closest medially implied boundary point. To allow the distinction by the sign of the distance of the inside and the outside of the figure, we take the distance to be relative to the medially implied boundary and to be negative on the interior of the figure. Pablo includes a procedure mapping arbitrary spatial positions $\underline{x} = (x,y,z)$ into figural coordinates $\underline{u} = (u,v,t, \tau)$ or (v,w,t, τ). As well, the correspondence under figural deformation given by figural coordinates is critically useful in computing the objective function used in segmentation (Fig. 8).

At each implied boundary vertex implied by the medial atoms in an m-rep, the m-rep also gives a displacement along the implied normal. This displacement is given as a τ value, i.e., in medial-radius-proportional units. Computing these displacements is done at the smallest scale level of the model deformation.

4. Details of the Segmentation Implemented in Pablo

4.1. Actually Used Terms in Objective Function

We are on a path at the end of which our objective function will be a log posterior probability. Indeed, as described in section 6, we have designed and implemented means of estimating probability function parameters for all of the objective function terms from a training set of images and their manual segmentations. We have also implemented trial forms to evaluate these log probabilities for a particular m-rep candidate and target image. When these computations become routine, the geometric typicality would be computed as a log prior, and the geometry-to-image match would be computed as a log likelihood.

However, in the form of Pablo that was used in the evaluation experiments sketched in section 5 and detailed in a companion paper, our geometric typicality term is measured in *r*-proportional squared distances from model-predicted positions and the geometry-to-image match term is measured in rms-proportional intensity squared units resulting from the correlation of a template stencil image and the target image stencil, normalized by local variability in these image intensities. While this strategy allows the objective function to change little with image intensity scaling or with geometric scaling, it leaves the necessity of setting the relative weights described in section 4.1.

The template to image match correlation measure that we presently use is an average, over the boundary vertex points, of the along-spoke intensity profile correlations at these sample points. These profiles are sampled with an equal spacing in the figural distance coordinate τ. For the geometry to correspond to the volume integral of these point-to-corresponding-point correlations, each profile must be weighted by the boundary surface area between it and its neighboring sample points, and the profile must be weighted by its r-proportional length. In addition, as indicated above, we weight each product in the correlation by a Gaussian in τ from the boundary (so the Gaussian in image space has a standard deviation increasing with the figural width). Also, to make the intensity profiles insensitive to offsets and linear compression in the intensity scale, each template intensity profile is offset to a mean of zero and both the template profiles and the target image profiles are rms-normalized within their respective stencil regions. For the kidney results sketched in section 5, the standard deviation in units of τ of the Gaussian template is 0.1, the stencil is cut off at $\tau = 0.3$, and the standard deviation of the weighting Gaussian in the intensity correlation is 0.15.

4.2. Weight parameter choices

At each scale level stage the objective function involves two weights. The first multiplies the geometric typicality, so choosing its value larger makes the object more resistant to geometric change. Typically its value is increased as the signal-to-noise ratio of the object to background contrast decreases. The second weight determines the relative importance of changes in the target entity shape and changes in its geometric relation to its neighboring entities, in forming the geometric typicality term. At the figural section (medial atom) scale level, this term maintains the shape of the atom mesh and as well maintains the relative orientation and magnification of neighboring atoms.

We have found that for a given target object and image type it is possible to fix the values of these weights. However, the user interface allows the user to change them for each geometric entity being optimized. When both the geometric typicality and the geometry-to-image match metrics become log probabilities, the use of the weights may well become unnecessary.

5. Segmentation results

Results of a controlled study on segmentation of kidneys from 3D CT images of clinical quality are given in a companion paper [Rao 2004]. Roughly, the results are as follows. Summarized over all the kidneys, the disagreement between our segmentation and that of either of two humans is about the same as that between the two humans. In 22 of the 24 cases, the automatic segmentation would require no editing for use in radiotherapy treatment planning. In fact, the automatic segmentations are frequently judged to be superior to the manual segmentations, and they have the additional benefit of being reproducible, even if the initialization is slightly different.

M-rep models have been built for both the liver [Han 2003], a multifigure object, and the heart [Pilgram 2003], a multi-object ensemble, and statistical description of these anatomic entities have generated. As illustrated in Fig. 9, 1) we have built a model for the multi-object ensemble of bladder, prostate, and pelvis in the male pelvis, all relative to m-rep models of the pubic bones; 2) we have fit it into a few dozen binary segmentations of these organs from fraction-by-fraction CT images in three patients' cases; 3) we have built multi-object statistical descriptions of the variability of the male pelvic anatomy across fractions within a particular patient; and 4) we have used these statistics to successfully segment these organs from CT images in later fractions. The details of this multi-object segmentation for adaptive radiotherapy must await a separate paper, but it was published as an abstract and presented orally in [Chaney 2004]. Our work on segmentation of these objects depends principally, now, on our ongoing work on training the object-relative image intensity statistics that are needed for the geometry-to-image match term in the objective function.

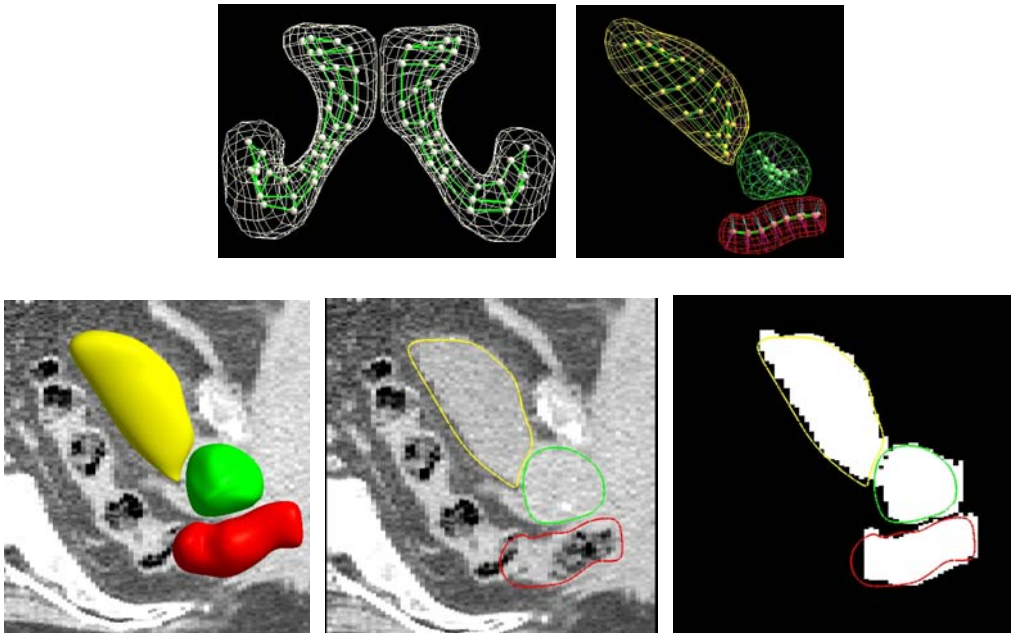

Fig. 9. M-reps for segmenting the male pelvis in CT images in later radiotherapy fractions. Top left: m-rep for pubic bones, used to register the later day fraction images with the first day fraction. Top-right: the m-rep for the bladder, prostate and rectum. Bottom left: a visualization of the bladder, prostate rectum m-rep's implied boundaries relative to a slice of the associated 3D CT image. Bottom middle and right: the segmentation result in a later fraction, shown in one of the image slices first vs. the greyscale CT image and then vs. the human segmentation shown in white.

5.2. Speed of computation

The speed of a 3D segmentation on a Pentium IV, 1.7GHz computer subdivides as follows.

- Preprocessing computations take less than 1 second.
- The largest scale stage (the object ensemble stage for an ensemble, the object stage for a single multifigure object, the figure stage for a single-figure object) takes less than 1 second per iteration and on the average requiring 6 iterations for a total time of under 10 seconds to determine the similarity transform and geometric warp coefficients. Similar numbers apply to each application of the object stages and figure stages, so the total time is proportional to the number of objects plus the number of figures.
- Each atom transformation stage takes on the average approximately 7 seconds per atom, and typically 3-4 passes through an average atom are required. In our kidney model there are 15 atoms, and in our liver model there are 33 atoms.
- The boundary displacement stage takes approximately 5 seconds.

Thus the total time for a kidney segmentation is typically 5-7 minutes.

While the method's speed has already benefited strongly from moving much of the computation from the deformation stage to the model building stage, there is still much room for speedup of integer multiples by more medial levels of coarse to fine, by medial deformation measured directly from the atoms without resort to the implied boundary, by having the gradients of the objective function relative to the changing parameters needed by the optimization steps be computed analytically rather than with numerical derivatives, and just by more careful coding.

6. Building M-rep Models and Associated Intensity Templates

In this section we briefly describe the means of building the models of m-reps and the associated intensity templates from training data. The reader interested in this object representation and segmentation method principally as a user can read this section lightly, if at all.

6.1. M-rep model building for anatomic objects from training segmentations

Model-building must name the figures making up an object or object ensemble, give the size of the mesh of each figure, and give the way the figures are related. It must also specify each medial atom in the model forming the mean object or object ensemble and the variability of these at many scale levels. Illustrated in the panels of Fig. 5 are m-rep models of a variety of anatomic structures that we have built. In the following we sketch our model building procedure, leaving the details of how we meet this challenging goal to other papers.

Because an m-rep is intended to allow the representation of a whole population of an anatomic object across patients, we build it based on a significant sample of instances of the object. Typically we use some tens of instances, say 50. For each instance we begin with both a 3D binary image representing the interior of the object, typically manually segmented, and an associated 3D greyscale image (CT or MRI or …).

Styner [2001] describes a tool for stably producing m-rep models from such binary image samples, based on the principle that effective segmentation depends on building a model that can easily deform into any instance of the object that can appear in a target image. We can use this tool to tell us which figures should be used, or we can choose the figures based on anatomic expertise to correspond to named anatomic structures. In either case,

given the figures, the tool chooses the number of atoms in each figure as the minimal number that can fit every training instance to a given error in volume overlap (typically 98%).

Once a base model is generated, we use Pablo to deform it into the binary segmented training images. Given the m-rep models for all the training cases (Fig. 10), we use a tool initially developed by Fletcher and Dam [2003] and further developed by Lu to compute the mean model and its variability (Fig. 10). Understanding these statistics requires a brief discussion of how to do statistics of geometric transformations that simultaneously describe local translations, rotations, and magnifications [Fletcher 2004]. The important point is that translations combine linearly, but rotations and magnifications combine nonlinearly, so the geometric transformation of a base m-rep describing each m-rep should be considered as a point on a nonlinear manifold, rather than the flat manifold in which linear statistics is normally done. Shortest distance paths between two m-reps on this nonlinear manifold are called geodesics, with geodesic distances related to integrated distances of the implied object boundary. If the m-reps' atoms correspond, the length of a geodesic path measures the amount of deformation between the m-reps, the mean is a point on the surface that is closest in sum of square geodesic distances to the training m-reps, and there are geodesic paths through the mean that are counterparts to principal components in linear statistics. As with linear statistics, each principal geodesic has an associated variance, and moving along that geodesic gives a principal mode of variation of the population of m-reps.

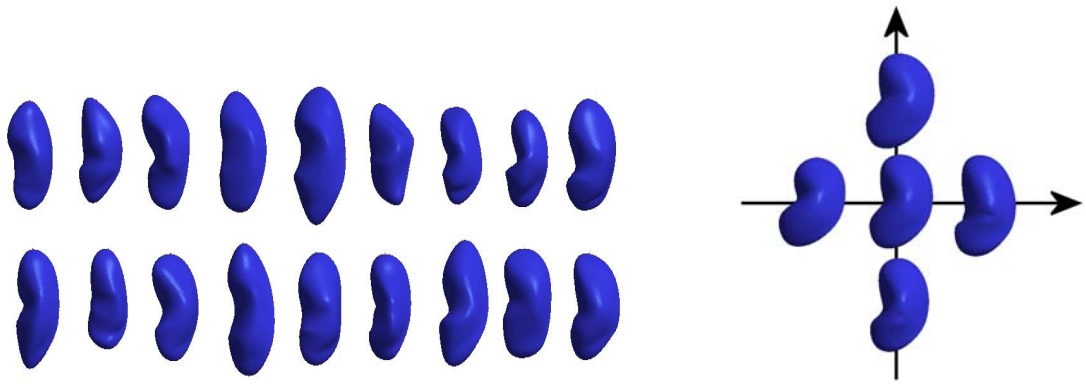

Fig 10. Left: A subset of our population of training kidneys. Right: the mean of the population and the mean ± 1 standard deviation in each of the first two principal geodesic modes.

Means and principal geodesics can be computed not only for m-reps but also for similarity transforms applied at a given scale level. Also, the statistics at one scale level need to describe the variability of the geometric entity at that scale level after the variability at the larger scale levels has been accounted for. Description of this residue statistics, based on theory of Markov random fields, is given in [Lu 2003].

With these means and a number of principal geodesics chosen to capture some fraction of the variance at that scale level, deforming a geometric entity at that scale level involves computing the parameters of the principal geodesics of the base transformation (e.g., a similarity transformation) plus the parameters of the principal geodesics of the deformation of that entity.

6.3. Generating intensity statistics

Objects that appear with high contrast over most of their boundary can be extracted using a Gaussian derivative template with the scale of the derivative proportional to the local object width *r*. For the remainder, in fact for almost every object, the pattern of image intensity profiles normal to the boundary in the boundary-centered stencil as a function of (figurally defined) boundary position are frequently characteristic. As a result a statistical intensity template as function of figural position drawn from the set of training images is needed.

We operate in a way analogous to the boundary point based segmentation methods of Cootes & Taylor [1993], except making intensity profile correspondences figurally defined boundary position by boundary position rather than designated boundary point by boundary point. Roughly, at profile position u we do statistics of the intensity profiles at u in the training greyscale images relative to m-reps fit to each case.

At present, we cluster all the normalized profiles, choose the dominant clusters, and represent each cluster by its mean. At each boundary vertex, the prevalence of each cluster is then recorded, where the vote of each case for the prevalence of a cluster at that boundary vertex is weighted by the normalized correlation value with the cluster mean. We presently simply use the mean of the most prevalent cluster in the template stencil, but we are developing a means of computing an actual log probability of image match, using the fits of the target image profile to each of the cluster means.

7. Discussion and Conclusions

The main objective of this paper was to sketch a method and software for 3D image segmentation intended for radiation treatment planning using m-reps both to capture prior knowledge of object geometry and as the basis of measurement of geometry-to-image match, and allowing efficient, multiscale operation. We have described the basic components of a method that is based on a) representation of the interior of an object through m-reps, a medial representation augmented by boundary displacements, b) the multiscale subdivision of an object ensemble into objects, then figures, then interior figural through sections, and then boundary locations, and c) the successive optimization of an objective function describing both geometric typicality and geometry-to-image match in the object-relative coordinate system that m-reps provides. We have given a view of the operation and user-interface of the program Pablo that implements this method. The initial successes of the segmentation method provide evidence of the usefulness of these ideas. Our quantitative validation on kidney segmentation is encouraging: robust and accurate enough for clinical use. We are beginning a test on segmentations from the daily clinical stream for radiation treatment planning. Controlled, quantitated validations on other objects are needed.

M-reps provide one means of modeling objects and collections of objects; boundary representations (b-reps) are a common alternative means of such object modeling. They share the limitations of all object modeling methods, namely that a single object model will not serve for a class of objects with mixed topologies at the figural level. However, because they explicitly model the interfigural relations, they have special weaknesses when these relations are variable over the population of objects. For example, an m-rep for a right kidney and a separate left kidney will not perform well for a horse-shoe kidney, in which the kidneys are joined. For such mixed classes, a separate m-rep is required for each exemplar. Another issue shared with other object models is instability for nearly spherically or circularly symmetric objects. In such cases the nearly degenerate geometry creates computational instabilities in discriminating among the three major axes which in turn can cause an m-

rep to "flip" during deformation in the image data in an unstable manner. However, m-reps share with other object models the particular strength of resolving these orientational instabilities via the relations among objects.

M-reps' special abilities relative to b-reps derive from their explicit representation of object orientation changes such as twisting and bending and of object size changes such as widening and narrowing. Thus statistics on rectal widenings due to gas, on the variability in the relative pose of the two lobes of the liver, and on the orientation of the bladder relative to the prostate are very effective in m-reps terms. The limitations not of m-reps by themselves but of m-reps with statistics come in situations when the orientational or magnificational relationships are very variable. Thus, like b-reps m-reps are well suited to complex slabs and tubes such as the cerebral cortex or the intestine, and both are well suited to intra-patient variations of these structures over time. But because in the population of humans the variability of the folding structure of the cerebral cortex is high and the variability of the curvature of the intestine is high, statistics on m-reps is a weak tool over that population for these structures.

Finally, because m-reps represent the interior of objects, they lose their effectiveness in image situations where only one side of an object appears in an image, and they have weakness relative to b-reps in situations where one side of an object boundary is statistically stable but the other side has great variability. In that situation b-reps can ignore the unstable or unimaged side, whereas m-reps inherently must represent both sides together.

In a separate paper [Crouch 2003] we have shown how the space parametrization provided by m-reps also allows the interior of the object to be divided into mesh elements useful for efficient mechanical modeling of intra-patient motion of anatomic structures due to such interventions as intrarectal imaging probes. The measures of mechanical energy computed in this approach could be used for segmentation of a patient whose segmented m-rep from an earlier (e.g., planning) image can be used as the model for a segmentation in a later (e.g., intra-treatment) image.

We are in the process of making the following improvements to our deformable m-reps segmentation method and software:

1) Automatic initialization based on a model of the body cross-section.
2) Sensing and reporting locations on the segmented object that do not have the expected level of geometry-to-image match, so that the user can take actions of relocating that object section and then restart the segmentation.
3) Grids of medial atoms at multiple levels of fineness, to handle variations at larger scales before refining them to handle more local variations.
4) Routinely using log probabilities and residues for both the geometric typicality and the geometry-to-image match terms in objective function being optimized.
5) Adding the option of computing a small scale diffeomorphism (space warp) both in the target object(s) and in the interstitial space between objects in place of the small scale boundary displacement.
6) Using deformable stencil profile curves that depend on the relative positions of neighboring objects in object ensembles.
7) Developing a form of our software intended for clinical use and thus being as automatic as possible, and making all interactions in clinical terms.

Acknowledgments

This work was done under the partial support of NIH grants P01 CA47982 and P01 EB02779, NSF SGER Grant CCR-9910419, and a fellowship from the Link Foundation. A gift from Intel Corp. provided computers on which some of this research was carried out. We thank Qiong Han, Conglin Lu, and Delphine Bull for help with models, figures, and references.

References

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

Catmull, E, J Clark (1978). Recursively generated B-spline surfaces on arbitrary topological meshes. *Computer Aided Design* **10**: 183-188.

Chaney, E, S Pizer, S Joshi, R Broadhurst, PT Fletcher, G Gash, Q Han, JY Jeong, S Kalra, J Levy, C Lu, D Merck, J Rosenman, J Stough, G Tracton (2004). Automatic male pelvis segmentation from 3D CT images via deformable m-rep models. Proc. ASTRO 2004.

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

Crouch, J, SM Pizer, EL Chaney, M Zaider (2003). Medial techniques to automate finite element analysis of prostate deformation. To appear in *IEEE-TMI*. Available at: http://midag.cs.unc.edu/pubs/papers/TMI04-Crouch-FEM.pdf

Dam, E, PT Fletcher, SM Pizer, G Tracton, J Rosenman (2003). Prostate Shape Modeling Based on Principal Geodesic Analysis Bootstrapping. IT University of Copenhagen technical report number D-2003-03. Chapter 9 of Erik Dam's dissertation entitled, Committing Medical Image Analysis. Available at: http://midag.cs.unc.edu/pubs/papers/Dam_Ch9.pdf

Delingette, H. (1999). General object reconstruction based on simplex meshes. *International Journal of Computer Vision*, **32**:111-146.

Fletcher, PT, SM Pizer, G Gash, S Joshi (2002). Deformable M-rep segmentation of object complexes. *IEEE International Symposium on Biomedical Imaging (ISBI)*, CD proceedings, 26-29. Also available at the bibliography webpage of http://midag.cs.unc.edu

PT Fletcher, C Lu, SM Pizer, and S Joshi, Principal geodesic analysis for the study of nonlinear statistics of shape (2004). *IEEE Transactions on Medical Imaging*, **23** (8): 995-1005, IEEE, Aug. 2004.

Han, Q, C Lu, G Liu, SM Pizer, S Joshi, A Thall (2003). Representing multi-figure anatomical objects. Submitted to *International Symposium on Biomedical Imaging 2004 (ISBI '04)*. Available at: http://midag.cs.unc.edu/pubs/papers/ISBI04-Han_multi-fig.pdf

Joshi, S, S Pizer, PT Fletcher, A Thall, G Tracton (2001). Multi-scale 3-D deformable model segmentation based on medial description. *Information Processing in Medical Imaging 2001 (IPMI '01)*. Lecture Notes in Computer Science, Springer **2082**: 64-77.

Kelemen, A, G Szekely, G Gerig (1999) Three-Dimensional Model-based Segmentation, *IEEE-TMI*, **18**(10):828-839.

Lu, C, S Pizer, S Joshi (2003). A Markov Random Field Approach to Multi-scale Shape Analysis. Scale Space Methods in Computer Vision (Scale Space 2003). LD Griffin and M Lillholm, eds. Lecture Notes in Computer Science **2695**: 416-431.

McInerny, T, D Terzopoulos (1996). Deformable models in medical image analysis: a survey. *Medical Image Analysis*, **1**(2): 91-108.

Pilgram, R, PT Fletcher, SM Pizer, O Pachinger, R Schubert (2003). Common shape model and inter-individual variations of the heart using medial representation: a pilot study. Tech report, Institute for Medical Knowledge Representation and Visualization, University for Health Informatics and Technology, Tyrol, Austria.

Pizer SM, D Fritsch, V Johnson, E Chaney (1999). Segmentation, registration, and measurement of shape variation via image object shape. *IEEE Transactions on Medical Imaging*, October 1999. **18**(10): 851-865.

Pizer, S, PT Fletcher, S Joshi, A Thall, Z Chen, Y Fridman, D Fritsch, G Gash, J Glotzer, M Jiroutek, K Muller, G Tracton, P Yushkevich, E Chaney (2003). Deformable M-reps for 3D Medical Image Segmentation. IJCV Special UNC-MIDAG issue, **55**(2/3): 85-106.

Rao M, Stough J, Chi Y-Y, Muller K., Tracton GS, Chaney EL, Pizer SM (2004). Comparison of Human and Automatic Segmentations of Kidneys from CT Images. *International Journal of Radiation Oncology, Biology, Physics*. In press.

Staib, LH, JS Duncan (1996). Model-based Deformable Surface Finding for Medical Images. *IEEE Trans. Med. Imaging* **15**(5): 1-12.

Stough, J, SM Pizer, EL Chaney, M Rao (2003). Clustering on image boundary regions for deformable model segmentation. Submitted to *International Symposium on Biomedical Imaging 2004 (ISBI '04)*. Available at: http://midag.cs.unc.edu/pubs/papers/ISBI04-Stough-clustering.pdf (2004).

Styner, M, G Gerig (2001). Medial models incorporating object variability for 3D shape analysis. *Information Processing in Medical Imaging 2001 (IPMI '01)*. Lecture Notes in Computer Science, Springer **2082**: 502-516.

Thall, A (2002). Fast C^2 interpolating subdivision surfaces using iterative inversion of stationary subdivision rules. Tech. Report 02-01, Univ. of NC, Dept of Computer Science, accessed at http://midag.cs.unc.edu/pubs/papers/Thall_TR02-001.pdf