

Construction of an Abdominal Probabilistic Atlas and its Application in Segmentation

Hyunjin Park, Peyton H. Bland, and Charles R. Meyer*, *Member, IEEE*

Abstract—There have been significant efforts to build a probabilistic atlas of the brain and to use it for many common applications, such as segmentation and registration. Though the work related to brain atlases can be applied to nonbrain organs, less attention has been paid to actually building an atlas for organs other than the brain. Motivated by the automatic identification of normal organs for applications in radiation therapy treatment planning, we present a method to construct a probabilistic atlas of an abdomen consisting of four organs (i.e., liver, kidneys, and spinal cord). Using 32 noncontrast abdominal computed tomography (CT) scans, 31 were mapped onto one individual scan using thin plate spline as the warping transform and mutual information (MI) as the similarity measure. Except for an initial coarse placement of four control points by the operators, the MI-based registration was automatic. Additionally, the four organs in each of the 32 CT data sets were manually segmented. The manual segmentations were warped onto the “standard” patient space using the same transform computed from their gray scale CT data set and a probabilistic atlas was calculated. Then, the atlas was used to aid the segmentation of low-contrast organs in an additional 20 CT data sets not included in the atlas. By incorporating the atlas information into the Bayesian framework, segmentation results clearly showed improvements over a standard unsupervised segmentation method.

Index Terms—Abdomen scan, probabilistic atlas, segmentation, thin plate spline (TPS).

I. INTRODUCTION

RECENTLY, significant effort has been directed toward the development of deformable templates typically for segmentation of the human brain [1]–[11]. Only one recent publication addresses segmentation of liver and kidneys using a deformable surface model [12]. Rich literature has been compiled on probabilistic atlases at the brain [13]–[23]; while most are extensible to abdominal organs, their extensions exist only in theory and have not been demonstrated on abdominal organs. Our earlier work with the construction of a probabilistic liver model taught us that the loci of liver edges were constrained by the presence of other organs as well, i.e., liver edges would not be found inside of the adjacent right kidney [24]–[26]. Thus, we

have pursued the simultaneous construction of a probabilistic atlas that includes a set of abdominal organs similar to that in [27]. By the term “probabilistic atlas,” we do not simply mean the average boundary of an organ, but the complete spatial distribution of probabilities that a voxel belongs to one or more organs, i.e., each voxel is an n -vector, where n depends on the number of organs in the modeling system. We continue to focus on probabilistic atlases in spite of the remarkable, even spectacular results of deformable templates, because of the belief that atlases, if well formed, bring more prior information to the process of defining complex organs in low signal-to-noise settings. After mapping a new patient onto the atlas using the same methods used to construct it, the computed inverse transform maps the probability distributions of the atlas back onto the patient. Then, the atlas can be used to find the most probable loci of edges of the patient’s organ. Furthermore, the derivative of the atlas defines a Bayesian weighting (i.e., probability density function), which provides very rich information to find the organ’s boundaries.

Here, we present the results of a full probabilistic atlas construction for 32 patients, where 31 of the patients have been mapped onto one individual using thin plate splines (TPSs) as the warping transform and mutual information (MI) as the similarity measure. Warping 31 patients onto one arbitrarily chosen reference patient can introduce bias toward that specific reference patient, especially if the reference patient is far from the average population the atlas was intended for. Thus, the reference patient was selected to best represent the population to our knowledge aided by an expert inspection. By iterating this construction process, i.e., using the resulting atlas as the reference target of another atlas construction phase, other groups have shown that the variance of the atlas and the dependence of the reference patient can be reduced further [14], [21]; we have not performed such iterations here and leave it to future work. While in the past many groups focused only on the mean surface, the Bayesian value of the probabilistic atlas in the segmentation task lies not only in the mean but also in the variance of the atlas. In general, the variance of the atlas will decrease as we increase the degrees of freedom (DOFs) of the warping transform.

Data sets used to construct our atlas come from anonymous, clinically ill patients. Other warping transforms like B-splines and radial basis function are commonly used in registration literatures [31], [32] for their local characteristics and computation efficiencies but TPS is highly endorsed by rich literatures in shape statistics and morphometrics [33]–[35]. Recently, it has been found that TPS is the maximum-likelihood estimator (MLE) for unspecified landmarks not identified as control

Manuscript received June 26, 2002; revised November 22, 2002. This work was supported in part by U.S. National Institute of Health under Grant 1P01 CA87634-01A2, Grant 1P01 CA85878-01A2, and Grant 1P30 CA46592. Recommended by Associate Editor N. Ayache. *Asterisk indicates corresponding author.*

H. Park is with the Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI 48109 USA.

P. H. Bland is with Department of Radiology, University of Michigan, Ann Arbor, MI 48109 USA.

*C. R. Meyer is with Department of Radiology, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: cmeyer@umich.edu).

Digital Object Identifier 10.1109/TMI.2003.809139

points [36]. Notions of landmarks, control points, and TPS can be found in [34], [35]. Similarity measures reflect how well two data sets of interests are registered [55]. There are many similarity measures other than MI, but MI has gained considerable support over the past years [56].

Segmentation can be thought as the process of assigning labels to individual voxels in the data set. Extensive review of the segmentation methods and deformable surface models can be found in [37], [57]. Most segmentation methods can be divided into two groups. The first one is supervised segmentation where certain volume of interest (VOI) has to be manually specified for training different tissue types. The other is unsupervised segmentation where the discovery of different tissue types is automated. Supervised segmentation typically uses k -nearest neighbors (k -NN) algorithm followed by some regularization [38], [39]. Unsupervised segmentation typically uses expectation maximization (EM) algorithm to estimate parameters for different tissue types and feeds those parameters to Bayes rule followed by some regularization [40]–[42]. Both segmentation algorithms include regularization in their steps. A common way to regularize or smooth is to incorporate a Markov random field (MRF) priors. Basically, the MRF model encourages adjacent voxels having similar labels. Extensive reviews and implementation on MRF regularization are available [43]–[47]. In the execution of Bayes rule of the unsupervised segmentation, some algorithms use maximum *a posteriori* (MAP) approach while others use “maximization of the posterior marginals” (MPM) approach [48]. The MPM approach was shown to be better suited for segmentation tasks [48]. Our work takes the route of the unsupervised segmentation with MAP. Our contribution is to extend the unsupervised segmentation with the additional information provided by the probabilistic atlas. Improvements in segmentation quality compared with cases with no atlas information are readily noticeable especially where there is little contrast between organs of interest. Specifically we were able to eliminate most of the soft tissue false positives around liver/kidneys/spinal cord and distinguish liver, spinal cord, left kidney, and right kidney. Note that in order to apply the atlas information, a CT data set should be mapped onto the atlas space by the same method that was originally used to construct the atlas.

We have applied our segmentation algorithm retrospectively to 20 abdominal CT scans of anonymous patients who have established lymphoma and received these scans as part of a standard clinical protocol to follow possible lymph node enlargement. These are all 7–10-mm-thick slices acquired during a breath-hold on a helical scanner. Only oral contrast material was administered; no intravenous (IV) contrast was used. Segmentation of such non-IV-contrast CT scans is very difficult because CT values for nonlipid soft tissues are primarily related to density, and all nonlipid soft tissues have the density of water. False positive rates and false negative rates are reported as a measure of segmentation performance. Also, comparisons with cases where no atlas information is used are made.

This paper is organized in three parts. The first part will include methods and results for constructing a probabilistic atlas. The second part will have methods and results for segmentation using the atlas information from the first part. Finally, we

will conclude with summary and discussion of atlas construction and segmentation. Primary contributions of this paper are: 1) to build a probabilistic atlas of abdominal organs and 2) to enhance unsupervised segmentation with added information from the probabilistic atlas.

II. METHODS

A. Mutual Information for Automatic Multimodality Image Fusion (MIAMI Fuse)

The registration of the individual data sets onto the target reference was implemented using MIAMI Fuse, which implements TPS as the warping transform and MI as the similarity measure [54]. Since the abdomen consists of organs that are compressible and have different sizes and shapes across individuals, the use of a warping transform is required instead of a rigid transform. TPS is expressed in terms of control points, which are placed in pairs in the corresponding loci of the related data sets. TPS deformation arises from surface interpolation literature where control points are constraints of the interpolating function. From the interpolating perspective, each constraint (control point) primarily affects interpolating function in its vicinity. Detailed visualization of local effects of TPS in two dimensions can be found in Appendix A. Thus, control points primarily affect deformations near the control points. Also, note that the density of control points is related to the DOF of the warping transform. Thus, when the region has large density of control points the DOF of the warping transform in that region is high. TPS is typically reserved for warping transforms but TPS reduces to affine or rigid transform if the number of control points is less than five in three dimension. Details of TPS and control points are found in these texts [34], [35]. MI is used as the similarity measure of choice. It basically exploits the fact that two co-registered data sets yield a joint probability density function (pdf) with tight clusters, whereas un-registered data sets yield a joint pdf with disperse clusters. Tighter clusters (i.e., more correlation) translate into higher MI values than more disperse ones (i.e., less correlation). In our implementation, all pdf's (both marginal and joint) are estimated by histograms with fixed bin width.

B. Distribution of Control Points

The liver, kidneys and spinal cord of the target reference data set were manually segmented by an expert using a window between -123 Hounsfield units (HUs) and 131 HUs. Boundaries were traced using hand controlled optical mouse and 36 control points were placed in the target reference data set. The distribution of control points in the reference data set were chosen as follows: 17 control points were chosen in the reference data set's liver, seven control points were chosen in and around the spinal cord, and six were chosen in each kidney as seen Fig. 1. The only criterion for the placement of these points is to achieve an approximately uniform density of control points for their distribution in each organ. The voxel dimensions of the target reference CT data set was $(1.29 \text{ mm})^2 \times 5 \text{ mm}$. Again, note that control points in each organ control deformations/transforms in their locality.

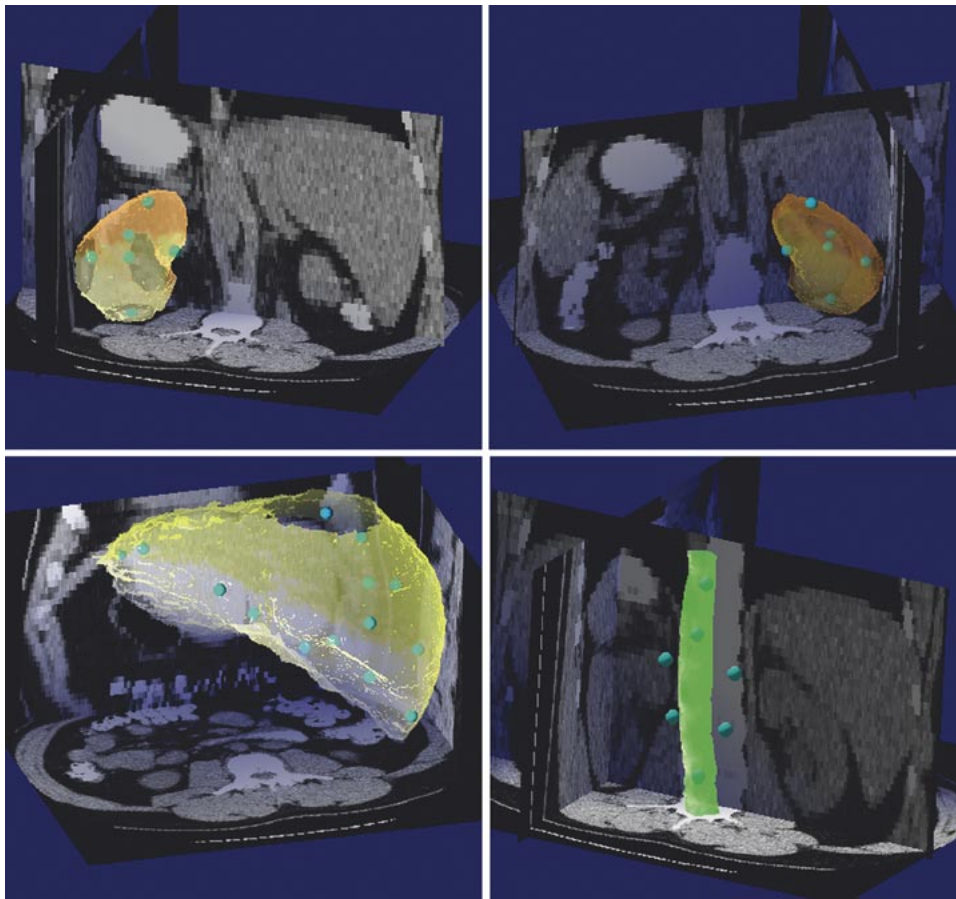


Fig. 1. Distribution of control points for organs of interests. From top left, left kidney model and associated six control points; (top right) right kidney model and associated six control points; (bottom left) liver model and associated 17 control points; and (bottom right) spinal cord model and associated seven control points in the reference volume.

C. Construction of the Atlas

We observed that if we registered the whole abdomen between reference and patient volume, liver primarily drives the registration since it has the largest voxel count of any organ and, therefore, the liver effects the joint histogram and the resulting MI more than other organs. Thus, smaller organs like kidneys were not accurately aligned if jointly registered with other organs. By registering each organ separately, better overall registration accuracy is obtained. Each organ was registered onto its homologous pair in the reference volume separately, and then one final optimization using all 36 control points was used to correct for any control point interactions that may have occurred between adjacent control points from the separate organ registrations.

Separate registrations for each organ were performed by masking out, i.e., zeroing, the remainder of the grayscale reference data so that it only contains the organ of interest. Since the usual implementation of our registration algorithm ignores zeros if they occur in either data set, the only data used to construct the joint histogram derives from the chosen reference organ and the mapping of the patient's data onto that organ only.

The registration of a single patient onto the target reference geometry begins by the operator's selection of just four homologous control points in the patient's data set corresponding

the loci of the first four control points in the target reference, i.e., two in the liver, and one at the center of each kidney. This process essentially identifies to the algorithm the pose of the patient in the scanner, i.e., head/feet first and lying prone/supine. The loci of these four points need not be exact since the optimizing process that follows will find the correct corresponding loci under the MI similarity measure. Then, the algorithm repeats the process of registering each organ separately. The optimizer moves the first four control points in the patient's data space to obtain the best isotropically scaled, rotate-translate mapping of the patient's organ onto the target reference organ. Then, based on the optimized geometry mapping just obtained, the remaining points (out of the first four plus the N for that organ) are mapped onto the patient's data set to initialize the $N + 4$ control point TPS transform of the patient's organ onto the target reference organ and the $N + 4$ point TPS optimization follows. After the fit of one organ has been optimized, the process of focusing the attention of the registration on each remaining organ separately continues until all of the specified organs have been fit. Basically, for each organ, first affine geometry is optimized and then based on the obtained affine geometry $N + 4$ point TPS is initialized, which removes the burden of specifying $N + 4$ control points in the patient's volume. The resulting loci of all the organ-specific control points are then simultaneously optimized using a 36 control-point TPS to correct for the small interplay that may

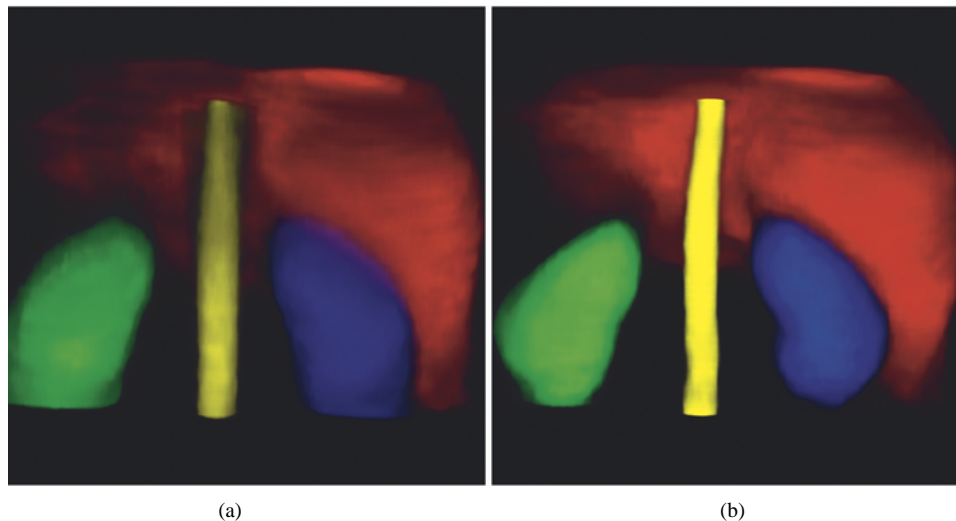


Fig. 2. Comparison of probabilistic atlases. (a) Probabilistic atlas constructed by a 36 control point TPS transform. (b) Probabilistic atlas constructed by four control point affine transform. Red corresponds to liver and green, blue, and yellow correspond to right kidney, left kidney, and spinal cord, respectively.

occur between relatively close control points in the global solution, which was initially ignored in the individual focused organ refinement process.

Except for the placement of the four initial control points in each patient, the process of mapping the patient onto the target reference geometry was completed by the algorithm totally automatically. Indeed, we may be able to achieve the first full affine registration fully automatically by placing the first four control points based only on the extents of the dataset and then implementing a limited DOF registration somewhere between full affine (12 DOFs) and that of the rigid registration (six DOFs), e.g., isotropically scaled (seven DOFs) or even anisotropically scaled (nine DOFs). We have not as yet tried such fully automatic approaches but they clearly are important topics for further research.

III. RESULTS—ATLAS

After registering each patient onto the reference, the probabilistic atlas was computed by applying the same registration transform to each patient's manually segmented data set. Then, for each voxel in the reference volume, a vector float value was computed where the value of each vector component for each voxel represented the fractional percentage of registered patient data sets that have a label at that reference voxel location corresponding the particular organ represented by that vector component. Volume rendering of the resulting probabilistic atlas is shown in Fig. 2(a). For the sake of comparison, the full affine atlas construction is shown in Fig. 2(b). Volume rendering for both is via the ray tracing algorithm where the final color and brightness of the rendered pixel depend on the integrated local attenuation of the ray as it passes through the data volume. There are no gradient lighting effects used here. In Fig. 2(a), the appearance of darker "shadows" around each of the individual organ models is due to the fact that at the edge of each model its probability, represented here by a particular intensity and hue corresponding to the organ, drops sharply to zero. The sharper the gradient, i.e. the better the registration process in constructing the probabilistic atlas, the more limited and thus

visible are the dark fringes. The difference between two construction methods is most evident between the right kidney and the liver. Note in Fig. 2(a) the sharp dark demarcation between the right kidney and liver where the probability that voxels in that region are either kidney or liver are low. In contrast, note in Fig. 2(b) how the kidney and liver blend into one another in the fuzzy, affine atlas construction.

In general, the variance of the atlas will decrease as we increase the DOF of the transform. Indeed, with sufficiently high DOF warping it might be possible to make the atlas variance approach zero. Constraints, i.e., guarantees of invertibility, would have to be added to assure that singularities do not occur in the warping as might happen in the case of creating/annihilating structures in the patient that do not exist in the atlas. Under such invertibility constraints, we are still assured that the atlas (with nearly no variance) can be reliably mapped onto the patient very accurately, and the task of segmentation is essentially completed by the warping with little need for further retrospective segmentation following registration by carrying the atlas information onto the target patient. At 36 control points (3×36 DOF) such concerns are not even within the realm of possibility for such moderate DOF. The variability of our atlas is largely affected by the presence/absence of features across patients, e.g., some patients do not have a left lobe of the liver, and shape differences that exceed the ability of our warping transform to compensate.

To quantify results, we computed average standard deviation over different slices of the resulting atlas. Occurrence of the organ edges were assumed at 50% threshold of the vector value and Gaussian curve with different variances were convolved with a step function occurring at the organ edges to best match in the least mean squared error (LMSE) sense the actual probabilistic atlas. The results are summarized in Table I.

In general, the variances of TPS based atlas construction are on the order of four to eight times better than the affine atlas construction. The structure that is a notable exception to this generalization is the left lobe of the liver. Here, the similarity in variance between the two methods may reflect more on the anatomical variation, i.e., presence or absence of this structure across patients.

TABLE I
VARIANCES OF THE PROBABILISTIC MAP IN SOME SLICES

Organ of Volumetric Probabilistic Atlas	TPS Construction, Average σ , mm	Affine Construction, Average σ , mm
Midhepatic slice, entire	6.8	11.2
Fossa of R. kidney	4.7	13.7
L. lobe	12.1	14.9
R. Kidney, midslice	6.1	13.0

IV. METHODS—SEGMENTATION

In this paper, the true label field (segmented field) will be denoted \mathbf{X} , the given observed data set will be denoted \mathbf{Y} and the probabilistic atlas is denoted by \mathbf{A} . Elements of \mathbf{X} , \mathbf{Y} and \mathbf{A} are arranged by spatial position denoted by $i \in I$, where I is the simple index in rectangular grid. Throughout the paper, $\mathbf{x} = (x_1, x_2, \dots, x_N)$, $\mathbf{y} = (y_1, y_2, \dots, y_N)$, and $\mathbf{a} = (a_1, a_2, \dots, a_N)$, where N is the total number of voxels, will represent sample realizations of \mathbf{X} , \mathbf{Y} and \mathbf{A} respectively. Sample space of \mathbf{X} will be denoted $\Omega_{\mathbf{X}}$ where $\Omega_{\mathbf{X}} = \{\mathbf{x} : x_i \in \{1, 2, 3, 4, 5\}, \forall i \in I\}$. Label 1 is the liver, 2 is the right kidney, 3 is the left kidney, 4 is the spinal cord and 5 is the “None of the above” label. Note that the probability atlas is 5-vector $a_i = (a_{i1}, a_{i2}, a_{i3}, a_{i4}, a_{i5})$, where each component corresponds to probability of a specific organ being there. The atlas information is realized as

$$P(x_i = k) = a_{ik}, \quad k = 1, \dots, 5$$

$$a_{i5} = 1 - \sum_{k=1}^4 a_{ik}, \quad i \in I.$$

As noted before data set \mathbf{Y} is already mapped onto the spatial frame of the probabilistic atlas using the same techniques that were used to construct the atlas.

A. Bayesian Framework

The problem is to estimate the label \mathbf{X} that best explains the given observation \mathbf{Y} according to some cost function. Different cost functions can be applied, such as MAP which aims to maximize the global *a posteriori* probability $\mathbf{P}(\mathbf{X}|\mathbf{Y})$ and MPM which aims to maximize the posterior marginal distribution $\mathbf{P}(\mathbf{X}_i|\mathbf{Y})$ for each voxel i . We chose MAP and the following is the formula for the realization of estimate of \mathbf{X} :

$$\hat{x} = \arg \max_x P(X = x | Y = y).$$

B. Tissue Models

We will assume that random variable Y_1, Y_2, \dots, Y_N are conditionally independent given the true label \mathbf{X} . Also, we will assume that probability density function of Y_i given \mathbf{X} depends only on X_i . Furthermore, the probabilities of Y_i 's are modeled as conditional Gaussians given mean and variance of the true label X_i . There are other advanced models for different tissue types especially for brain tissues [51] but still many papers use

this simple Gaussian model and have been successful [41], [42]. Formulation for the conditional probability follows:

$$p(\mathbf{Y}|\mathbf{X}) = \prod_{i=1}^N p(Y_i|X_i = k)$$

$$p(y_i|X_i = k) = \left\{ \frac{1}{\sqrt{2\pi\sigma_{x_i}^2}} \exp\left(-\frac{(y_i - \mu_k)^2}{2\sigma_{x_i}^2}\right) \right.$$

for $k = 1, 2, 3, 4, 5$;
 histogram for $k = 5$, i.e., “None of the above”}.

However, for the tissue type “None of above,” there may be multiple clusters in the gray scale distribution since it includes background, all intestine, spleen, stomach, body wall, and bones. Thus, using a unimodal Gaussian will not be sufficient. Note that our typical abdomen CT scans of slice thickness 5 mm have $512 \times 512 \times 40 \approx 10^8$ voxels. If we can use only 0.1% of all voxels (i.e., 10^5 counts) to train the tissue type “None of above” using a histogram, we have enough samples to make the 256 bin histogram closely approximate the true distribution. Using a very conservative estimate, which assumes “None of above” tissue type occupies 5% of the abdomen volume (i.e., 5×10^6 counts), we are guaranteed to have a reasonable estimate of the “None of above” tissue type.

C. Estimation of Gaussian Parameters

Mean μ_k and variance σ_k^2 of the Gaussian modeling have to be estimated for the first four tissue types. A simple sample mean and variance over VOIs are used. The selection of VOI is automated by the aid of the probabilistic atlas. With the atlas information, we have a very strong prior knowledge where the organs are. Thus, the VOI for the training of a specific organ can be identified where the atlas predicts nearly 100% organ presence. We used above 96% of being the desired organ for specifying VOIs. Lowering the threshold for the VOIs may induce better robustness of hyper parameters (i.e. mean and variance) but it may also bias the estimates, e.g., lowered threshold VOIs may included multiple organs. Actual segmentation results with different threshold for VOIs will be discussed later. For the tissue type “None of above,” a histogram was calculated over the automatically selected VOI.

The standard method used in unsupervised segmentation is the combination of EM/MAP. In those cases iteration between EM and MAP to ensure convergence of the hyper parameters (i.e., mean and variance) is required [41], [42]. However, our method is estimating hyper parameters over the areas where the likelihood of a specific organ presence is very high, so the estimates of hyper parameters are very stable. In result our method need not iterate to achieve stable hyper parameter estimates and segmentation. We have confirmed with our data sets that our segmentation results did not improve with iteration.

D. MRF Regularization

We included the usual penalty for dissimilar adjacent labels in the form of MRF regularization where the underlying label is modeled as Gibbs distribution. The MRF penalty favors adjacent label fields to have similar labels, which is the same general spatial correlation that exists in medical data sets. A vast amount

of literature exists on MRF models [43], [44]. Here, we use a very simple implementation of the MRF model. A six-voxel neighborhood system that is called clique is used. Our Clique consists of four nearest neighbor voxels in the slice plane direction and two nearest neighbor voxels in the out of plane direction. In case of corner voxels and outer edge voxels, Cliques are accordingly reduced to a smaller neighborhood. There can be other complicated Cliques that span many more voxels than six, which is tailored for other applications. According to Hammerley–Clifford theorem, the probability density function of the MRF model takes the form of Gibbs distribution [49]

$$p(x) = Z^{-1} \exp(-U(x_i))$$

$$= Z^{-1} \exp\left(-\beta \sum_{(i,j) \in C} t(x_i, x_j)\right)$$

$$t(x_i, x_j) = \begin{cases} 0; & x_i = x_j \\ 1; & x_i \neq x_j \end{cases}.$$

Z is the normalizing factor, β is the positive constant which controls regularization, C is the Clique and $U(x)$ is the energy function that increases with inhomogeneous adjacent labels. We assume that the value β is known *a priori*, which is assumed in many MRF-based segmentations [52], [53]. Also, it is found that the optimal value of β is not highly dependent of the image content and the performance of the segmentation algorithm remains relatively unchanged with respect to β [41].

E. Formula and Optimization

Optimizing the posterior probability is not an easy task, especially because there are so many possible realizations of the MRF model and the optimization is prone to be caught in local maximas. There are many methods for optimizing the posterior probability in the MRF setup. Typical methods are Simulated Annealing [45] and genetic algorithms [46] that are theoretically proven to converge to the global maximum, but all these methods are extremely computationally expensive. In this chapter, we have used a method similar to Iterated Conditional Mode (ICM) by Besag [47], which is a greedy approach to those slower methods. Additional information available from the atlas is reflected on the Markovian assumption of the label field $p(X_i = k | x_{\delta i}) = P(X_i = k) \frac{e^{\beta s_i k}}{Z}$

$x_{\delta i}$ clique of voxel i ;

$s_i(k)$ number of voxels with the same label as voxel i in the clique;

Z normalizing factor.

Thus, we are favoring similar adjacent labels in addition to the most probable label from the atlas. In ICM, maximizing the posterior probability is maximizing the product of two probabilities; $p(y_i | x_i)$ and $p(x_i | x_{\delta i})$ [47]. The $p(x_i | x_{\delta i})$ term is further divided into atlas term and the Markov regularization term. Basically, we have included one additional term into the posterior probability. The first term $p(y_i | x_i)$ is the difference in the observation from the model mean in Gaussian sense, the second term is the contribution from the probabilistic atlas and the third term

is the MRF regularization term. Typical methods have only two terms, the first and the third term

$$\arg \max p(y_i | X_i = k) p(X_i = k | x_{\delta i})$$

$$= \arg \max p(y_i | X_i = k) P(X_i = k) e^{\beta s_i k}, \quad i \in I.$$

The ICM method needs an initial segmentation to begin with. The initial segmentation is provided by maximizing the posterior probability without the MRF term. In detail, it is optimizing the only two terms excluding the MRF term in the posterior probability formula provided above. In the implementation of ICM, regular ICM requires that all current labels (i.e. labels upto index $i - 1$ and labels from the previous iteration) are used to update the new label. Our implementation only uses labels from the previous iteration. Our modified ICM iteration stops when label field changes between iterations are less than a given threshold.

V. RESULTS—SEGMENTATION

Our segmentation method was tested on 20 non-IV-contrast abdominal CT scans. Patient CT voxels that map onto probabilistic atlas organ regions of 96% or greater were used for training, the threshold for the ICM iteration was set to 5% label change and β , the smoothness parameter, was set to 1.5. Performance of the segmentation is measured using false positive rate and false negative rates. We assume that the ground truth segmentation comes from an expert labeling, the null hypothesis is the absence of a specific organ and the alternative hypothesis is the presence of a specific organ. Here is one example of segmentation out of 20 cases.

Note in Fig. 3(g) that there is very little contrast among different Gaussian modelings of organs. The four Gaussian curves for different tissue types have very large overlapping regions. Thus, without the atlas information the segmentation algorithm cannot distinguish well among organs of interest. Without the atlas information as in Fig. 3(c), right kidney, left kidney, and spinal cord are all assigned the same label and there are lots of other soft tissues that are still classified as organs of interest. With the added atlas information as in Fig. 3(d), we now gain contrast over kidneys and spinal cord and soft tissues that are not of interest (e.g., intestines, spleen, etc) are eliminated. Fig. 3(f) shows the difference between Fig. 3(c) and Fig. 3(d). In Fig. 3(f), the background gray is the zero difference and all other values reflect differences between the two segmentation methods. From Fig. 3(f), mis-classification of spinal cord and left kidney is evident and differences in the two segmentation methods are found at body walls near the liver, organs of noninterest and centers of kidneys.

The false negative rates for left kidney and spinal cord without the atlas are over 90%; it comes as no surprise that the segmentation algorithm without the atlas cannot distinguish left kidney and spinal cord from other organs. From the first two rows of Table II, false negative rate values were generally larger than false positive values, which is the result of a conservative segmentation. By overlaying Fig. 3(a) and Fig. 3(d) or by inspecting Fig. 3(e), you can observe that the segmentation results are not

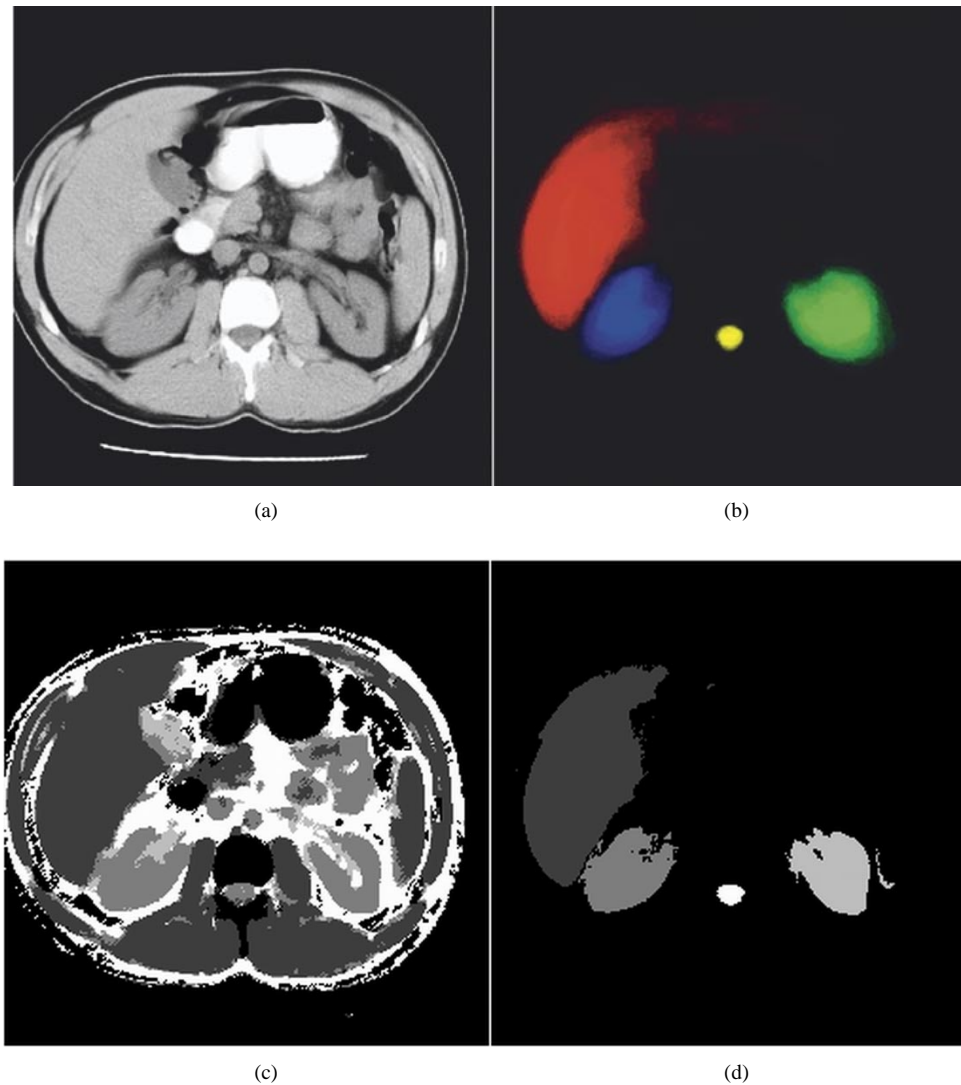


Fig. 3. (a) One slice of the data set. (b) Atlas information. Segmentation results (c) without and (d) with atlas information.

really extended toward kidney and liver boundaries. In other words, our results are conservatively biased, i.e., we are more likely to call a voxel not an organ when there is an organ present especially around the edges of the organs. Overall performance measures with the atlas information were very close to the performance measures of the liver, which is not surprising since liver is the biggest organ in our four-organ abdomen model and thus has the most voxel counts. The last two rows of Table II show the effect of changing the training threshold from 96% to 90%. The resultant false positive rates and false negative rates for the 90% case compared to 96% are slightly worse, but possibly within the realm of no statistical significance. If the threshold was further lowered for instance 60%, it is likely that we would observe biased estimates for the Gaussian parameters leading to worse segmentation results. However, for a moderate change in threshold (i.e., from 96% to 90%), changes in segmentation results were minimal.

There remain fragments of organs around the liver and the kidneys in the segmented results [Fig. 3(d)], these fragments can be further eliminated if we had an atlas with lower variance. Atlases with lower variances will fall off very rapidly near the

organ boundaries, thus the algorithm will be unlikely to classify tissue fragments that are not only similar to the organ itself in grayscale values but also spatially close, as organs of interest. Still, the atlas constructed from 32 subjects was sufficient to remove most of the fragments. A subsection of the receiver operating characteristic (ROC) curve is presented in Fig. 3(h). In this case, we were not able to adjust the false positive rate at our will since the change of the only maneuverable parameter yielded very limited range of false positive rates. Thus, in this implementation we are operating at the low false positive rate and high true positive rate region. Performance measures for all 20 data sets are given in Table III. For Table III, ICM threshold is set to 0.5% and other parameters are the same as before.

VI. CONCLUSION

We have successfully constructed a full vector valued probabilistic atlas by registering 31 patient data sets onto a single patient. The value of probabilistic atlases increase inversely with their variance. In general the variances of TPS based atlas construction are on the order of four to eight times better than the

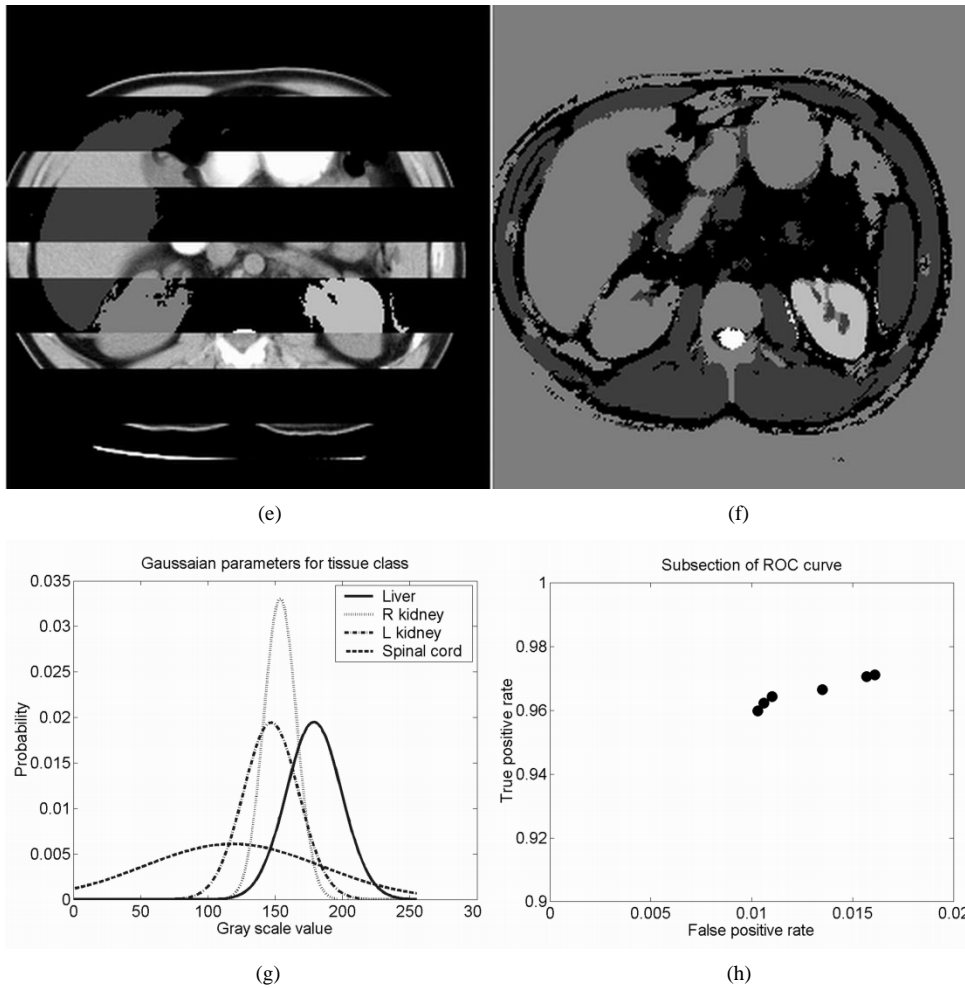


Fig. 3. (Continued.) (e) Comparison between grayscale data set and segmentation with atlas information. (f) Difference between segmentation with and without atlas information. (g) Gaussian modeling for different tissue types. (h) Subsection of the ROC curve.

TABLE II
PERFORMANCE MEASURES FOR THE SAMPLE CASE

	Liver	Right Kidney	Left Kidney	Spinal Cord	Overall
False positive rate (with atlas)	0.0064	0.0004	0.0004	0.0002	0.0054
False negative rate (with atlas)	0.0639	0.1447	0.1081	0.2116	0.0759
False positive rate (w/o atlas)	0.1452	0.0476	0.0130	0.1200	0.1271
False negative rate (w/o atlas)	0.0579	0.1278	0.9195	0.9843	0.1405
False positive rate (with atlas and lowered threshold for training)	0.0072	0.0004	0.0004	0.0002	0.0060
False negative rate (with atlas and lowered threshold for training)	0.0641	0.1365	0.1166	0.1982	0.0758

affine atlas construction. The structure that is a notable exception to this generalization is the left lobe of the liver. Here, the similarity in variance between the two methods may reflect more on the anatomical variation, i.e., presence or absence of this structure across patients. In order to reduce the effects of such variation at least two approaches are possible. One is to use a few subcategories of atlases with or without the missing structure, and the other is to use many DOF in the registration method [28]–[30].

While we do not claim that the 36 control points chosen here are optimal in their location, it is clear from the excellent regis-

tration results across a widely varying population that we are able to choose a sufficient set of control points in the target reference to support the computation of a useful probabilistic atlas. Recall that these results are computed from noncontrast CT scans of 32 patients, which are more difficult to register than contrast enhanced CT scans. Future work will address issues such as iterative atlas construction and optimal control point placement.

Segmentation is one of the obvious applications where atlas information can help. In the case of noncontrast CT scans, atlas information aided the segmentation algorithm to support differentiation between similar gray valued organs where previously no other differentiations were possible. Also, atlas information led us to remove organs of lesser interest. All these will be helpful in clinical setups like radiation therapy treatment planning. Our algorithm showed underestimation of organs largely due to inability to assign correct labels near organ boundaries. This effect can be remedied by having an atlas with lower variance or a better modeling of “None of above” label other than a simple histogram. Other methods of finding specific organ edges in the patient’s data set using the model information may further improve the results. In summary we have demonstrated successful integration of the atlas information into the standard unsupervised segmentation and the results for the noncontrast CT cases are excellent. Further research will include assimi-

TABLE III
PERFORMANCE MEASURES FOR 20 DATA SETS

	Liver	Right Kidney	Left Kidney	Spinal Cord	Overall
False positive rate	0.0077	0.0005	0.0007	0.0002	0.0064
False negative rate	0.0782	0.1053	0.1066	0.1355	0.0839

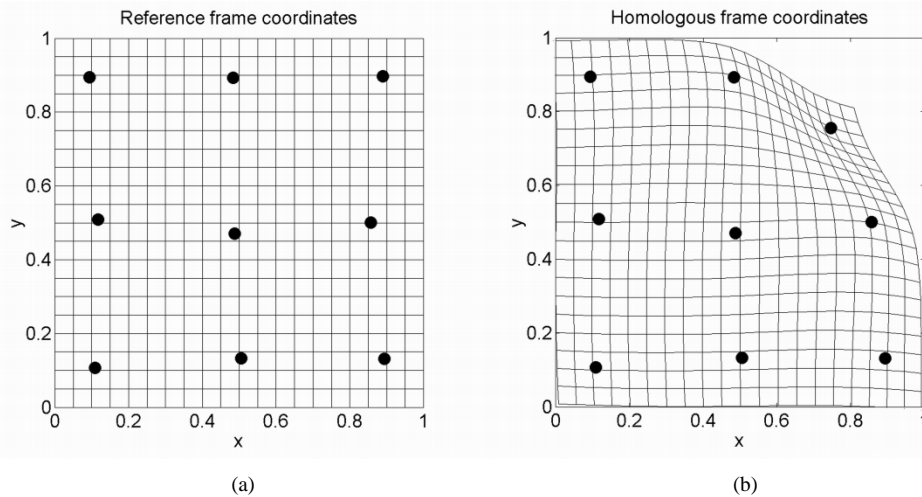


Fig. 4. Local effects of control points. (a) Reference frame coordinates. (b) Homologous frame coordinates.

tion of the atlas information into other well established segmentation algorithms including level set theory [50].

APPENDIX

LOCAL EFFECTS OF CONTROL POINTS IN TWO DIMENSION

Nine control point pairs are placed uniformly across the $[0, 1] \times [0, 1]$ square as seen in Fig. 4. Control point of the northeast side in the homologous side was moved toward southwest direction. Effects of control points of moving the control point in the northeast region are clearly localized to the northeast region. Local effects of control points are harder to visualize in three dimension still effects of control points are stronger near control points.

REFERENCES

- [1] A. Kelemen, G. Szekely, and G. Gerig, "Elastic model-based segmentation of 3-D neuroradiological data sets," *IEEE Trans. Med. Imag.*, vol. 18, pp. 828–839, Oct. 1999.
- [2] D. MacDonald, D. Avis, and A. C. Evans, "Proximity constraints in deformable models for cortical surface identification," in *Proc. Medical Image Computing Computer-Assisted Intervention*, vol. 1496, 1998, pp. 650–659.
- [3] D. MacDonald, N. Kabani, D. Avis, and A. C. Evans, "Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI," *Neuroimage*, vol. 12, pp. 340–356, 2000.
- [4] T. McNerney and D. Terzopoulos, "Topology adaptive deformable surfaces for medical image volume segmentation," *IEEE Trans. Med. Imag.*, vol. 18, pp. 840–850, Oct. 1999.
- [5] J. Montagnat and H. Delingette, "Volumetric medical images segmentation using shape constrained deformable models," in *Proc. CVRMed-MRCAS'97*, vol. 1205, 1997, pp. 13–22.
- [6] W. J. Niessen, B. M. T. Romeny, and M. A. Viergever, "Geodesic deformable models for medical image analysis," *IEEE Trans. Med. Imag.*, vol. 17, pp. 634–641, Aug. 1998.
- [7] D. G. Shen and C. Davatzikos, "An adaptive-focus deformable model using statistical and geometric information," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, pp. 906–913, Aug. 2000.
- [8] A. Garrido and N. P. de la Blanca, "Applying deformable templates for cell image segmentation," *Pattern Recogn.*, vol. 33, pp. 821–832, 2000.
- [9] H. Y. Wang and B. Ghosh, "Geometric active deformable models in shape modeling," *IEEE Trans. Image Processing*, vol. 9, pp. 302–308, Feb. 2000.
- [10] X. L. Zeng, L. H. Staib, R. T. Schultz, and J. S. Duncan, "Segmentation and measurement of the cortex from 3-D MR images using coupled-surfaces propagation," *IEEE Trans. Medical Imag.*, vol. 18, pp. 927–937, Oct. 1999.
- [11] X. C. Zhao and F. H. Qi, "Automatic image segmentation based on deformable models and pixel affinity statistic," *J. Infrared Millimeter Waves*, vol. 19, pp. 33–37, 2000.
- [12] L. M. Gao, D. G. Heath, and E. K. Fishman, "Abdominal image segmentation using three-dimensional deformable models," *Invest. Radiol.*, vol. 33, pp. 348–355, 1998.
- [13] I. D. Dinov, M. S. Mega, P. M. Thompson, and L. Lee *et al.*, "Analyzing functional brain images in a probabilistic atlas: a validation of subvolume thresholding," *J. Comput. Assist. Tomogr.*, vol. 24, pp. 128–138, 2000.
- [14] A. Guimond, J. Meunier, and J. P. Thirion, "Average brain models: a convergence study," *Comput. Vision Image Understand.*, vol. 77, pp. 192–210, 2000.
- [15] G. Le Goualher, E. Procyk, D. L. Collins, and R. Venugopal *et al.*, "Automated extraction and variability analysis of sulcal neuroanatomy," *IEEE Trans. Med. Imag.*, vol. 18, pp. 206–217, Feb. 1999.
- [16] G. Le Goualher, D. L. Collins, C. Barillot, and A. C. Evans, "Automatic identification of cortical sulci using a 3D probabilistic atlas," in *Proc. Medical Image Computing Computer-Assisted Intervention*, vol. 1496, 1998, pp. 509–518.
- [17] P. Thompson and A. W. Toga, "Visualization and mapping of anatomic abnormalities using a probabilistic brain atlas based on random fluid transformations," *Visual. Biomed. Comput.*, vol. 1131, pp. 383–392, 1996.
- [18] P. M. Thompson, D. MacDonald, M. S. Mega, and C. J. Holmes *et al.*, "Detection and mapping of abnormal brain structure with a probabilistic atlas of cortical surfaces," *J. Computer Assist. Tomogr.*, vol. 21, pp. 567–581, 1997.

- [19] P. M. Thompson, J. Moussai, S. Zohoori, and A. Goldkorn *et al.*, "Cortical variability and asymmetry in normal aging and Alzheimer's disease," *Cerebral Cortex*, vol. 8, pp. 492–509, 1998.
- [20] P. M. Thompson, M. S. Mega, R. P. Woods, and C. I. Zoumalan *et al.*, "Cortical change in Alzheimer's disease detected with a disease-specific population-based brain atlas," *Cerebral Cortex*, vol. 11, pp. 1–16, 2001.
- [21] A. W. Toga and P. M. Thompson, "The role of image registration in brain mapping," *Image Vision Comput.*, vol. 19, pp. 3–24, 2001.
- [22] K. L. Narr, P. M. Thompson, T. Sharma, and J. Moussai *et al.*, "Three-dimensional mapping of gyral shape and cortical surface asymmetries in schizophrenia: gender effects," *Amer. J. Psychiatry*, vol. 158, pp. 244–255, 2001.
- [23] K. L. Narr, P. M. Thompson, T. Sharma, and J. Moussai *et al.*, "Mapping morphology of the corpus callosum in schizophrenia," *Cerebral Cortex*, vol. 10, pp. 40–49, 2000.
- [24] J. L. Boes, P. H. Bland, T. E. Weymouth, and L. E. Quint *et al.*, "Generating a normalized geometric liver model using warping," *Invest. Radiol.*, vol. 29, pp. 281–286, 1994.
- [25] J. L. Boes, C. R. Meyer, and T. E. Weymouth, "Liver definition in CT using a population-based shape model," in *Computer Vision, Virtual Reality and Robotics in Medicine*, N. Ayache and F. R. Niçe, Eds. Berlin, Germany: Springer-Verlag, 1995, vol. 905, Lecture Notes in Computer Science, pp. 506–512.
- [26] J. L. Boes, T. E. Weymouth, and C. R. Meyer, "Multiple organ definition in CT using a Bayesian approach for 3D model fitting," in *SPIE Vision Geometry IV*. San Diego, CA: SPIE, 1995, vol. 2753, pp. 244–251.
- [27] A. C. Evans. [Online] <http://www.bic.mni.mcgill.ca/demos/animal/>
- [28] M. I. Miller, G. E. Christensen, Y. Amit, and U. Grenander, "Mathematical Textbook of Deformable Neuroanatomies," in *Proc. Nat. Acad. Sci.*, 90, 1993, pp. 11944–11948.
- [29] G. E. Christensen, R. D. Rabbitt, and M. I. Miller, "Deformable templates using large deformation kinetics," *IEEE Image Processing*, vol. 5, pp. 1435–1447, Oct. 1996.
- [30] G. E. Christensen, S. C. Joshi, and M. I. Miller, "Volumetric transformation of brain anatomy," *IEEE Trans. Medical Imag.*, vol. 16, pp. 864–877, Dec. 1997.
- [31] N. Arad and D. Reissfeld, "Image warping using few anchor points," *Comput. Graph. Forum*, vol. 14, pp. 35–46, 1995.
- [32] S. Lee, G. Wolberg, K.-Y. Chwa, and S. Y. Shin, "Image metamorphosis with scattered feature constraints," *IEEE Trans. Visual. Comput. Graph.*, vol. 2, pp. 337–354, Dec. 1996.
- [33] F. L. Bookstein, "Shape and the information in medical images: a decade of the morphometric synthesis," *Comput. Vision Image Understand.*, vol. 66, no. 2, pp. 97–118, 1997.
- [34] F. L. Bookstein, *Morphometric Tools for Landmark Data*. Cambridge, U.K.: Cambridge Univ. Press, 1991.
- [35] I. L. Dryden and K. V. Mardia, *Statistical Shape Analysis*. New York: Wiley, 1998.
- [36] J. T. Kent and K. V. Mardia, "Link between kriging and thin plate splines," *Probab., Stat., Optim.*, pp. 325–339, 1994.
- [37] J. C. Bezdek, L. O. Hall, and L. P. Clark, "Review of MR segmentation images using pattern recognition," *Med. Phys.*, vol. 20, pp. 1033–1048, 1993.
- [38] S. Warfield, "Fast k-NN classification for multichannel image data," *Pattern Recogn. Lett.*, vol. 17, pp. 713–721, 1996.
- [39] S. M. Belkasim, M. Shridhar, and M. Ahmadi, "Pattern classification using an efficient KNNR," *Pattern Recogn.*, vol. 25, pp. 1269–1274, 1992.
- [40] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Mag.*, pp. 47–60, Nov. 1996.
- [41] M. L. Corner and E. J. Delp, "The EM/MPM algorithm for segmentation of textured images: analysis and further experiment results," *IEEE Trans. Image Processing*, vol. 9, pp. 1731–1744, Oct. 2000.
- [42] L. A. Christopher, E. J. Delp, C. R. Meyer, and P. L. Carson, "3-D Bayesian ultrasound breast image segmentation using the EM/MPM algorithm," presented at the Int. Symp. Biomedical Imaging, Washington, DC, 2002.
- [43] P. Barone, A. Frigessi, and M. Piccioni, *Stochastic Models, Statistical Methods, and Algorithms in Image Analysis*. Berlin, Germany: Springer-Verlag, 1992.
- [44] G. R. Cross and A. K. Jain, "Markov random field texture models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-5, pp. 25–39, Jan. 1983.
- [45] T. Simchony, R. Chellappa, and Z. Lichtenstein, "Relaxation algorithms for MAP estimation of gray-level images with multiplicative noise," *IEEE Trans. Inform. Theory*, vol. 36, pp. 608–613, Mar. 1990.
- [46] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading, MA: Addison-Wesley, 1989.
- [47] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Stat. Soc. B*, vol. 48, no. 3, pp. 259–302, 1986.
- [48] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *J. Amer. Statist. Assoc.*, vol. 82, pp. 76–89, 1987.
- [49] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Roy. Statist. Soc. B*, vol. 36, no. 2, pp. 192–236, 1974.
- [50] X. Zeng, L. H. Staib, R. T. Schultz, and J. S. Duncan, "Segmentation and measurement of the cortex from 3-D MR images using coupled-surfaces propagation," *IEEE Trans. Med. Imag.*, vol. 18, pp. 927–937, Oct. 1999.
- [51] D. H. Lawrence, K. W. Fleischer, and A. H. Barr, "Partial volume Bayesian classification of material mixtures in MR volume data using voxel histograms," *IEEE Trans. Med. Imag.*, vol. 17, pp. 74–86, Jan. 1998.
- [52] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and Bayesian restoration of image," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721–741, Jan. 1984.
- [53] J. Zhang, J. W. Modestino, and D. A. Langan, "Maximum-likelihood parameter estimation for unsupervised model based image segmentation," *IEEE Trans. Image Processing*, vol. 3, pp. 404–420, Mar. 1994.
- [54] C. R. Meyer, J. L. Boes, B. Kim, and P. Bland *et al.*, "Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin plate spline warped geometric deformations," *Med. Image Anal.*, vol. 3, pp. 195–206, 1997.
- [55] M. Holden, D. L. G. Hill, E. R. E. Denton, and J. M. Jarosz *et al.*, "Voxel similarity measures for 3D serial MR brain image registration," *IEEE Trans. Med. Imag.*, vol. 19, pp. 94–102, Jan. 2000.
- [56] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, "Medical image registration," *Phys. Med. Biol.*, vol. 46, pp. r1–r45, 2001.
- [57] J. Montagnat, H. Delingette, and N. Ayache, "A review of deformable surfaces: topology, geometry and deformation," *Image Vision Comput.*, vol. 19, no. 14, pp. 1023–1040, 2001.