MCMC Curve Sampling for Image Segmentation

Ayres C. Fan[1], John W. Fisher III[1,2], William M. Wells III[2,3], James J. Levitt[3,4], and Alan S. Willsky[1]

[1] Laboratory for Information and Decision Systems, MIT, Cambridge, MA
[2] Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA
[3] Brigham and Women's Hospital, Harvard Medical School, Boston, MA
[4] Dept. of Psychiatry, VA Boston HCS, Harvard Medical School, Brockton, MA
fan@mit.edu

Abstract. We present an algorithm to generate samples from probability distributions on the space of curves. We view a traditional curve evolution energy functional as a negative log probability distribution and sample from it using a Markov chain Monte Carlo (MCMC) algorithm. We define a proposal distribution by generating smooth perturbations to the normal of the curve and show how to compute the transition probabilities to ensure that the samples come from the posterior distribution. We demonstrate some advantages of sampling methods such as robustness to local minima, better characterization of multi-modal distributions, access to some measures of estimation error, and ability to easily incorporate constraints on the curve.

1 Introduction

Curve evolution methods are a class of algorithms which seek to segment an image I with a curve $\boldsymbol{C}$ by finding a local optimum of a given energy functional $E(\boldsymbol{C}; I)$. In general, having a single local optimum provides little insight as to how close the result is to the global optimum or how confident one should be in the answer. For low signal-to-noise ratio (SNR) or ill-posed problems, there are many local optima, and there can be multiple answers that plausibly explain the data. A common alternative is to view the problem as one of probabilistic inference by viewing $E(\boldsymbol{C}; I)$ as the negative log of a probability density:

$$\mathrm{p}(\boldsymbol{C} \mid I) \propto \exp(-E(\boldsymbol{C}; I)) \ . \tag{1}$$

Having a probabilistic interpretation allows the use of many standard inference algorithms such as stochastic optimization [1], particle filtering [2], or Markov chain Monte Carlo (MCMC) methods [3] to avoid local minima.

We propose an algorithm to draw samples from $\mathrm{p}(\boldsymbol{C} \mid I)$ which is, in general, a complex distribution and non-trivial to sample from. Samples are useful because not only can they help avoid local minima, they can also be used to characterize multi-modal distributions and estimation uncertainty by more fully exploring the configuration space. We will show examples of noisy images where the global maximum *a posteriori* estimate does not provide a satisfactory segmentation due

N. Ayache, S. Ourselin, A. Maeder (Eds.): MICCAI 2007, Part II, LNCS 4792, pp. 477–485, 2007.

to the large amount of noise, whereas a constellation of samples can help provide greater information as to likely locations for the true segmentation.

MCMC methods [4,5] were developed for situations when one wishes to draw samples from a distribution, but it is not possible to do so directly. Instead a proposal distribution q is defined, and samples from q are accepted in such a way as to guarantee that samples from p are generated asymptotically. They have been widely used for image segmentation since Geman and Geman [6] used a MCMC approach to segment images with a Markov random field (MRF) model. The advantage of sampling curves instead of from MRFs is that curve sampling is an inherently geometric process that enables one to work explicitly in the space of shapes and encode statistical properties of shape directly into the model such as global object characteristics. Tu and Zhu [3] also propose an approach using MCMC, but their primary focus is on finding global optima using a simulated annealing approach, and they do not generate large numbers of samples from the posterior distribution. One of our key results is to show how to ensure that detailed balance holds when sampling from the space of closed curves (a necessity to ensure that we asymptotically generate true samples from the posterior) and how to adapt these sampling methods to use user input to perform conditional simulation in order to reduce the estimation variance.

2 Curve Evolution Methods

Given an image domain $\Omega \subset \mathbb{R}^2$, a scalar-valued image $I : \Omega \to \mathbb{R}$, and a closed curve $\boldsymbol{C} : [0,1] \to \Omega$, active contour methods are formulated by specifying an energy functional $E(\boldsymbol{C} \,|\, I)$ and evolving the curve according to the gradient descent of that functional. Introducing an artificial time variable t, this results in a geometric PDE of the form $\frac{\partial \boldsymbol{C}}{\partial t}(s) = f(s)\boldsymbol{\mathcal{N}}_{\boldsymbol{C}}(s)$ where $f(s)$ is a force function and $\boldsymbol{\mathcal{N}}_{\boldsymbol{C}}(s)$ is the outward normal to the curve. A classical energy functional is Euclidean curve length: $E(\boldsymbol{C} \,|\, I) = \oint_{\boldsymbol{C}} \mathrm{d}s$ with $\mathrm{d}s$ being the differential arc length along the curve. The resulting force function is $f(s) = -\kappa_{\boldsymbol{C}}(s)$ where $\kappa_{\boldsymbol{C}}$ is curvature. This flow has a smoothing effect on the curve [7] and is typically used as a regularization term. Region-based energy functionals (*e.g.*, Chan-Vese [8] and Mumford-Shah [9]) separate regions using the image statistics and are now widely used for image segmentation due to their robustness to noise.

The earliest curve evolution methods by Kass *et al.* [10] tracked discrete marker points on the contour. Level set methods [7] were later introduced to more naturally handle topological changes and reduce reinitialization problems. With level sets, a surface $\Psi(\boldsymbol{x})$ is created whose zeroth level set is the curve: $\Psi(\boldsymbol{C}(s)) = 0 \;\forall\; s \in [0,1]$. By convention, Ψ is negative inside the curve and positive outside the curve. To ensure that the zeroth level set of Ψ tracks $\boldsymbol{C}$, we need: $\frac{\partial \Psi}{\partial t} = -\frac{\partial \boldsymbol{C}}{\partial t} \cdot \nabla\Psi$. As $\frac{\partial \boldsymbol{C}}{\partial t}$ is only defined on the zeroth level set of Ψ, the PDE is extended to the rest of Ω by a technique known as *velocity extension* [7].

3 Formulation

For MCMC methods, an ergodic Markov chain with $\mathrm{p}(\boldsymbol{C} \,|\, I)$ as its stationary distribution is constructed [5], so simulating the chain has the probability distribution of the state asymptotically approach $\mathrm{p}(\boldsymbol{C} \,|\, I)$ for any initial state $\boldsymbol{C}_0$. The chain's transition probability $\mathrm{T}(\boldsymbol{C}^{(t)} \rightarrow \boldsymbol{C}^{(t+1)})$ is the product of a proposal distribution $\mathrm{q}(\boldsymbol{\Gamma}^{(t+1)} \,|\, \boldsymbol{C}^{(t)})$ and an acceptance probability function $\mathrm{a}(\boldsymbol{\Gamma}^{(t+1)} \,|\, \boldsymbol{C}^{(t)})$. A sample from $\mathrm{T}(\boldsymbol{C}^{(t)} \rightarrow \boldsymbol{C}^{(t+1)})$ is drawn by accepting a sample $\boldsymbol{\Gamma}^{(t+1)}$ from $\mathrm{q}(\boldsymbol{\Gamma} \,|\, \boldsymbol{C}^{(t)})$ with probability $\mathrm{a}(\boldsymbol{\Gamma}^{(t+1)} \,|\, \boldsymbol{C}^{(t)})$ otherwise $\boldsymbol{C}^{(t+1)} = \boldsymbol{C}^{(t)}$.

The proposal distribution is chosen so as to be easy to sample from, as MCMC methods change the problem of sampling from p to one of drawing many samples from q. For discrete state spaces, a sufficient condition for $\mathrm{p}(\boldsymbol{C}|I)$ to be the stationary distribution of the chain is *detailed balance*:

$$\mathrm{p}(\boldsymbol{C}^{(t)}|I)\mathrm{T}(\boldsymbol{C}^{(t)} \rightarrow \boldsymbol{C}^{(t+1)}) = \mathrm{p}(\boldsymbol{C}^{(t+1)}|I)\mathrm{T}(\boldsymbol{C}^{(t+1)} \rightarrow \boldsymbol{C}^{(t)}) \ . \tag{2}$$

For continuous state spaces, a similar statement can generally be made [5]. A common acceptance rule is Metropolis-Hastings [11]. For an iterate $\boldsymbol{C}^{(t)}$ and a candidate sample $\boldsymbol{\Gamma}^{(t+1)}$, the Metropolis-Hastings acceptance probability is defined as $\mathrm{a}(\boldsymbol{\Gamma}^{(t+1)}|\boldsymbol{C}^{(t)}) = \min\left(1, \eta(\boldsymbol{\Gamma}^{(t+1)}|\boldsymbol{C}^{(t)})\right)$ where the Hastings ratio η is $\eta(\boldsymbol{\Gamma}^{(t+1)}|\boldsymbol{C}^{(t)}) = \mathrm{p}(\boldsymbol{\Gamma}^{(t+1)})\mathrm{q}(\boldsymbol{C}^{(t)} \,|\, \boldsymbol{\Gamma}^{(t+1)})/\mathrm{p}(\boldsymbol{C}^{(t)})\mathrm{q}(\boldsymbol{\Gamma}^{(t+1)} \,|\, \boldsymbol{C}^{(t)})$.

The algorithmic steps for a Metropolis-Hastings sampler involve: 1) sample from $\mathrm{q}(\boldsymbol{\Gamma}^{(t+1)}|\boldsymbol{C}^{(t)})$; 2) evaluate $\mathrm{a}(\boldsymbol{\Gamma}^{(t+1)}|\boldsymbol{C}^{(t)})$; 3) accept or reject $\boldsymbol{\Gamma}^{(t+1)}$.

3.1 Proposal Distribution

We implicitly define q by explicitly defining how to sample from it. To generate a candidate sample $\boldsymbol{\Gamma}^{(t+1)}$, we randomly perturb the previous iterate $\boldsymbol{C}^{(t)}$:

$$\boldsymbol{\Gamma}^{(t+1)}(s) = \boldsymbol{C}^{(t)}(s) + f^{(t+1)}(s)\boldsymbol{\mathcal{N}}_{\boldsymbol{C}^{(t)}}(s)\delta t \tag{3}$$

where $f^{(t+1)}(s)$ is a random field. The problem of generating $\boldsymbol{\Gamma}^{(t+1)}$ is now the problem of generating $f^{(t+1)}(s)$.

In this work, we focus on generating $f^{(t)}(s)$ composed of a correlated zero-mean Gaussian random process $r^{(t)}(s)$ and a mean function $\mu^{(t)}(s)$: $f^{(t)}(s) = \mu^{(t)}(s) + r^{(t)}(s)$. We construct $r^{(t)}(s)$ by circularly convolving white Gaussian noise $n^{(t)}(s)$ with a smoothing kernel $h(s)$ (*e.g.*, a Gaussian kernel). Many other choices for generating $r(s)$ are possible such as using Fourier or wavelet bases.

The mean process is chosen to increase the convergence rate of the sampling algorithm. Here, we define it as $\mu^{(t)}(s) = -\kappa_{\boldsymbol{C}^{(t)}}(s) + \gamma^{(t)}$ where $\gamma^{(t)}$ is a positive inflation term counteracting the curve shortening term $\kappa_{\boldsymbol{C}}$. As discussed earlier, $f(s) = -\kappa_{\boldsymbol{C}}(s)$ is a regularizing flow which creates smooth curves, so this biases our proposal distribution to create smooth curves.

3.2 Detailed Balance

Metropolis-Hastings sampling requires that we be able to evaluate both the forward and reverse proposal distributions $\mathrm{q}(\boldsymbol{\Gamma}^{(t+1)}|\boldsymbol{C}^{(t)})$ and $\mathrm{q}(\boldsymbol{C}^{(t)}|\boldsymbol{\Gamma}^{(t+1)})$.

This computation needs to be understood in order to ensure detailed balance and guarantee that our samples come from the posterior. For our curve perturbations, this is non-trivial because q is asymmetric due to the mean component.

The perturbation defined in equation (3) is a differential in the direction of the normal, so each random perturbation maps one curve uniquely to another. This remains approximately true for small finite δt. Thus evaluating $\mathrm{q}(\boldsymbol{\Gamma} \,|\, \boldsymbol{C})$ is equivalent to evaluating the probability of generating $f(s)$. To implement (3), we generate a noise vector $\boldsymbol{n}$ of fixed length, multiply it by a circulant matrix $\boldsymbol{H}$ (which implements the circular convolution), and add a mean vector $\boldsymbol{\mu}$. This results in a Gaussian random vector $\boldsymbol{f} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{H}\boldsymbol{H}^{\mathrm{T}})$. Note that $\boldsymbol{f}$ is deterministically generated from $\boldsymbol{n}$ so $\mathrm{p}_{\mathbf{f}}(\boldsymbol{H}\boldsymbol{n} + \boldsymbol{\mu}) = \mathrm{p}_{\mathbf{n}}(\boldsymbol{n}) \propto \exp(-\frac{1}{2}\boldsymbol{n}^{\mathrm{T}}\boldsymbol{n})$.

To compute the probability of the reverse perturbation, we construct the analog to equation (3):

$$\boldsymbol{C}^{(t)}(s) = \boldsymbol{\Gamma}^{(t+1)}(s) + g^{(t+1)}(s)\boldsymbol{\mathcal{N}}_{\boldsymbol{\Gamma}^{(t+1)}}(s)\delta t \ , \tag{4}$$

and the reverse perturbation probability is the probability of generating $g^{(t+1)}$. For small δt, a reasonable estimate is $g^{(t+1)}(s) \approx -f^{(t+1)}(s)/\boldsymbol{\mathcal{N}}_{\boldsymbol{C}^{(t)}}(s)\cdot\boldsymbol{\mathcal{N}}_{\boldsymbol{\Gamma}^{(t+1)}}(s)$ which is obtained using locally-linear approximations to $\boldsymbol{\Gamma}^{(t+1)}$ and $\boldsymbol{C}^{(t)}$. Note that this explicit correspondence we construct here means that even though we implement the perturbations using level sets, topological change is not valid for our chain. To allow splitting or merging of regions, a jump-diffusion process must be used [3]. Once we have computed $g(s)$, we can write it as $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{n}' + \boldsymbol{\mu}'$ (obtaining $\boldsymbol{\mu}'$ from $\boldsymbol{\Gamma}^{(t+1)}$ exactly as we obtain $\boldsymbol{\mu}$ from $\boldsymbol{C}^{(t)}$) and compute its probability as $\mathrm{p}_{\mathbf{g}}(\boldsymbol{H}\boldsymbol{n}' + \boldsymbol{\mu}') = \mathrm{p}_{\mathbf{n}'}(\boldsymbol{n}') \propto \exp(-\frac{1}{2}\boldsymbol{n}'^{\mathrm{T}}\boldsymbol{n}')$.

We can then use the analysis we just performed (a detail that most implementations ignore) to ensure detailed balance by computing the ratio of the forward and reverse proposal distributions (for use in the acceptance rule) as

$$\frac{\mathrm{q}(\boldsymbol{C}^{(t)} \,|\, \boldsymbol{\Gamma}^{(t+1)})}{\mathrm{q}(\boldsymbol{\Gamma}^{(t+1)} \,|\, \boldsymbol{C}^{(t)})} = \exp\left(-\tfrac{1}{2}\left[\boldsymbol{n}'^{\mathrm{T}}\boldsymbol{n}' - \boldsymbol{n}^{\mathrm{T}}\boldsymbol{n}\right]\right) \tag{5}$$

with $\boldsymbol{n} = \boldsymbol{H}^{-1}(\boldsymbol{f} - \boldsymbol{\mu})$ and $\boldsymbol{n}' = \boldsymbol{H}^{-1}(\boldsymbol{g} - \boldsymbol{\mu}')$.

3.3 Conditional Simulation

In many application domains of interest, segmentations are currently performed by an expert. Rather than trying to remove them completely from the loop, we can create a feedback system that allows them to focus their expertise and knowledge on the most difficult portions of the problem. With an optimization-based approach, this would require one to do constrained optimization which is hard for high-dimensional problems. With a sampling approach, we can use a technique known as *conditional simulation* where we fix part of the state space and sample from the distribution conditioned on the known part.

Let $\boldsymbol{C}_k : [0, \beta] \rightarrow \Omega$ be the known part of the curve ($\beta \in [0, 1]$) and $\boldsymbol{C}_u : [\beta, 1] \rightarrow \Omega$ be the unknown part of the curve. Then $\boldsymbol{C}(s) = \boldsymbol{C}_k(s)$ and $\boldsymbol{C}(s) = \boldsymbol{C}_u(s)$ on $[0, \beta]$ and $[\beta, 1]$ respectively. It is straightforward to generalize this approach for multiple fixed intervals. We wish to sample from $\mathrm{p}(\boldsymbol{C}_u | I, \boldsymbol{C}_k) = \mathrm{p}(\boldsymbol{C}_u, \boldsymbol{C}_k | I)/\mathrm{p}(\boldsymbol{C}_k) \propto \mathrm{p}(\boldsymbol{C} \,|\, I)$. Thus, we can see that computing the conditional target distribution is unchanged (except part of $\boldsymbol{C}$ no longer changes).

To ensure that our samples from our proposal distribution stay on the manifold of curves that contain $\boldsymbol{C}_k$, we need to modify our proposal distribution to impose zero variance on $\boldsymbol{C}_k$. A simple way to implement this is to multiply the random perturbations defined in Sec. 3.1 by a scalar field: $\tilde{r}(s) = d(s)r(s)$ with $d(s) = 0$ for $s \in [0, \beta]$, $d(s) = 1$ for $s \in [\beta + \epsilon, 1 - \epsilon]$ and $\epsilon > 0$. From $(\beta, \beta + \epsilon]$ and $[1 - \epsilon, 1)$, $d(s)$ smoothly transitions from 0 to 1 so there is not a strong variance mismatch at the end points of $\boldsymbol{C}_k$.

Computationally, this is equivalent to multiplying our random vector $\boldsymbol{r}$ by a diagonal matrix $\boldsymbol{D}$ resulting in $\tilde{\boldsymbol{r}} \sim \mathrm{N}(\boldsymbol{D}\boldsymbol{\mu}, \boldsymbol{D}\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{D})$. This is a degenerate probability distribution as some entries of $\boldsymbol{r}$ have zero variance, so we should only evaluate $\mathrm{q}(\cdot \,|\, \cdot)$ using the the perturbation $\boldsymbol{r}_u$ on the unknown part of the curve. Otherwise the computation is identical to that described in Sec. 3.2.

4 Results

In this section we present results on a prostate magnetic resonance (MR) example and a thalamus MR segmentation problem. For each application, we generated 1000 samples from $\mathrm{p}(\boldsymbol{C} \,|\, I)$. Computation time per sample ranged from 10-30 seconds for 256×256 images on a 2 GHz Opteron workstation. Each sample is generated independently of the others, so sample throughput can be easily increased using parallel computers.

For both examples, we assume that pixels are independent and identically distributed (iid) given the curve and learn (from segmented training data) non-parametric histogram distributions $\mathrm{p}(I(\boldsymbol{x})|0)$ and $\mathrm{p}(I(\boldsymbol{x})|1)$ for the intensity distribution outside and inside the curve respectively (shown in in Figs. 1(a) and 2(a)). Using the Heaviside (or indicator) function $\mathcal{H}$ and a curve length prior, this results in an overall posterior probability of:

$$\mathrm{p}(\boldsymbol{C} \,|\, I) = \exp(-\alpha \oint_{\boldsymbol{C}} \mathrm{d}s) \prod_{\boldsymbol{x}} \mathrm{p}\left(I(\boldsymbol{x}) \,|\, \mathcal{H}(-\Psi_{\boldsymbol{C}}(\boldsymbol{x}))\right) \quad . \tag{6}$$

To display the samples, we will use three main visualization approaches that are only possible because we are able to draw a large number of statistically-correct samples from the posterior distribution:

1. Displaying the highest probability samples (*e.g.*, Fig. 1(b)). The most likely samples can be viewed as proxies for what a global optimizer would find.
2. Histogram images (*e.g.*, Fig. 1(c)). For each $\boldsymbol{x}$ we count the number of $\boldsymbol{C}_i$ for which $\boldsymbol{x}$ is inside the curve (*i.e.*, $\Psi_{\boldsymbol{C}_i}(\boldsymbol{x}) < 0$). This is thus the marginal distribution over segmentation labels at each $\boldsymbol{x}$.

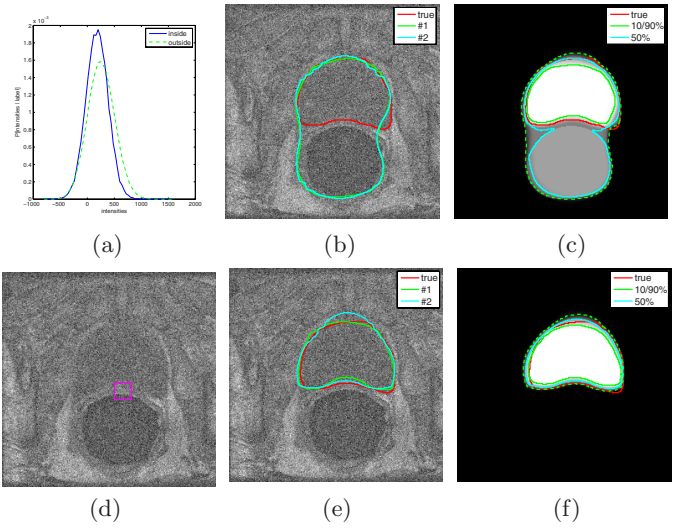

Fig. 1. Prostate segmentation using non-parametric intensity distributions. (a) Pixel intensities for each class. (d) Initial curve. (b) Two most likely samples (very different from the correct curve). (c) Marginal confidence bounds and histogram image. (e)-(f) Most likely samples and marginal bounds for prostate-only cluster.

3. Marginal confidence bounds (*e.g.*, Fig. 1(c)). Given a histogram image, we plot the level contours. These can be viewed as confidence bounds (*e.g.*, the 10% confidence bound is the contour outside of which all pixels were inside fewer than 10% of the samples). The 50% confidence bound can be viewed as being analogous to a median contour.

Confidence bounds have been used in some previous image segmentation or reconstruction approaches [12], but those dealt with parametric shape representations (so the uncertainty was over a finite set of parameters). It is important to note that our confidence representations are marginal statistics from infinite-dimensional non-parametric shape distributions.

4.1 Prostate Segmentation

In Fig. 1, we show results from a noisy T1-weighted prostate MR image. The histogram image and the marginal confidence bounds in Fig. 1(c) show this distribution has three primary modes: one around the correct prostate segmentation (the red contour); one containing only the rectum (the dark region beneath the prostate); and one encompassing both the prostate and the rectum. As can be seen in Fig. 1(b), the most likely mode contains the curves that segment the two regions together, and this is what a gradient-based curve evolution implementation of (6) also finds. The reason for this can be seen in the image intensity likelihoods in Fig. 1(a). Due to the noise and our simple iid model, the model prefers all pixels with intensity below some threshold (including the rectum) to

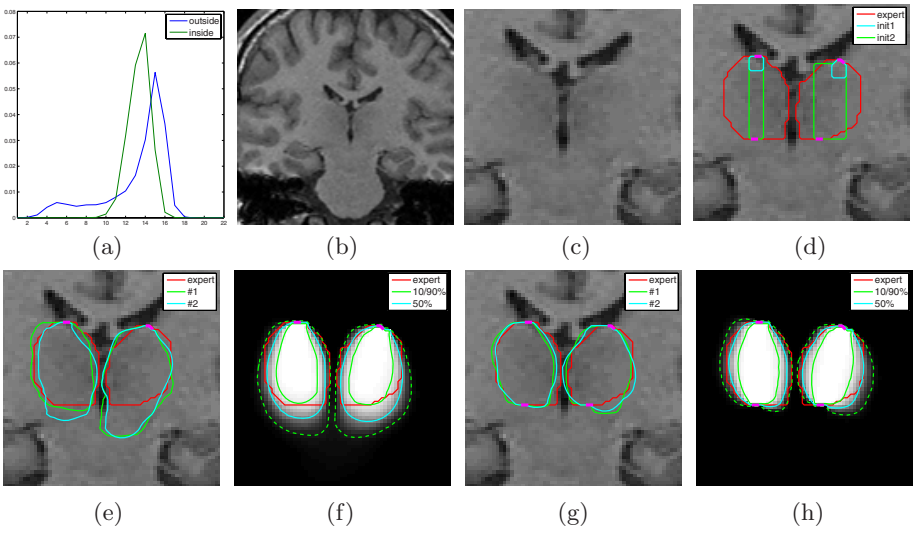

Fig. 2. Conditionally-simulated thalamus segmentation using non-parametric intensity distributions. (a) Pixel intensities for each class. (b)-(c) Observed image (original and zoomed). (d) Expert and initial curves. Two most likely samples ((e) and (g)) and marginal confidence bounds and histogram image ((f) and (h)) with a point on the top fixed and points on the top and the bottom fixed respectively.

be inside the curve. The sampling process enables us to see the multiple possible solutions.

Without having any additional *a priori* information, it would be difficult to say which of these three scenarios is the correct one. In fact, it is possible in some applications where multiple modes all provide reasonable explanations of the data. One approach we can take here is to utilize the information our sampling procedure provides to us. While the aggregate marginal statistics do not appear to be providing very useful information (though the 90% confidence boundary is located within the true prostate boundary), it is easy to create three clusters of samples. An expert user or a shape-driven classifier could then pick the correct cluster. We show the most-likely samples and the marginal confidence boundaries for the prostate-only cluster in Fig. 1(e) and (f).

4.2 Thalamus Segmentation

Segmenting sub-cortical structures is a challenging task due to the low amount of contrast between tissue types. One approach to reduce the ill-posedness involves using strong prior shape models [13]. As there is too little contrast for an unconstrained approach to succeed here, we apply the conditional simulation version of our approach and specify small portions of the curve *a priori* (indicated with the magenta line segments in Fig. 2).

We begin by fixing a small portion of the top of each half of the thalamus and generating samples conditioned on that information. Two separate level sets are evolved for each half of the thalamus. The two most likely samples in Fig. 2(e) correctly segment most of the thalamus except the bottom which is least constrained by the fixed portion at the top. Note, though, that the marginal confidence bounds in Fig. 2(f) show that the expert contour location is mostly bracketed between the 90% and 10% confidence contours, and the median contour is quite close to the expert-segmented boundary location.

Note that the sampling method actually provides information about where the greatest uncertainty is and, thus, where expert assistance is most needed. We can see in Fig. 2(f) that there is a more diffuse histogram image (and a larger gap between the confidence bounds) at the bottom of the thalamus indicating a greater amount of sample variability. In Fig. 2(g)-(h), we take the knowledge gained from the first experiment and interactively revise the information provided to the sampler by specifying a location on the bottom of the thalamus as well. With this additional information, the most likely samples are now both reasonable, and the estimation variance is greatly reduced.

5 Conclusion

In this paper, we presented an approach to generate samples from probability distributions defined on spaces of curves by constructing a MCMC algorithm and showing how to properly compute the proposal distribution probabilities to ensure detailed balance and asymptotic convergence to a desired posterior distribution. The sampling approach provided robustness to local minima in low-SNR and ill-posed problems, and we showed how a large number of curve samples can be used to provide useful aggregate statistics (such as non-parametric marginal confidence bounds) about the likely location of the true curve locations. We demonstrated the usefulness of this aggregate information even when the most likely curves were not providing satisfactory segmentations, and we showed how constraints can be easily imposed on the samples (unlike a gradient-based optimization method) to provide a semi-automatic segmentation approach.

Future work in this space involves developing faster sampling algorithms by utilizing better proposal distributions or multiresolution methods; extending the framework to non-closed curves, unknown topology, and volumetric segmentation; and creating uncertainty measures that provide information about the local characteristics of the shape manifold.

Acknowledgments. We wish to thank Clare Tempany and Martha Shenton for their assistance in acquiring image data. Our research was primarily supported by a grant from Shell International Exploration and Production, Inc. with additional support from NSF grant EEC9731748 and NIH grants U41-RR019703, P41-RR13218, R01-CA109246, R01-CA111288, K02-MH01110, R01-MH50747, and U54-EB005149.

References

1. Juan, O., Keriven, R., Postelnicu, G.: Stochastic motion and the level set method in computer vision: Stochastic active contours. Intl. J. Comp. Vis. 69(1) (2006)
2. de Bruijne, M., Nielsen, M.: Shape particle filtering for image segmentation. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3216, pp. 168–175. Springer, Heidelberg (2004)
3. Tu, Z., Zhu, S.C.: Image segmentation by data-driven Markov chain Monte Carlo. IEEE Trans. Patt. Anal. Mach. Intell. 24(5), 657–673 (2002)
4. Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E.: Equations of state calculations by fast computing machines. J. Chem. Phys. 21(6) (1953)
5. Neal, R.M.: Probabilistic inference using Markov chain Monte Carlo methods. Technical Report CRG-TR-93-1, Univ. of Toronto (1993)
6. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE PAMI 6, 721–741 (1984)
7. Sethian, J.: Level Set Methods and Fast Marching Methods. Cambridge University Press, Cambridge (1999)
8. Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Trans. Imag. Proc. 10, 266–277 (2001)
9. Tsai, A., Yezzi, A., Willsky, A.: Curve evolution implementation of the Mumford-Shah functional. IEEE Trans. Imag. Proc. 10(8), 1169–1186 (2001)
10. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. Intl. J. Comp. Vis. 1(4), 321–331 (1988)
11. Hastings, W.K.: Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57(1), 97–109 (1970)
12. Ye, J.C., Bresler, Y., Moulin, P.: Asymptotic global confidence regions in parametric shape estimation problems. IEEE Trans. Inf. Theory 46(5), 1881–1895 (2000)
13. Pohl, K.M., Fisher, J., Grimson, W.E.L., Kikinis, R., Wells, W.: A Bayesian model for joint segmentation and registration. Neuroimage 31, 228–239 (2006)