Markov Chain Monte Carlo Shape Sampling using Level Sets

Siqi Chen and Richard J. Radke*

Department of Electrical, Computer and Systems Engineering

Rensselaer Polytechnic Institute, Troy, NY 12180

chens@rpi.edu, rjradke@ecse.rpi.edu

Abstract

In this paper, we show how the Metropolis-Hastings algorithm can be used to sample shapes from a distribution defined over the space of signed distance functions. We extend the basic random walk Metropolis-Hastings method to high-dimensional curves using a proposal distribution that can simultaneously maintain the signed distance function property and the ergodic requirement. We show that detailed balance is approximately satisfied and that the Markov chain will asymptotically converge. A key advantage of our approach is that the shape representation is implicit throughout the process, as compared to existing work where explicit curve parameterization is required. Furthermore, our framework can be carried over to 3D situations easily. We show several applications of the framework to shape sampling from multimodal distributions and medical image segmentation.

1. Introduction

In radiotherapy planning, a planning Computed Tomography (CT) scan is usually given to a patient on the first day of his/her arrival. The target organ is outlined by a physician and a comprehensive treatment plan is calculated based on these contours- a time-consuming process. Due to the organ's deformation, on each day of the radiotherapy, the contours will no longer be accurate, resulting in a suboptimal delivery of radiation. Our motivating problem is to define a probability distribution over possible shapes of the target organ, sample shapes from it, precompute corresponding treatment plans, and select the best matching plan on the day of treatment. We hope that this approach will improve the accuracy and efficacy of treatment.

Therefore, we are interested in the problem of generating likely shapes from a target distribution $\pi(x)$. Generating samples from arbitrary distributions $\pi(x)$ plays a

fundamental role in Bayesian inference, optimization and simulation problems. Such problems have important applications in machine learning, physics, statistics and economics [14]. For relatively simple and low-dimensional distributions, standard sampling techniques such as matching cumulative distributions [3] can be used. However, when the distribution is complex and higher-dimensional, iterative techniques such as rejection sampling and importance sampling need to be applied [3]. These techniques typically require a "proposal distribution" that is both easy to sample from and similar to $\pi(x)$. It is often difficult to satisfy both requirements, and when such a proposal distribution is not obtainable, a more complicated algorithm known as Markov Chain Monte Carlo (MCMC) is used [14].

In this paper, we propose an MCMC-based method of sampling simple closed curves from an arbitrary distribution defined over the space of signed distance functions. The main advantage of our technique is that the shape representation is implicit throughout the sampling process, in contrast to existing work where explicit curve parameterization is required.

The rest of the paper is organized as follows. In Section 2 we review related work on geometric level sets and MCMC in computer vision. Section 3 gives a brief overview of MCMC theory and the Metropolis-Hastings algorithm. Section 4 presents the main shape sampling formulation using our new approach. Section 5 describes several experiments on shape sampling and image segmentation using the new formulation. Section 6 concludes the paper with discussion and ideas for future work.

2. Related Work

2.1. Geometric Level Sets

To avoid problems associated with explicit shape representations (e.g., reparametrization [19]), it is common to represent a contour C in a domain Ω using the zero level set of a higher level embedding function $\phi : \Omega \to \Re$. In order to guarantee a unique embedding, we assume that ϕ is a signed distance function (SDF) with $\phi < 0$ inside the shape,

$\phi > 0$ outside the shape and $|\nabla\phi| = 1$ almost everywhere [27].

Much work on geometric level set methods has been proposed to solve a large number of computer vision problems. Leventon et al. [22] proposed to represent shapes by SDFs and perform Principal Component Analysis (PCA) on the set of training SDFs. Various groups have proposed to use a distance measure between the evolving level set and a reference level set as a shape prior term [5, 10, 31] to guide the curve evolution in segmentation problems. Recently, Cremers et al. [8] and Kim et al. [20] modeled shape variation with such distance measures and the nonparametric technique of Kernel Density Estimation (KDE). Zhao et al. [36] addressed shape reconstruction from unorganized points with a weighted minimal-surface-like model. Breen and Whitaker [4] optimized a similarity measure between a deforming surface and a target to achieve 3D shape metamorphosis. Variational level set methods have also been used in image inpainting, denoising, registration and many other computer vision problems [26]. Typically, these approaches all consider the problem of local optimization, using the gradient descent technique to optimize a given functional starting from an initial shape. This is in significant contrast to the MCMC method, in which not only can a global optimum be reached, but also the underlying distribution can be explored.

2.2. Markov-Chain Monte Carlo

Metropolis et al. [23] and Hastings [17] formed the groundbreaking Markov-Chain Monte Carlo (MCMC) work known as the Metropolis-Hastings (MH) algorithm. Geman and Geman [13] were the first to apply MCMC to computer vision problems using the Gibbs Sampler in Markov Random Fields (MRFs). Grenander and Miller [16] proposed the jump and diffusion MCMC approaches. Green [15] further proposed the reversible jump MCMC, which was applied to sampling and learning by Zhu and Yuille [37]. Storvik [33] used simulated annealing to solve active contour problems. Later, Tu and Zhu [34] proposed a data-driven MCMC-based image segmentation framework. Allassonniere et al. [1] used a MCMC based stochastic version of the EM algorithm to estimate deformable template parameters in shape modeling problems.

Recently, Juan et al. [18] and Law et al. [21] both proposed stochastic optimization methods for a level-set-based segmentation algorithm in order to avoid local minima. However, the primary focus of these approaches was looking for a global optimum instead of drawing samples from the shape distribution, so they did not need to maintain the detailed balance condition.

The most related work to what we propose here was presented by Fan et al. [11, 12]. Both approaches use the MH algorithm to sample curves from an arbitrary distribution. Compared to their work, our framework has two advantages:

1. The proposal distribution in their work requires explicit parametrization of the curve, which is both complicated and time-consuming. In our work, sampling is done entirely in the implicit representation. Due to the explicit parametrization, their method would be difficult to extend to full 3D (e.g., volumetric medical images). Working in the implicit representation allows us to easily extend our work to 3D.

2. We present a new proposal distribution that can simultaneously satisfy the signed distance function property and the requirements of MCMC. This opens up the possibility of incorporating more advanced proposal distributions from MCMC theory into our method that can give better mixing.

3. MCMC Basics

To estimate a target distribution $\pi(x)$, MCMC uses the previous sample x^{i-1} to randomly generate the next sample x^i, exploring the entire state space based on a Markov chain mechanism. The Markov chain is carefully constructed so that it will asymptotically converge to a stationary distribution (designed to be $\pi(x)$ up to a normalizing constant). The time before the stationary distribution is reached is called the "mixing time". A chain is said to be "poorly mixing" if it remains in one region of the state space for a long period of time, as opposed to a "well mixing" chain where all the modes are visited. In order to guarantee the asymptotic convergence to an invariant distribution, the Markov transition matrix T must obey two properties: irreducibility and aperiodicity, together known as the ergodic property. Furthermore, a sufficient, but not necessary, condition to ensure that $\pi(x)$ is the desired invariant distribution is the so-called detailed balance condition:

$$\pi(x^i)T(x^{i-1}|x^i) = \pi(x^{i-1})T(x^i|x^{i-1}) \qquad (1)$$

Typically, the transition matrix T is the composition of two operations. First, a candidate sample y^i is generated from a proposal distribution $q(y^i|x^{i-1})$, which is easy to sample from, but may not exhibit any resemblance to $\pi(x)$. Then y^i is accepted or rejected according to an acceptance function $a(y^i, x^{i-1})$ where $x^i = y^i$ if the sample is accepted or $x^i = x^{i-1}$ if the sample is rejected. Among MCMC methods, the Metropolis-Hastings (MH) algorithm [17, 23] given in Algorithm 1 is the most frequently used.

Algorithm 1 Metropolis-Hastings Algorithm

Initialize x^0
for $i = 1$ to N **do**
    Sample $u \sim U_{[0,1]}$
    Sample $y^i \sim q(y^i|x^{i-1})$
    if $u < a(y^i, x^{i-1}) = \min\left(1, \frac{\pi(y^i)q(x^{i-1}|y^i)}{\pi(x^{i-1})q(y^i|x^{i-1})}\right)$ **then**
        $x^i = y^i$
    else
        $x^i = x^{i-1}$
    end if
end for
---

4. Formulation

Our MCMC curve sampling framework is similar to the basic MH method called the symmetric random-walk Metropolis algorithm (RWM) [23]. In this case, the proposal distribution $q(y|x) = q(y - x)$ where q is a symmetric function about 0 such as a Gaussian. In other words, $y = x + r$ where $r \sim N[0, \sigma_{mag}]$ and σ_{mag} (or Σ_{mag} in the multivariate case) controls the perturbation, called the "scaling" in MCMC theory. Since we embed the curve C using a SDF ϕ, in a real implementation, the dimensionality of the domain Ω is the number of the grid points used to sample ϕ. It is impossible to perturb each point independently with a Gaussian while maintaining the SDF property that $|\nabla\phi| = 1$. In the following, we present a new proposal distribution in which different regions of the curve are updated simultaneously in order to maintain the SDF property, and show that the perturbation approximately satisfies the detailed balance and ergodic requirements.

4.1. Proposal Distribution

The proposal distribution in our work is similar to the RWM method. Let ϕ^{i-1} and ϕ^i be the level set representations of the curves C^{i-1} and C^i respectively, where C^i is locally perturbed from C^{i-1}. The perturbation can be expressed as:

$$\phi^i = \phi^{i-1} + r^{i-1} \qquad (2)$$

where r^{i-1} is also a function: $\Omega \rightarrow \Re$. It is critical to maintain the SDF property of ϕ^i when ϕ^{i-1} is a valid SDF. Zhao et al. [35] proved the following lemma:

Lemma 1: If ϕ^{i-1} is a SDF, then $\phi^i = \phi^{i-1} + r^{i-1}$ remains a SDF if

$$r^{i-1}(x) = r^{i-1}(F(x)) \qquad (3)$$

where

$$F(x) = x - \phi^{i-1}(x)\nabla\phi^{i-1}(x) \qquad (4)$$

That is, $r^{i-1}(x)$ must be a function of the point on C^{i-1} closest to x, also called the "foot point" of x. We specify a particular perturbation by choosing a source point S^{i-1} on C^{i-1} and locally "pushing" the curve in the normal direction at S^{i-1}. That is, we choose

$$r^{i-1}(x) = m\exp\left(-\frac{D^2(S^{i-1}, F(x))}{2\sigma_{bump}^2}\right) \qquad (5)$$

where D is the geodesic distance between S^{i-1} and $F(x)$ on the curve, and m is the magnitude of the perturbation. This means that points on the curve further away from S^{i-1} are perturbed less, depending on the variance σ_{bump}. m is drawn from another Gaussian distribution with mean 0 and variance σ_{mag}, which controls the magnitude of the noise and the direction of the perturbation, either inward or outward. We illustrate the perturbation with two examples in Fig. 1.
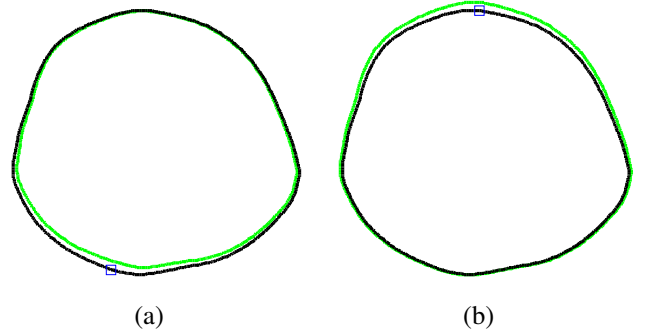

      (a)   (b)

Figure 1. Perturbation examples. The black curve is C^{i-1}, the green curve is C^i, and the source point is plotted as a blue square. (a) $m = 5$. (b) $m = -5$.

The ergodic condition requires both irreducibility and aperiodicity, and both of these properties are generally difficult to prove [14]. In our work, because the state space (simple closed curves) is so large and the perturbation is random, it is virtually impossible that the chain is periodic. For irreducibility, we need to show there is a positive probability of visiting any state starting from any other state. We heuristically argue that our MCMC framework is irreducible for 3 reasons. First of all, σ_{bump} is not fixed and allowed to change from perturbation to perturbation, so we can approximate sharp corners when $\sigma_{bump} \rightarrow 0$ and flat regions when $\sigma_{bump} \rightarrow \infty$. Second, the "foot point" S^{i-1} is chosen over the entire curve, which means any "part" of the curve can be reached. Finally, the perturbation magnitude m comes from a Gaussian distribution that extends from $-\infty$ to ∞, which means any "scale" of the curve can be reached. However, rigorous mathematical analysis is challenging and is left for future work.

4.2. Detailed Balance

In the MH algorithm, we must calculate both $q(\phi^i|\phi^{i-1})$ and $q(\phi^{i-1}|\phi^i)$. Given σ_{bump} at each perturbation, the only thing that affects ϕ^i given ϕ^{i-1} is the source point location S^{i-1} and the noise m. Since they are independent of each other, we have the following forward proposal:

$$q(\phi^i|\phi^{i-1}) = p(S^{i-1}) \cdot p(m) \qquad (6)$$

Computing the backward proposal is more difficult and we make some simplifying approximations. We construct the analog to (2):

$$\phi^{i-1} = \phi^i + r^i \qquad (7)$$

We conjecture that r^i can be approximated by $-m \exp\left(-\frac{D^2(S^i, F(x))}{2\sigma_{bump}^2}\right)$, where S^i is the foot point of S^{i-1} with respect to C^i.

Recall that

$$\phi^i = \phi^{i-1} + m \exp\left(-\frac{D^2(S^{i-1}, F(x))}{2\sigma_{bump}^2}\right) \qquad (8)$$

If the conjecture is correct, we can plug the above equation into (7) and show that

$$\phi^{i-1}(x) = \phi^{i-1}(x) + m \exp\left(-\frac{D^2(S^{i-1}, F_{i-1}(x))}{2\sigma_{bump}^2}\right)$$

$$-m \exp\left(-\frac{D^2(S^i, F_i(x))}{2\sigma_{bump}^2}\right) \qquad (9)$$

that is,

$$\exp\left(-\frac{D^2(S^{i-1}, F_{i-1}(x))}{2\sigma_{bump}^2}\right) = \exp\left(-\frac{D^2(S^i, F_i(x))}{2\sigma_{bump}^2}\right) \qquad (10)$$

where $F_{i-1}(x)$ and $F_i(x)$ are the foot points of x with respect to C^{i-1} and C^i. It can be shown that the line connecting S^i and S^{i-1} is orthogonal to both C^{i-1} and C^i [25]. Therefore, if x is on this line (like x_1 in Fig. 2), $F_{i-1}(x) = S^{i-1}$ and $F_i(x) = S^i$. So $D(S^i, F_i(x)) = D(S^{i-1}, F_{i-1}(x)) = 0$ and (10) holds. If x is far away from S^i and S^{i-1} (like x_2 in Fig. 2), although $D(S^i, F_i(x)) \neq D(S^{i-1}, F_{i-1}(x))$, they are both large compared to σ_{bump} since they approach the tail of the Gaussian. So (10) also approximately holds since both exponentials vanish to zero. If x is in the intermediate region between x_1 and x_2, as x approaches x_1, $D(S^{i-1}, F_{i-1}(x))$ approaches $D(S^i, F_i(x))$; as x approaches x_2, both exponentials approach zero. This shows why the conjecture is approximately correct.

The experiments in Fig. 3 verify the conjecture in practice. In Fig. 3(a), we take C^{i-1} and perturb it to C^i with
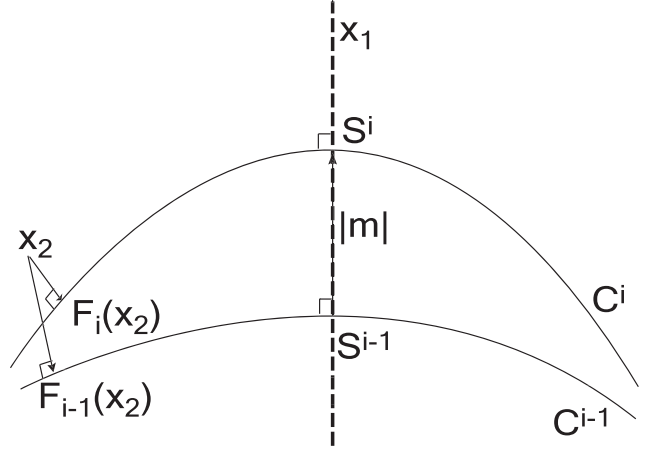

Figure 2. Sketch proof of detailed balance equations

noise magnitude m using source point S^{i-1}. In Fig. 3(b), we perturb C^i to C^{i+1} with magnitude $-m$ using the source point S^i. As Fig. 3(c) shows, C^{i-1} is almost exactly the same as C^{i+1}. This shows equation (10) approximately holds.
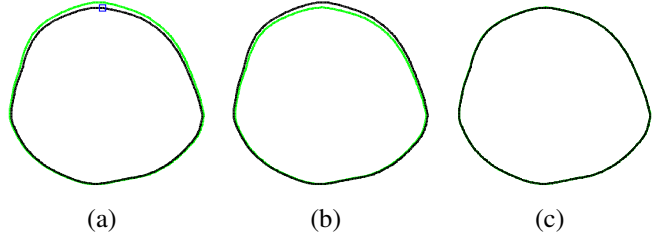

Figure 3. The detailed balance examples. (a) The black curve is C^{i-1}, the green curve is C^i, and the blue square is S^{i-1}. (b) The black curve is C^i, the green curve is C^{i+1}. (c) We overlay C^{i-1} and C^{i+1}; the result shows they are almost the same.

Given the above analysis, the backward proposal can be approximated as:

$$q(\phi^{i-1}|\phi^i) = p(S^i) \cdot p(-m) \qquad (11)$$

If the source point is randomly chosen along the curve under a uniform distribution on the arc length, we can show

$$p(S^{i-1}) = \frac{1}{Length(C^{i-1})} \quad p(S^i) = \frac{1}{Length(C^i)} \qquad (12)$$

The overall proposed MCMC curve sampling algorithm is given in Algorithm 2.

We mention that in computer vision applications, the target distribution $\pi(\phi)$ is sometimes so complex that we only know it up to a constant of proportionality. This is not a problem in the MH algorithm since the acceptance ratio will cancel the normalizing constants.

Algorithm 2 MCMC curve sampling algorithm
___
Initialize ϕ^0
for $i = 1$ to N **do**
    Sample $u \sim U_{[0,1]}$
    Randomly pick S^{i-1} over the zero level set of ϕ^{i-1}
    Sample $m \sim N[0, \sigma_{mag}]$
    Generate the candidate curve $\psi^{(i)}$ as described in Section 4.1
    Calculate

$$a(\psi^i, \phi^{i-1}) = \min\left(1, \frac{\pi(\psi^i)q(\phi^{i-1}|\psi^i)}{\pi(\phi^{i-1})q(\psi^i|\phi^{i-1})}\right)$$

    where $q(\psi^i|\phi^{i-1})$ and $q(\phi^{i-1}|\psi^i)$ are given by (6) and (11).
    if $u < a(\psi^i, \phi^{i-1})$ **then**
        $\phi^i = \psi^i$
    else
        $\phi^i = \phi^{i-1}$
    end if
end for
___

4.3. Implementation

Our implementation is similar to that of the narrowband level set [28], which considerably reduces computational complexity. We maintain re-initialization [28] as an important step in our algorithm, even though Lemma 1 shows that ϕ^i is a valid SDF, for two reasons:

1. Numerical errors, (e.g, $|\nabla \phi|$ is not strictly equal to 1 everywhere, and the foot point calculated in (4) is only approximately accurate as we describe below), will tend to accumulate and re-initialization is necessary.

2. In gradient descent curve evolution, where the curve only moves a short distance each time step, the band that holds the zero level set does not need to be reconstructed very often. However, in our work, since perturbation may lead to large jumps, reconstructing the narrow band is done quite often and re-initialization is a standard technique for this problem.

Due to numerical errors in the computation of $\nabla \phi$, the foot point calculated in (4) may not lie on the curve, and errors may be introduced in the geodesic distance calculation. Since the curve is numerically quantized into grid vertices $x_1, x_2, \ldots, x_N$, we can approximate $F(x) = \arg\min_i ||x_i - x||$, which can be implemented by fast searching algorithms such as Qhull [2].

The choices of σ_{mag} and σ_{bump} depend on the application. A large σ_{bump} will give curves without fine details, and a small σ_{bump} allows for high-frequency variation. For example, in our shape simulation experiment below, we

chose $\sigma_{bump} = 0.05 \cdot length(C^i) \cdot \sqrt{\lceil |m| \rceil}$, which guarantees the curves generated are smooth.

Choosing σ_{mag} is more subtle; if σ_{mag} is too big, large jumps are encouraged, but these jumps are also likely to be rejected and the overall effect is that the state doesn't change much. If σ_{mag} is small, although perturbations are likely to be accepted, it will take a tremendous amount of time to explore the whole state space. Both cases are examples of a poorly mixing chain. Various research in pure MCMC analysis provides some intuitive methods for how to choose this "scaling" [29, 30]. For example, Roberts et al. [29] observed that for RWM, the optimal value should scale with dimension d as $O(d^{-1})$, with an optimal acceptance rate of 0.234. In our experiments, we chose the σ_{mag} that leads to an acceptance rate of 0.25.

5. Experiments

In this section, we present numerical results for our curve sampling algorithm in several different situations. For each case, several hundred or thousand curves are typically generated from the target distribution. In order to abstract useful information out of these large number of samples, we use the techniques introduced by Fan [11] to summarize this information:

1. The sample curves with the highest probability among all the samples generated. These can be viewed as the modes of the target distribution, and also the globally optimal solutions that an optimization method could possibly reach.

2. Histogram images. $H(x)$ is the sample probability that x is inside all the curves generated.

3. Marginal confidence bounds. The level contours where $H(x) = 0.1$, $H(x) = 0.9$.

In our experiments with C++ code on an Intel Core 2.66G CPU, 3.5GB RAM computer, our algorithm runs about 0.3 second for each iteration, which includes the curve generation process and the acceptance ratio evaluation.

5.1. Shape Simulations

As a first experiment, we show the application of curve sampling in shape simulation. Given one training curve or multiple training curves $\tilde{C}_i$, we would like to generate new curves C that are "similar" to the given ones. We define the similarity by a nonparametric shape distribution [8, 20]:

$$\pi(C|\tilde{C}) \propto \frac{1}{N} \sum_{i=1}^{N} \exp\left(\frac{-D^2(C, \tilde{C}_i)}{2\sigma_\phi^2}\right) \quad (13)$$

where the distance measure D is the symmetric area difference. Unlike the approach of Fan [11], we do not require a
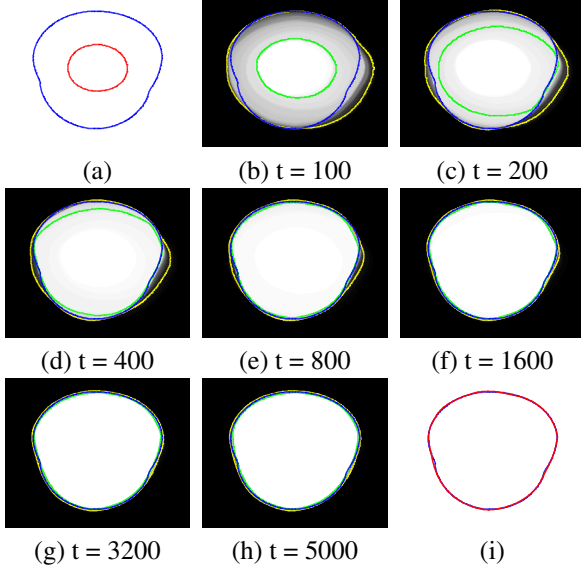
Figure 4. Single target curve simulation results. (a) The initial curve in red and the target curve in blue. (b)–(g) The histogram images at time stamps 100, 200, 400, 800, 1600, 3200, 5000 respectively. The 10% and 90% confidence bounds are shown as yellow and green curves respectively. (i) The red curve is the most likely curve generated among the 5000 samples. This figure is best viewed in color.
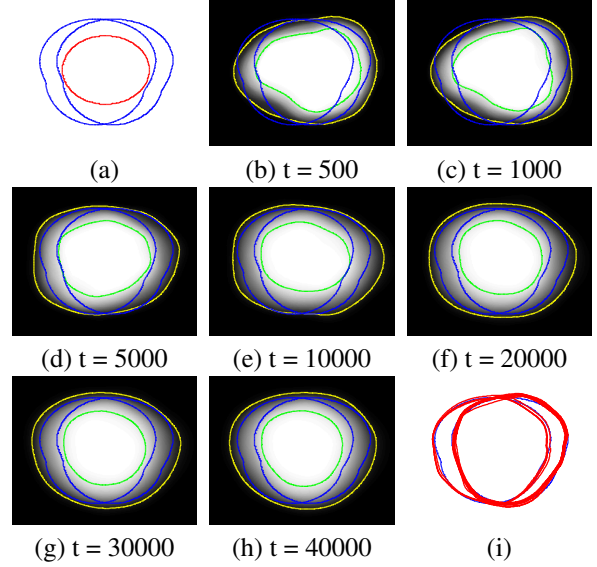

Figure 5. Double target curve simulation results. (a) The initial curve in red and the target curves in blue. (b)–(g) The histogram images at time stamps 500, 1000, 5000, 10000, 20000, 30000, 40000 respectively. The 10% and 90% confidence bounds are shown as yellow and green curves respectively. (i) The red curves are the 20 most likely curves generated among the 40000 samples. This figure is best viewed in color.

separate term in π to enforce smoothness, since the proposal distribution is designed to generate smooth curves.

In Fig. 4, we show the results of using a single curve as the target curve. The target curve and the initial curve are illustrated in Fig. 4(a). In Fig. 4(b)–(g), we show the evolving histogram image with respect to time t. We collect all the samples up to time t and display the confidence bounds of 10/90% with yellow and green curves. The result indicates that the marginal confidence bounds and the histogram image do not change much after $t = 1600$, which shows that the samples converge to a final distribution. The target curve is well bounded by the confidence bounds, which also indicates that the final distribution is the desired distribution. Fig. 4(i) shows the most likely shape among the 5000 generated samples.

In Fig. 5, we show the results of using two curves in (13). In Fig. 5(b)–(g), we show the histogram image at various times t. The mixing time in this case is around $t = 20000$, which is much longer compared with the single target curve experiment. The reason is that since the distribution has two modes, the random walk algorithm will take many steps of very low acceptance probability to make the trip. In Fig. 5(i), we show the 20 most likely curves among the 40000 samples, demonstrating that the stationary distribution is indeed bimodal. In both experiments, we note that increasing σ_ϕ would make the target distributions less sharp or bimodal, and therefore the histogram images

would span a larger region.

5.2. Medical Image Segmentation

As a second experiment, we apply our curve sampling framework to a medical image segmentation application. Direct gradient descent optimization methods are often trapped in regions of local minima. By formulating the cost functional as a probability distribution and drawing curves from this distribution accordingly, we have more informative information regarding the solution's possible locations.

In this experiment, we assume that there are multiple training shapes $\tilde{C}_i$ available, and a reference histogram for the foreground object $\tilde{h}_{in}$ and the background $\tilde{h}_{out}$ respectively. We formulate our cost functional as a probability distribution [9]:

$$p(C|I) \propto p(I|C) \cdot p(C|\tilde{C}_i) \qquad (14)$$

where $p(C|\tilde{C}_i)$ is given by the nonparametric shape distribution (13) and $p(I|C)$ is given by

$$p(I|C) = \exp\left(-\frac{D(h_{in}(C), \tilde{h}_{in})^2}{2\sigma^2}\right) \times$$
$$\exp\left(-\frac{D(h_{out}(C), \tilde{h}_{out})^2}{2\sigma^2}\right) \qquad (15)$$

which drives the object and background histograms to the

reference histograms [6]. D is taken as the Wasserstein cumulative distance measure between histograms [6].

In Fig. 6, we show a segmentation result for the prostate in the CT imagery. This problem is considered to be difficult due to the prostate's indistinguishable intensity distribution compared with surrounding organs such as the bladder. In this example, we used 9 other prostate shapes from the same patient on different days as training data in (13), and used the known reference inside/outside histograms in (15). In Fig. 6(a), we show the 9 training shapes. In Fig. 6(b), we show the initial curve and the ground truth. In Fig. 6(c), we show the confidence bounds after 20000 samples are generated. The ground truth is well bounded by the confidence level contours as expected. In Fig. 6(d), we show the most likely curve among 20000 samples in cyan, which overlays almost exactly with the ground truth.

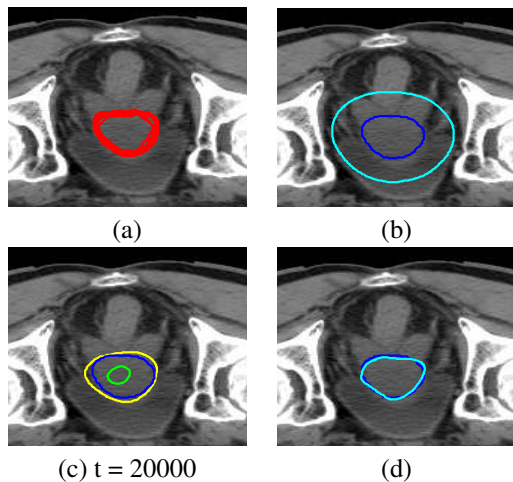(a)   (b)

(c) t = 20000   (d)

Figure 6. Image segmentation results. (a) The training shapes in red. (b) The initial curve in cyan and the ground truth in blue. (c) The histogram image at time stamp 20000. The 10% and 90% confidence bounds are shown as yellow and green curves respectively. (d) The cyan curve is the most likely curve generated among 20000 samples.

6. Conclusions and Future Work

In this paper, we proposed to use the MH algorithm to draw simple closed curves from a distribution defined over the space of signed distance functions. By representing the shapes implicitly, we avoid the problem of complex re-parametrizations typical of explicit shape representations. We illustrated our algorithm with shape simulations and an image segmentation application, and experimentally demonstrated that the Markov chain will asymptotically converge to the desired distribution. We believe our method is an important first step in making MCMC on high-dimensional level set functions viable.

The success or failure of any MCMC algorithm often depends on the choice of the proposal distribution, especially for a complex distribution that has multiple modes. In this work, we illustrated the basic idea of using an RWM-like algorithm for the curve sampling problem. In future work, we plan to investigate other proposal distributions such as Hybrid Monte Carlo (HMC), which incorporates the gradient of the target distribution to improve mixing [24].

To extend our curve sampling algorithm to 3D, we note that while the geodesic distance is easy to compute in 2D, it is more complex in higher dimensions. However, the Fast Marching method can be used to calculate it [32].

Although the level set shape representation is ideal to handle topology change issues, our current framework is restricted to simple closed curves due to the construction of the foot point and its induced perturbation. We plan to investigate whether it is possible to extend our method to different topology situations.

Finally, we are investigating improved segmentation algorithms based on the curve sampling approach. In medical image segmentation problems, the image intensity often has a highly inhomogeneous distribution, and objects to be segmented contain very weak and diffuse boundary edges [11]. Fig. 6 indicates the promise of the idea, but we plan to refine the segmentation cost function in (15) based on a multimodal distribution learned from training data [7].

References

[1] S. Allassonniere, E. Kuhn, A. Trouve, and Y. Amit. Generative model and consistent estimation algorithms for non-rigid deformable models. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 5, 2006.

[2] C. Barber, D. Dobkin, and H. Huhdanpaa. The Quickhull Algorithm for Convex Hulls. *ACM Transactions on Mathematical Software*, 22(4):469–483, 1996.

[3] C. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag, 2006.

[4] D. Breen and R. Whitaker. A level-set approach for the metamorphosis of solid models. *IEEE Transactions on Visualization and Computer Graphics*, 7(2):173–192, 2001.

[5] X. Bresson, P. Vandergheynst, and J. Thiran. A Priori Information in Image Segmentation: Energy Functional Based on Shape Statistical Model and Image Information. In *Proc. International Conference on Image Processing*, 2003.

[6] T. Chan, S. Esedoglu, and K. Ni. Histogram Based Segmentation Using Wasserstein Distances. In *Proc. International Conference Scale Space and Variational Methods in Computer Vision*, 2007.

[7] S. Chen and R. Radke. Level Set Segmentation with Both Shape and Intensity Priors. In *Proc. International Conference on Computer Vision*, 2009.

[8] D. Cremers, S. Osher, and S. Soatto. Kernel Density Estimation and Intrinsic Alignment for Shape Priors in Level Set Segmentation. *International Journal of Computer Vision*, 69(3):335–351, 2006.

[9] D. Cremers, M. Rousson, and R. Deriche. A Review of Statistical Approaches to Level Set Segmentation: Integrating Color, Texture, Motion and Shape. *International Journal of Computer Vision*, 72(2):195–215, 2007.

[10] D. Cremers and S. Soatto. A Pseudo-Distance for Shape Priors in Level Set Segmentation. In *Proc. IEEE Workshop on Variational, Geometric and Level Set Methods in Computer Vision*, 2003.

[11] A. Fan. *Curve Sampling and Geometric Conditional Simulation*. PhD thesis, Massachusetts Institute of Technology, 2007.

[12] A. Fan, J. Fisher III, W. Wells III, J. Levitt, and A. Willsky. MCMC Curve Sampling for Image Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2007.

[13] S. Geman and D. Geman. Stochastic Relaxation, Gibbs Distributions, and The Bayesian Restoration of Images. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 6(6):721–741, 1984.

[14] W. Gilks, S. Richardson, and D. Spiegelhalter. *Markov Chain Monte Carlo In Practice*. Chapman & Hall/CRC, 1996.

[15] P. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732, 1995.

[16] U. Grenander and M. Miller. Representations of knowledge in complex systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 549–603, 1994.

[17] W. Hastings. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57(1):97–109, 1970.

[18] O. Juan, R. Keriven, and G. Postelnicu. Stochastic Motion and the Level Set Method in Computer Vision: Stochastic Active Contours. *International Journal of Computer Vision*, 69(1):7–25, 2006.

[19] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active Contour Models. *International Journal of Computer Vision*, 1(4):321–331, 1988.

[20] J. Kim, M. Çetin, and A. Willsky. Nonparametric Shape Priors for Active Contour-Based Image Segmentation. *Signal Processing*, 87(12):3021–3044, 2007.

[21] Y. Law, H. Lee, and A. Yip. A Multiresolution Stochastic Level Set Method for Mumford-Shah Image Segmentation. *IEEE Transactions on Image Processing*, 17(12):2289–2300, 2008.

[22] M. Leventon, W. Grimson, and O. Faugeras. Statistical Shape Influence in Geodesic Active Contours. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2000.

[23] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, E. Teller, et al. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087, 1953.

[24] R. Neal. Probabilistic inference using Markov chain Monte Carlo methods (Technical Report CRG-TR-93-1). *Department of Computer Science, University of Toronto*, 1993.

[25] S. Osher and R. Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*. Springer, 2003.

[26] S. Osher and N. Paragios. *Geometric level set methods in imaging, vision, and graphics*. Springer-Verlag New York Inc, 2003.

[27] S. Osher and J. Sethian. Fronts propagating with curvature-dependent speed- Algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79:12–49, 1988.

[28] D. Peng, B. Merriman, S. Osher, H. Zhao, and M. Kang. A PDE-Based Fast Local Level Set Method. *Journal of Computational Physics*, 155(2):410–438, 1999.

[29] G. Roberts, A. Gelman, and W. Gilks. Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms. *The Annals of Applied Probability*, 7(1):110–120, 1997.

[30] G. Roberts and J. Rosenthal. Optimal Scaling for Various Metropolis-Hastings Algorithms. *Statistical Science*, 16(4), 2001.

[31] M. Rousson and N. Paragios. Shape Priors for Level Set Representations. In *Proc. European Conference on Computer Vision*, 2002.

[32] J. Sethian. *Level Set Methods and Fast Marching Methods*. Cambridge University Press, 1999.

[33] G. Storvik. A Bayesian approach to dynamic contours through stochastic sampling and simulated annealing. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 16:976–986, 1994.

[34] Z. Tu and S. Zhu. Image Segmentation by Data-Driven Markov Chain Monte Carlo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):657–673, 2002.

[35] H. Zhao, T. Chan, B. Merriman, and S. Osher. A Variational Level Set Approach to Multiphase Motion. *Journal of Computational Physics*, 127(1):179–195, 1996.

[36] H. Zhao, S. Osher, B. Merriman, and M. Kang. Implicit and nonparametric shape reconstruction from unorganized data using a variational level set method. *Computer Vision and Image Understanding*, 80(3):295–314, 2000.

[37] S. Zhu and A. Yuille. Region competition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(9):884–900, 1996.