Efficient MCMC Sampling with Implicit Shape Representations

Jason Chang
Joint work with John W. Fisher III
Massachusetts Institute of Technology
April 20, 2011
Image Segmentation

- Separate the image into separate regions
Implicit Level Set Representation

- Implicitly define a curve on the image with a surface in 3D
Implicit Level Set Representation

- Implicitly specify the curve
- Define a height at every pixel in the image

The Surface \( \varphi \)

The Level Sets / Contours of the Surface
Implicit Level Set Representation

- The zero level set represents the 2D curve
Implicit Level Set Representation

- Signed Distance Function
Sampling Motivation

• Segmentation is often formulated as energy minimization

\[
\arg \min_L E(X, L)
\]

• Exponentiated Mutual Information under some prior is equivalent to posterior:

\[
\exp \left[ -E(X, L) \right] = \exp \left[ I(X; L) - \int_C ds \right] \equiv \pi(\varphi|x)
\]

• Why would we want to sample from posterior of curves \( \pi(\varphi|x) \)?
  – More robust results
  – Multimodal distributions
  – Calculating marginal probabilities
    • Probability that a pixel is on the boundary
    • Probability that a pixel is within a certain region
    • Probability that a pixel is in the same region as another pixel
Sampling Motivation
Metropolis-Hastings Sampling

• The space of segmentations is huge: \( M^{|\Omega|} \)
• Use Metropolis-Hastings MCMC to sample
  – Sample from a proposal distribution

\[
q \left( \hat{\varphi}^{(t+1)} | \varphi^{(t)} \right)
\]

  – Accept the proposal with probability

\[
\min \left( \frac{\pi \left( \hat{\varphi}^{(t+1)} \right)}{\pi \left( \varphi^{(t)} \right)} \cdot \frac{q \left( \varphi^{(t)} | \hat{\varphi}^{(t+1)} \right)}{q \left( \hat{\varphi}^{(t+1)} | \varphi^{(t)} \right)}, 1 \right)
\]

  – Samples will eventually converge if the Markov chain is ergodic because the Hastings ratio ensures detailed balance.
Switches between implicit and explicit representations

Previous Sampling Methods

Preserves signed distance function

Previous Sampling Methods

• [4] alternates between implicit and explicit domain
• [3] generates small, smooth proposal perturbations that maintain the signed distance function

Limitations

– Single simply connected shapes (and no topological changes)
– Only binary segmentation
– Complicated proposals – very slow to sample from and evaluate
– Small proposal perturbations – poor mixing-times
– Unbiased (or curvature biased) proposal perturbations – poor mixing-times
Key Ideas

- Eliminating **signed distance** constraint
  - Proposal easy to sample from
  - Forward-backward ratio simple to evaluate
- Bias proposals with **gradient** of energy functional
  - Increases the posterior-sample ratio and the acceptance ratio
• Assume proposals are generated with some additive perturbation

$$\hat{\phi}(t+1) = \phi(t) + f(X)$$

• A look into the forward-backward ratio

$$\frac{q(\phi(t) | \hat{\phi}(t+1))}{q(\hat{\phi}(t+1) | \phi(t))} = \frac{p_X(-4)}{p_X(4)} \ll 1$$
Biased Proposal Distributions

- Assume proposal is generated from $N$ i.i.d. biased Gaussian RVs
  \[ X_1, X_2, \ldots, X_N \sim \mathcal{N}(\mu, \sigma^2) \]
- How does the distribution of forward-backward ratios look?

\[
Z_N = \log(\text{FBR}) = \log \prod_{i=1}^{N} \frac{p_X(-X_i)}{p_X(X_i)} \sim \mathcal{N} \left( \frac{-2N\mu^2}{\sigma^2}, \frac{4N\mu^2}{\sigma^2} \right)
\]

Biased proposals produce smaller forward-backward ratios!
A Quick Recap

• Ultimate Goal: Increase Hastings ratio
  – Want to bias with gradient to increase the PSR
  – Bias decreases FBR a lot

\[
\frac{\pi(\hat{\varphi}(t+1))}{\pi(\varphi(t))} \cdot \frac{q(\varphi(t)|\hat{\varphi}(t+1))}{q(\hat{\varphi}(t+1)|\varphi(t))}
\]

Posterior-Sample Ratio (PSR)  Forward-Backward Ratio (FBR)
Our Proposal Distribution

\[ h \ast \left( \begin{array}{cc} c^{(t)} & \circ \end{array} n^{(t)} \right) = f^{(t)} \]

\[ \varphi^{(t)} + f^{(t)} = \hat{\varphi}^{(t+1)} \]
Our Proposal Distribution

CSAIL

- Biased proposal tradeoff – increased DLR and decreased FBR
  - Exploit the fact that nearby pixels tend to have same label
- Our proposal

\[
\hat{\varphi}(t+1) = \varphi(t) + f(t)
\]

\[
f(t) = h \ast (c(t) \circ n(t))
\]

LPF allows sparse points to influence PSR a lot
Sparse points only influence the FBR a little
Biased noise tends to increase the PSR

\[
p_{C_i}^{(t)}(1) \propto \exp \left[ -v_i \cdot \text{sign} \left( \varphi_i^{(t)} \right) \right] \quad N_i \sim \mathcal{N} \left( v_i, \sigma^2 \right)
\]

\[h \triangleq \text{LPF with Random Bandwidth}\]

\[v_i \triangleq \text{Gradient Velocity at Pixel } i\]
We show segmentation results in 3 ways:

- **Histogram image** – A count of times pixels are labeled with the same region across all samples

- **Probability of Boundary image** – A normalized count of times pixels are labeled on the edge

- **Segmentation Quantiles** – Thresholding the histogram image to provide confidence bounds (e.g. this pixel belongs to the “inside” region 50% of the time)

- **Best Segmentation** – The sample path with the highest energy. This is a proxy for what the best optimization technique could achieve
Topological Changes

- Other algorithms either catch the inside or outside (depending on initialization), but never both

Comparing Sampling Algorithms

Fan et al. [4]

Chen et al. [3]

Ours

Iteration: 000000
Time: 000000.00

Iteration: 000000
Time: 000000.00

Iteration: 000000
Time: 000000.00
### Computation Time

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Biased</th>
<th>Number of Iterations</th>
<th>Seconds per Iteration</th>
<th>Total Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ours</td>
<td>Yes</td>
<td>150</td>
<td>0.030</td>
<td>x1</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>40,000</td>
<td>0.025</td>
<td>x222</td>
</tr>
<tr>
<td>Chen et al. [3]</td>
<td>Yes</td>
<td>254,000</td>
<td>0.30</td>
<td>x16,933</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>896,000</td>
<td>0.26</td>
<td>x51,769</td>
</tr>
<tr>
<td>Fan et al. [4]</td>
<td>Yes</td>
<td>321,000</td>
<td>5.0</td>
<td>x356,667</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>336,000</td>
<td>5.0</td>
<td>x373,333</td>
</tr>
</tbody>
</table>

![Graph](image)
• Synthetic example with varying SNR
• When images have high SNR (i.e. are very separable), sampling makes less of a difference
M-Ary Sampling

- M-ary segmentation typically achieved with multiple level sets
  - Have to ensure following conditions do not occur
    - Vacuum – pixels are not represented by any region
    - Overlap – pixels are represented by multiple regions
M-ary segmentation typically achieved with multiple level sets
  - Have to ensure following conditions do not occur
    • Vacuum – pixels are not represented by any region
    • Overlap – pixels are represented by multiple regions
  • Use (M) level sets to represent (M+1) regions
    \[
    R_0 = \bigcap_{\ell \in \mathcal{L}} \{ \varphi_\ell < 0 \} \\
    R_\ell = \{ \varphi_\ell \geq 0 \} , \quad \forall \ell \in \mathcal{L} = \{1, 2, \ldots, M\}
    \]
  • Vacuum impossible by construction
Choose a random level set, $\ell$

- Pixels belong in 3 categories:
  1. Belongs to $R_\ell$ and has non-negative height only in $\varphi_\ell$
  2. Belongs to $R_0$ and has negative height in all level sets
  3. Belongs to $R_\ell$ and has non-negative height only in $\varphi_l$ ($l \neq \ell$)
- Only allow moves between pixels of type (1) and (2)
- M-Ary proposal:

$$\hat{\varphi}_{\ell}^{(t+1)} = \varphi_{\ell}^{(t)} + f_{\ell}^{(t)}$$

$$f_{\ell}^{(t)} = (h_\ell \ast (c_{\ell}^{(t)} \circ n_{\ell}^{(t)})) \circ 1 \{R_\ell \cup R_0\}$$
M-Ary Sampling

• For a pixel to move from $R_\ell$ to $R_l$ it must go through $R_0$
• This must be reflected in our bias
  \[ \mathbf{v}(\ell, l) \triangleq \text{Gradient velocity between } \varphi_\ell \text{ and } \varphi_l \]

• Proposal only looks at $\mathbf{v}(\ell, 0)$
• Instead of biasing with gradient, bias with minimal gradient
  \[ m_i(\ell) \triangleq \min_{l \in \{0, 1, 2, \ldots, M\}} \min_{l \neq \ell} \mathbf{v}_i(\ell, l) \]

• When using mutual information, the minimal gradient is
  \[ m_i(\ell) = \log \frac{p_\ell^X(x_i)}{p_X^{\max}(x_i)} \quad p_X^{\max}(i) = \max_{l \in \{0, 1, 2, \ldots, M\}} \min_{l \neq \ell} p_\ell^X(x_i) \]
The green line in the plot shows the energy for the sample path that produces the optimal energy after the chain has converged. Clearly, not all samples reach this extrema; however, the marginal statistics of these samples provide a much richer characterization of the probabilistic space of shapes.
Example Sampling vs. Optimization
Results on the BSDS

Results from the Berkeley Segmentation Dataset. (‘X’ on the Precision-Recall curve correspond to the probability of boundary image. ‘+’ on the curve corresponds to the best sample path)
Results on the BSDS

Contributions

- Effortlessly allow for **topological** changes
- Extension to **M-ary** sampling
- Improves convergence **speed** by orders of magnitude
- Demonstrate **versatility** of sampling methods for segmentation