MCMC Sampling with Implicit Shape Representations

CSAIL

MIT COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE LABORATORY

Jason Chang

Joint work with John W. Fisher III

Massachusetts Institute of Technology

April 21, 2010

Sampling Motivation

- Segmentation is usually formulated with energy maximization
- Exponentiated Mutual Information under some prior is equivalent to the posterior:

$$\exp\left[I\left(X;L\right) - \oint_{\mathcal{C}} ds\right] \equiv \pi\left(\varphi|x\right)$$

- Why would we want to sample from posterior of curves $\pi\left(\varphi|x\right)$?
  - More robust results
  - 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
    - Etc.

Metropolis-Hastings Sampling

- The space of segmentations is infinite
- 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.

Previous Sampling Methods

- [1] pioneered the sampling of segmentation space but required expensive alternating between an implicit and explicit representation
- [2] extended the work by staying in the implicit domain and perturbing the level set function such that the signed distance function was preserved
- Both [1] and [2] are limited to a single simply connected shape, and do not allow for topological changes

[1] Fan, A.C., J. W. Fisher III, W. M. Wells, J. J. Levitt, A. S. Willsky. MCMC curve sampling for image segmentation. In: MICCAI (2007).

[2] Chen, S., R. J. Radke. Markov chain monte carlo shape sampling using level sets. Second Workshop on non-Rigid Shape Analysis and Deformable Image Alignment, in conjunction with ICCV 2009.

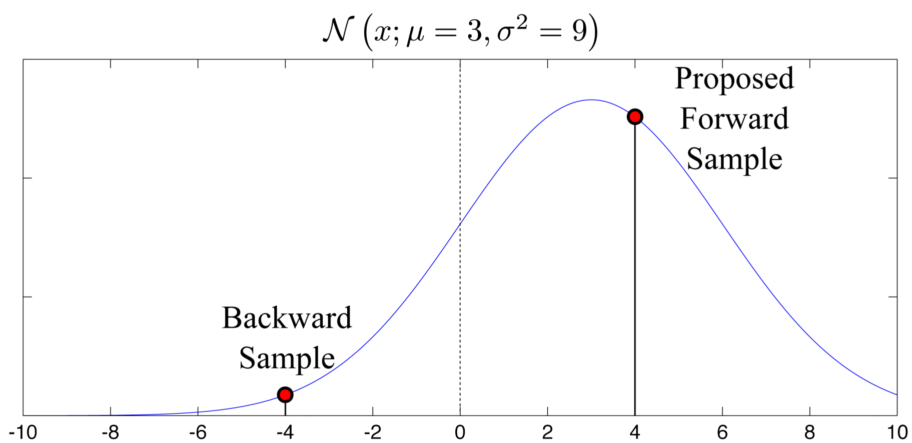
Good Proposal Distributions

- Probability of accepting a sample is increased if posterior increases

$$\frac{\pi\left(\hat{\varphi}^{(t+1)}\right)}{\pi\left(\varphi^{(t)}\right)}$$

- Tradeoff with biased proposals
  - The proposal can be biased to increase the posterior
  - Decreases the forward-backward ratio in the acceptance ratio

$$\frac{q\left(\varphi^{(t)} | \hat{\varphi}^{(t+1)}\right)}{q\left(\hat{\varphi}^{(t+1)} | \varphi^{(t)}\right)}$$



Our Proposal Distribution

- **Goal**: design a proposal distribution that biases samples towards likely configurations (increasing the likelihood ratio) without decreasing the forward-backward ratio much.
- **Solution**: a small number of biased random variables move many pixels into a more probable configuration
- Proposal perturbation

$$\hat{\varphi}^{(t+1)} = \varphi^{(t)} + \underline{f}^{(t)}$$

$$\underline{f}^{(t)} = \underline{h} * (\underline{c} \circ \underline{n})$$

$$\Pr[C_i = 1] \propto \exp\left[v_i \cdot \operatorname{sign}\left(-\varphi_i^{(t)}\right)\right] \qquad N_i \sim \mathcal{N}\left(v_i, \sigma^2\right)$$

$$\underline{h} \triangleq \text{LPF with Random Bandwidth}$$

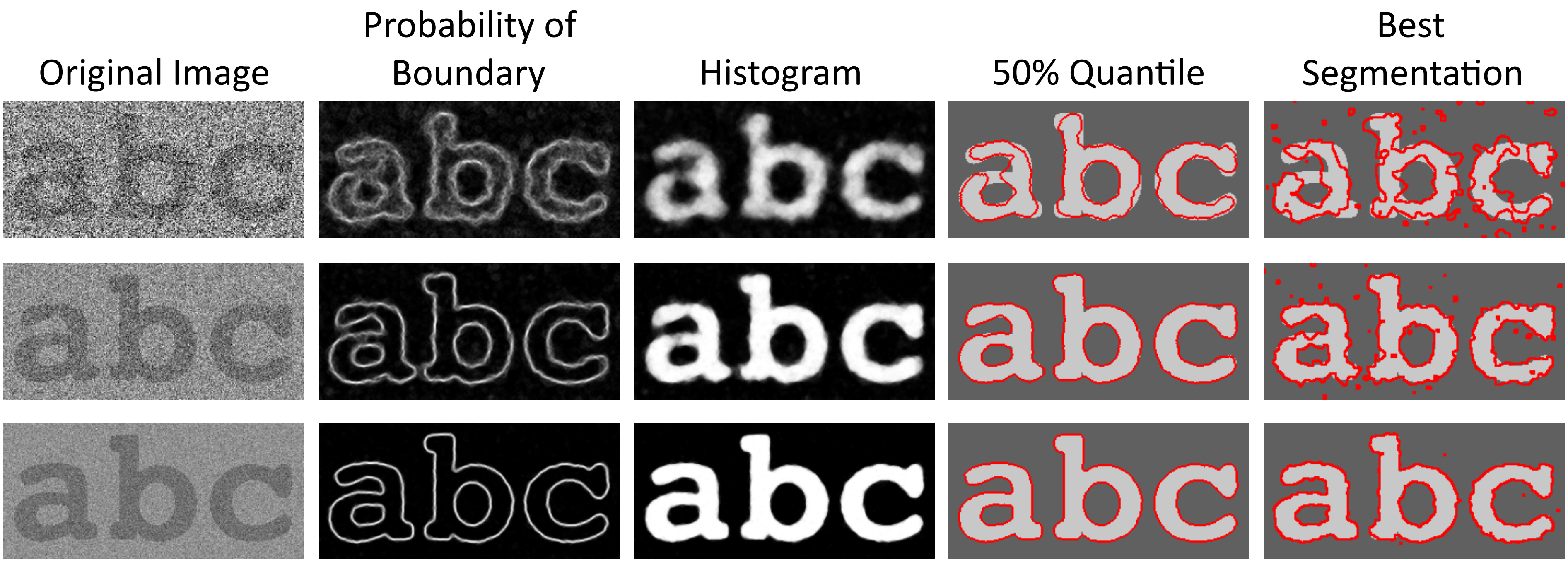
$$v_i \triangleq \text{Gradient Descent Velocity of Pixel } i$$

Results

- 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

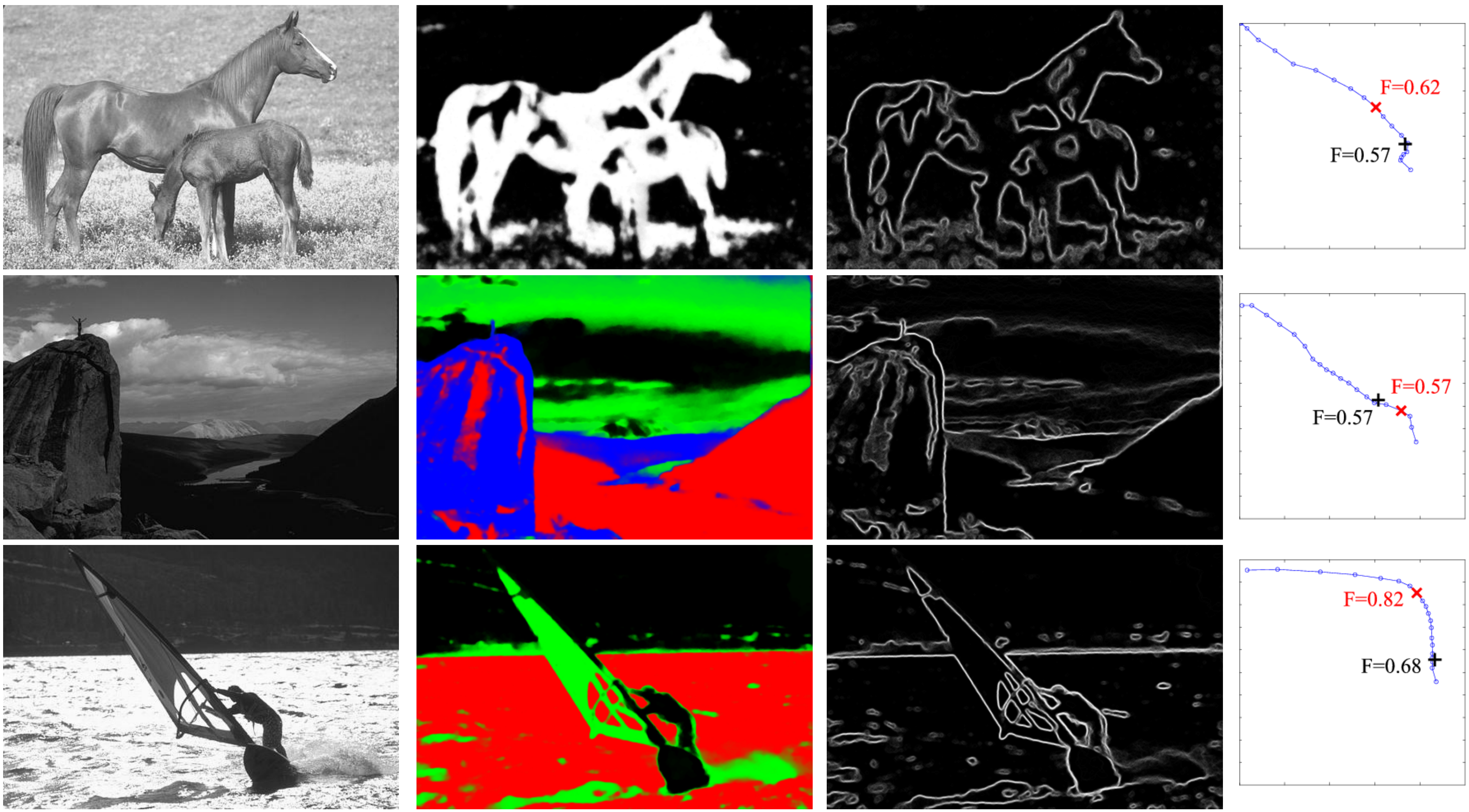
Synthetic Results

- Synthetic example with varying SNR
- When images have high SNR (i.e. are vary separable), sampling makes less of a difference

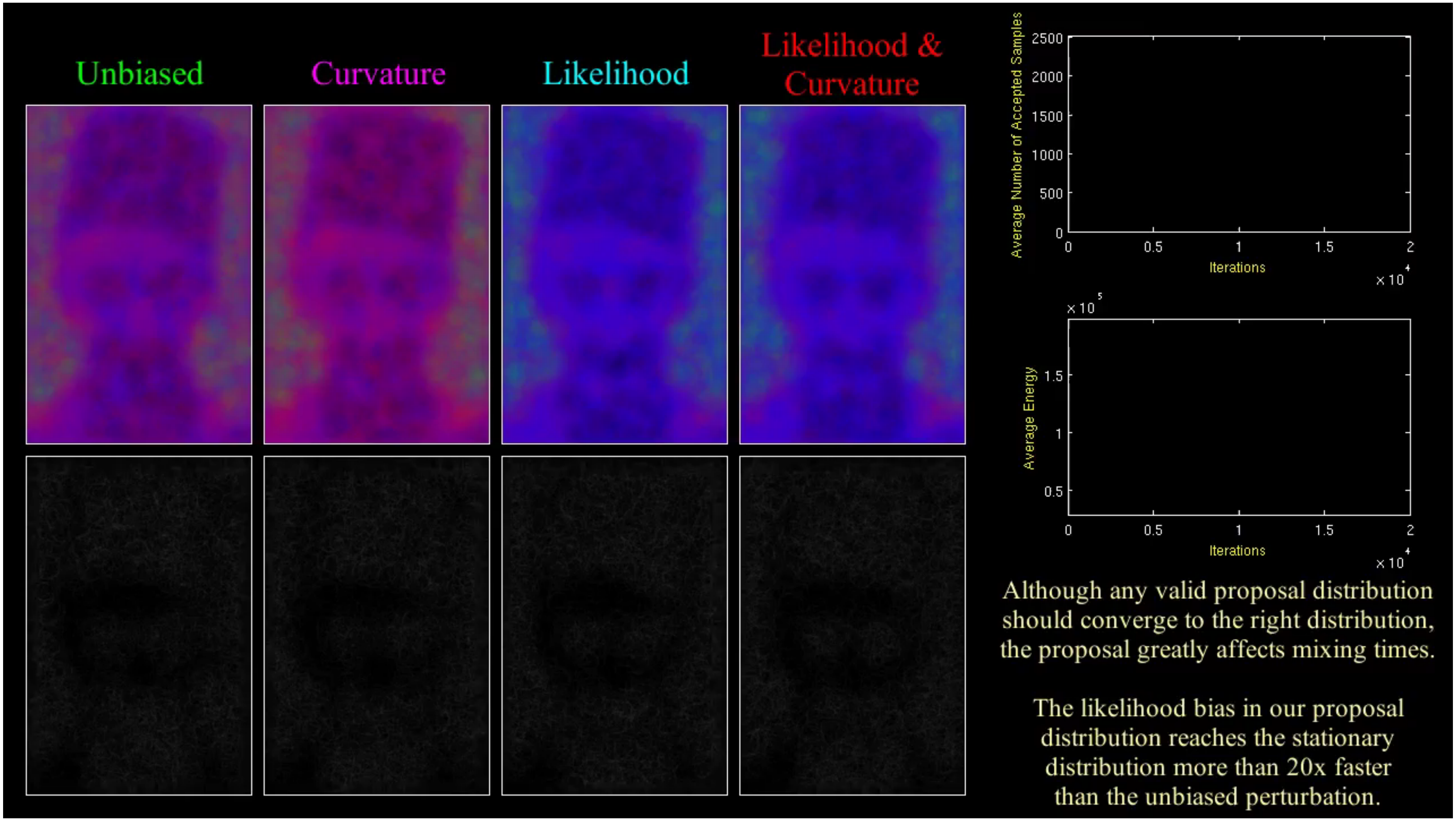


Natural Image Results

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)



Mixing Rate Comparison

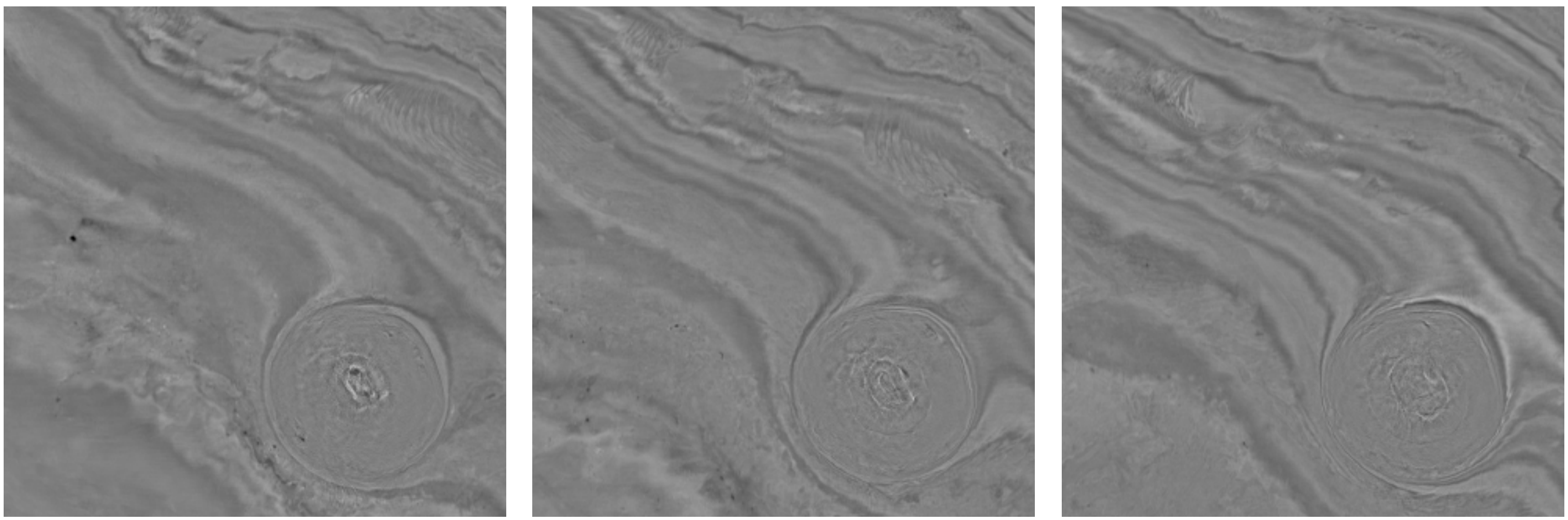


Although any valid proposal distribution should converge to the right distribution, the proposal greatly affects mixing times.

The likelihood bias in our proposal distribution reaches the stationary distribution more than 20x faster than the unbiased perturbation.

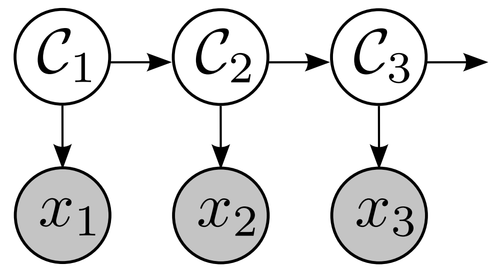
Dynamics of Curves

- We want to incorporate the 3D structure of the salt dome into the sampling of successive frames



Dynamics of Curves

- We could use Symmetric Area Difference (SAD) as a prior on future frames



$$\mathrm{SAD}\left(\mathcal{C}_1, \mathcal{C}_2\right) = \sum_{i \in R_2^+} \left[\mathbb{1}\left\{\ell_{1,i} = -\right\}\right] + \sum_{i \in R_2^-} \left[\mathbb{1}\left\{\ell_{1,i} = +\right\}\right]$$

$$p\left(\varphi_2 | \varphi_1\right) \propto \exp\left[-\mathrm{SAD}\left(\mathcal{C}_1, \mathcal{C}_2\right)\right]$$

$$E = I\left(X; L\right) - \oint_{\mathcal{C}} ds + \frac{1}{N} \sum_{j=1}^{N} \exp\left[-\mathrm{SAD}\left(\mathcal{C}, \mathcal{C}_j\right)\right]$$

* We use the notation $\mathcal{C}$ and φ interchangeably because a level set function defines a curve
$\ell_{j,i}$ corresponds to the label assigned to pixel i in image j

Sampling Dynamics on Curves

- Hastings Ratio requires calculating the energy at every sample

$$E = I\left(X; L\right) - \oint_{\mathcal{C}} ds + \underbrace{\frac{1}{N}\sum_{j=1}^{N} \exp\left[-\mathrm{SAD}\left(\mathcal{C}, \mathcal{C}_j\right)\right]}_{E_{SAD}}$$

- We typically have over 1000 previous samples. Thus, calculating the SAD term can become computationally intensive

$$E_{SAD} = \underbrace{\frac{1}{N}\sum_{j=1}^{N} \exp \underbrace{\left[-\sum_{i\in R^+} \mathbb{I}\left\{\ell_{j,i} = -\right\} - \sum_{i \in R^-} \mathbb{I}\left\{\ell_{j,i} = +\right\}\right]}_{\text{Loop over current sample}}}_{\text{Loop over } \underline{\textbf{all}} \text{ previous samples}}$$

Sampling Dynamics on Curve

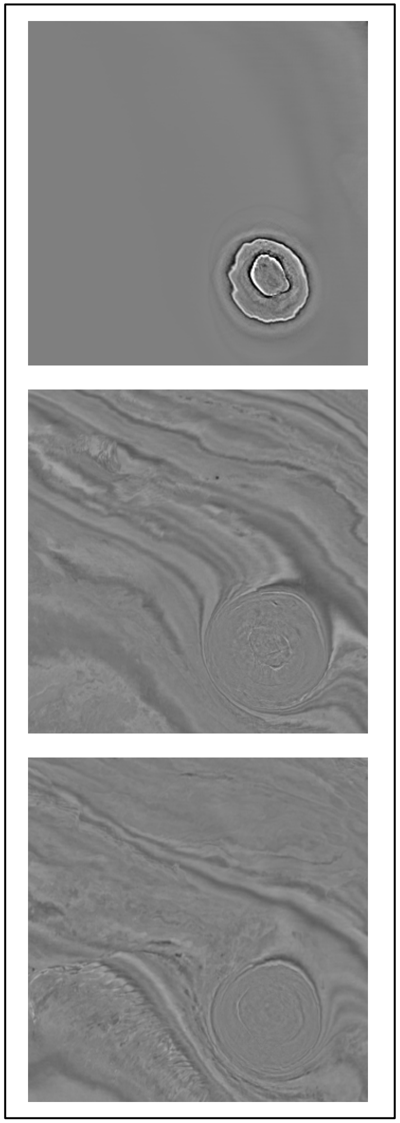
- We approximate this energy prior with a Mixture of SADs (MSAD)

$$E = I(X;L) - \oint_{\mathcal{C}} ds + \underbrace{\frac{1}{N}\sum_{j=1}^{N} -\mathrm{SAD}(\mathcal{C}, \mathcal{C}_j)}_{E_{MSAD}}$$

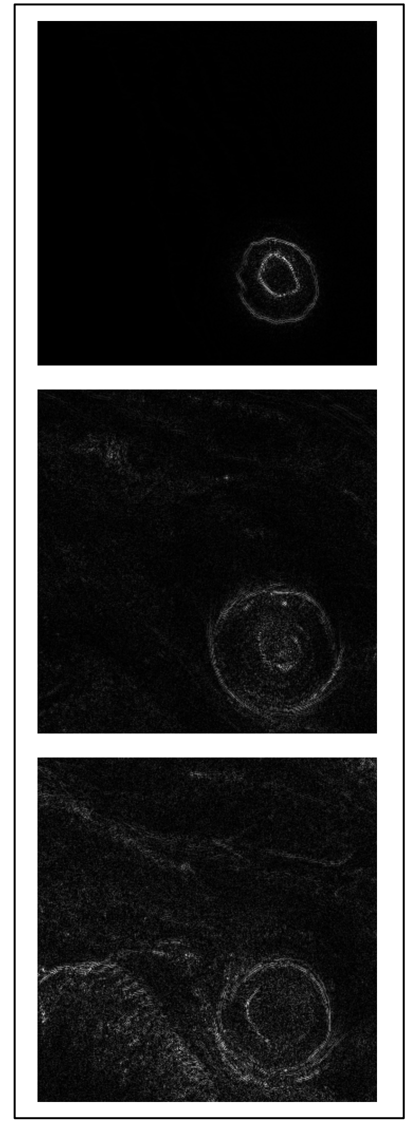
- MSAD can be computed more efficiently with a precomputed histogram image, $\underline{h}$

$$\begin{aligned} E_{MSAD} &= \frac{1}{N}\sum_{j=1}^{N}\left[-\sum_{i\in R^+} \mathbb{1}\{\ell_{j,i} = -\} - \sum_{i\in R^-} \mathbb{1}\{\ell_{j,i} = +\}\right] \\ &= -\frac{1}{N}\left[\sum_{i\in R^+}\sum_{j=1}^{N} \mathbb{1}\{\ell_{j,i} = -\} + \sum_{i\in R^-}\sum_{j=1}^{N} \mathbb{1}\{\ell_{j,i} = +\}\right] \\ &= -\frac{1}{N}\left[\sum_{i\in R^+}(N - h_i) + \sum_{i\in R^-} h_i\right] \end{aligned}$$

Simple Feature



$$* \begin{array}{|c|c|c|}\hline & -1/4 & \\ \hline -1/4 & 1 & -1/4 \\ \hline & -1/4 & \\ \hline \end{array} \rightarrow \boxed{|\cdot|} =$$



Comparison without Dynamics