Efficient MCMC Sampling with Implicit Shape Representations

CSAIL
MIT COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE LABORATORY

Jason Chang

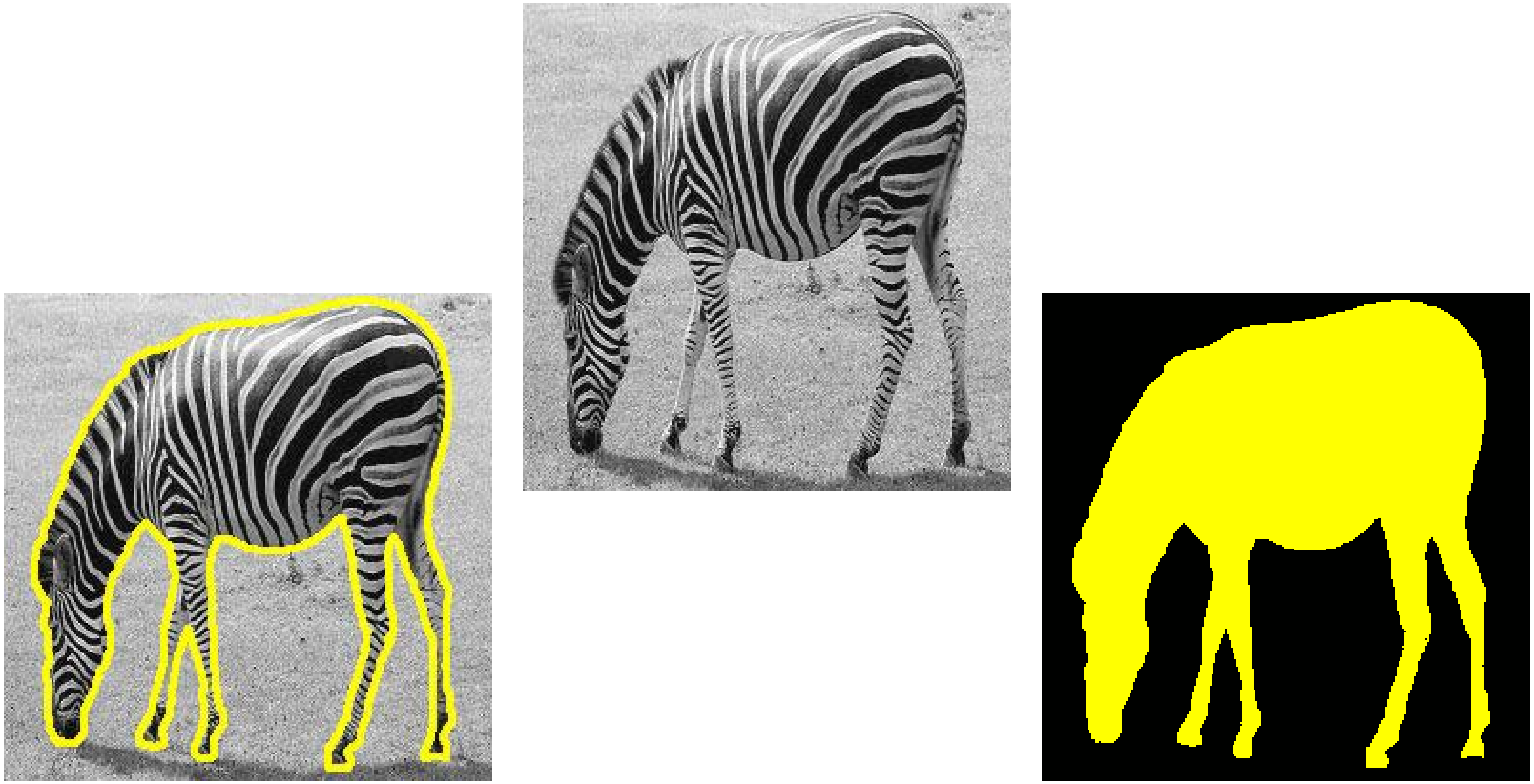
Joint work with John W. Fisher III

Massachusetts Institute of Technology

June 6, 2011

Image Segmentation

- Separate the image into separate regions

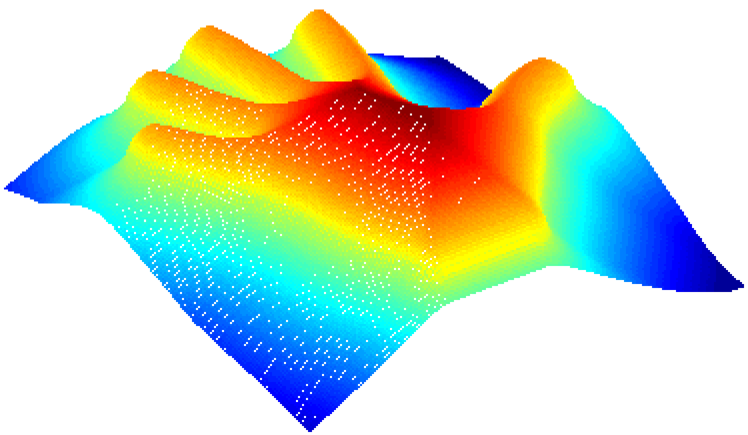


Implicit Level Set Representation

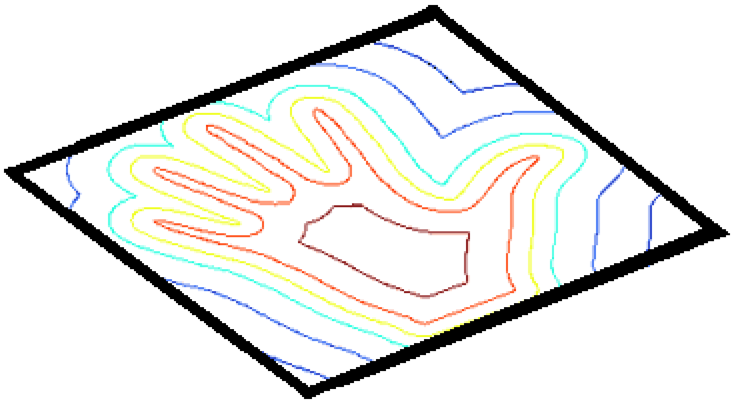
- Implicitly define a curve on the image with a surface in 3D

Implicit Level Set Representation

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



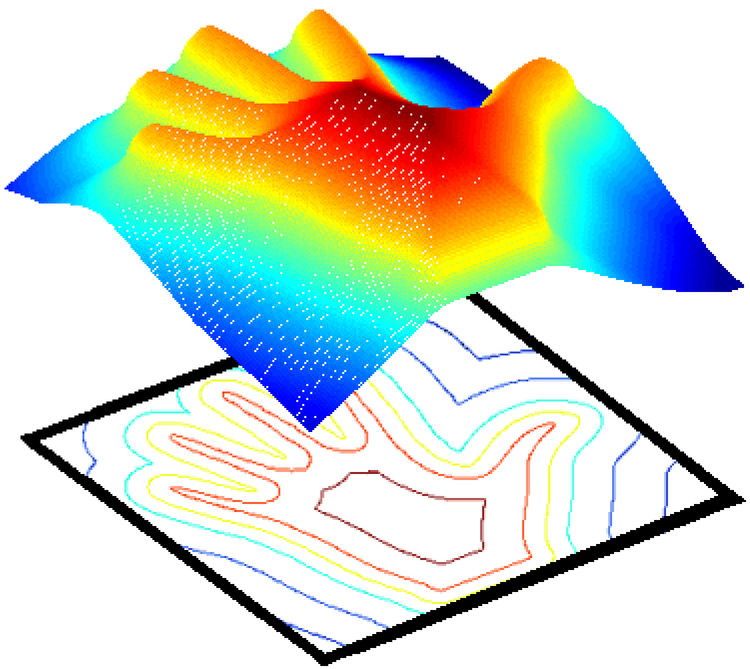
The Surface φ

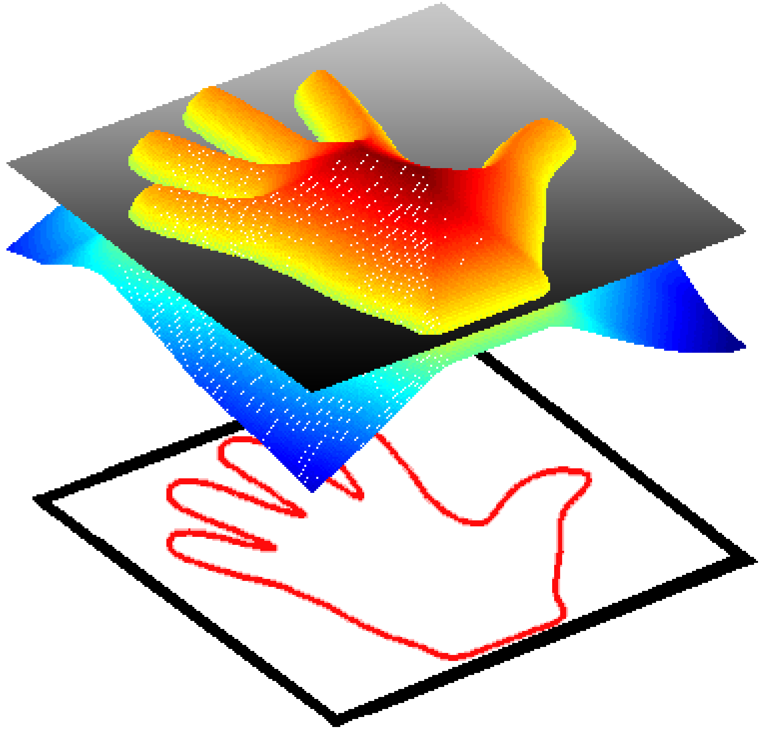


The Level Sets / Contours of the Surface

Implicit Level Set Representation

- The zero level set represents the 2D curve

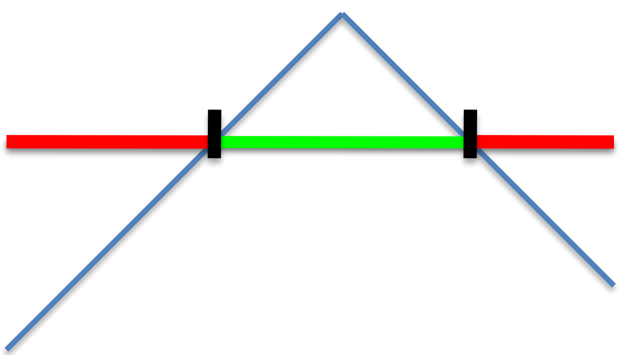




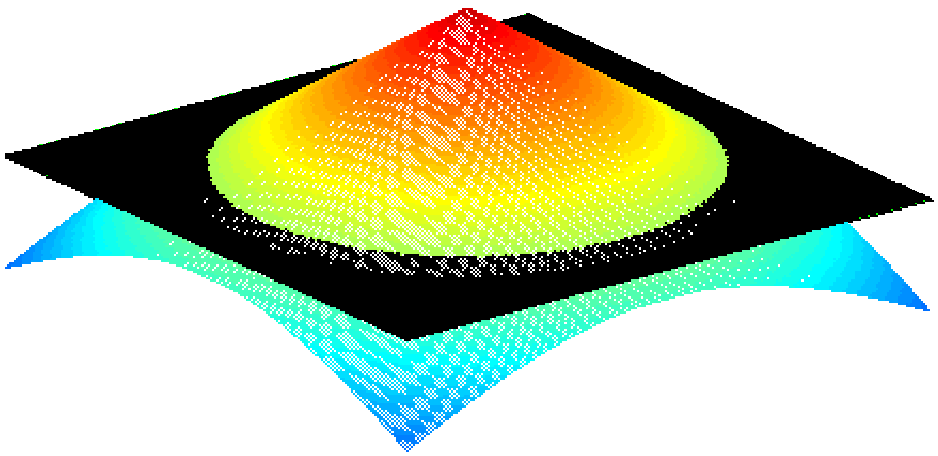
Implicit Level Set Representation

- Signed Distance Function

1D



2D



Sampling Motivation

- Segmentation is often formulated as energy minimization

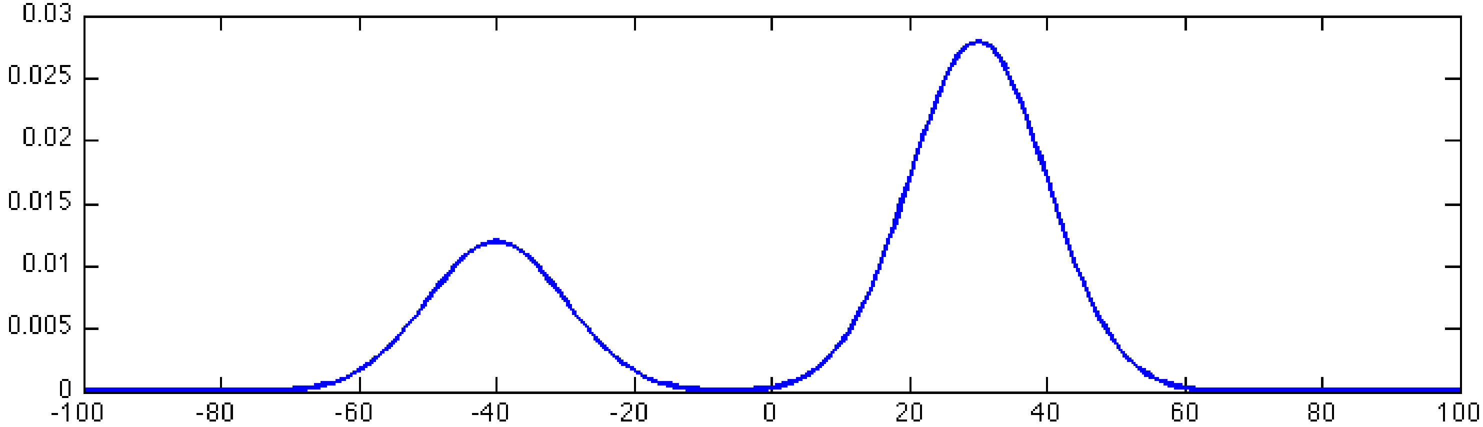
$$\arg\min_{L} E(X,L)$$

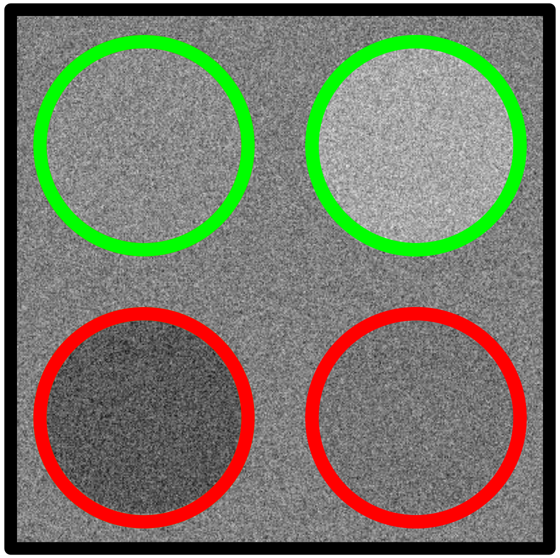
- Exponentiated Mutual Information under some prior is equivalent to posterior:

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

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

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

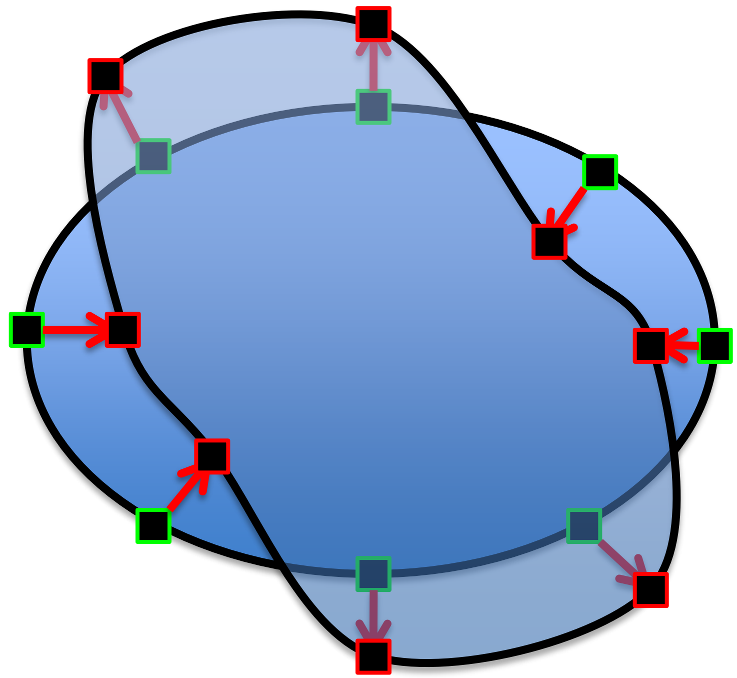
  - Accept the proposal with probability

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

  - Samples will eventually converge if the Markov chain is ergodic because the Hastings ratio ensures detailed balance.

Previous Sampling Methods

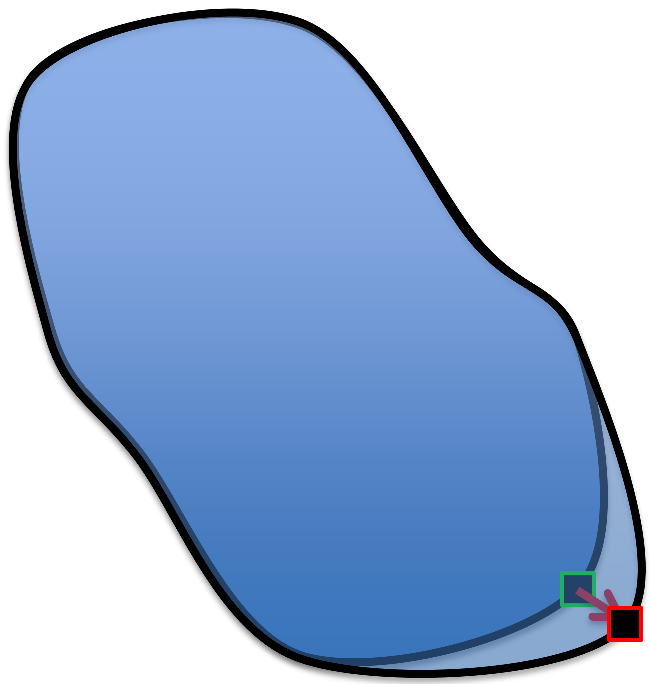
Switches between implicit and explicit representations



[4] 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).

Previous Sampling Methods

Preserves signed distance function



[3] 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.

Previous Sampling Methods

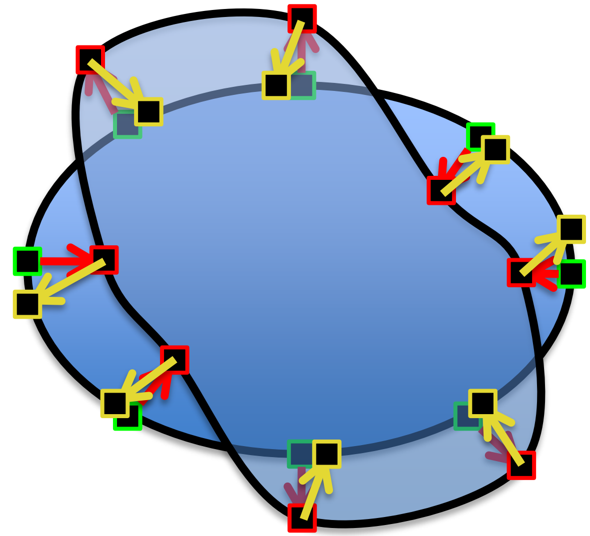
- [4] alternates between implicit and explicit domain
- [3] generates small, smooth proposal perturbations that maintain the signed distance function
- Limitations
  - Single **simply connected** shapes (and no topological changes)
  - Only **binary** segmentation
  - **Complicated** proposals – very slow to sample from and evaluate
  - **Small** proposal perturbations – poor mixing-times
  - **Unbiased** (or curvature biased) proposal perturbations – poor mixing-times

Key Ideas

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

Posterior-Sample Ratio (PSR)

Forward-Backward Ratio (FBR)



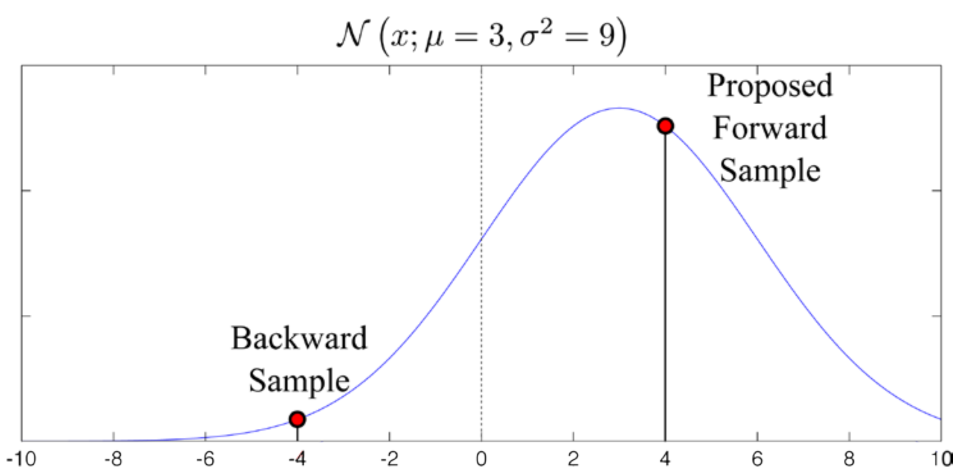
- Eliminating **signed distance** constraint
  - Proposal easy to sample from
  - Forward-backward ratio simple to evaluate
- Bias proposals with **gradient** of energy functional
  - Increases the posterior-sample ratio and the acceptance ratio

Biased Proposal Distributions

- Assume proposals are generated with some additive perturbation

$$\hat{\varphi}^{(t+1)} = \varphi^{(t)} + f\left(X\right)$$

- A look into the forward-backward ratio



$$\frac{q\left(\varphi^{(t)} | \hat{\varphi}^{(t+1)}\right)}{q\left(\hat{\varphi}^{(t+1)} | \varphi^{(t)}\right)} = \frac{p_X(-4)}{p_X(4)} \ll 1$$

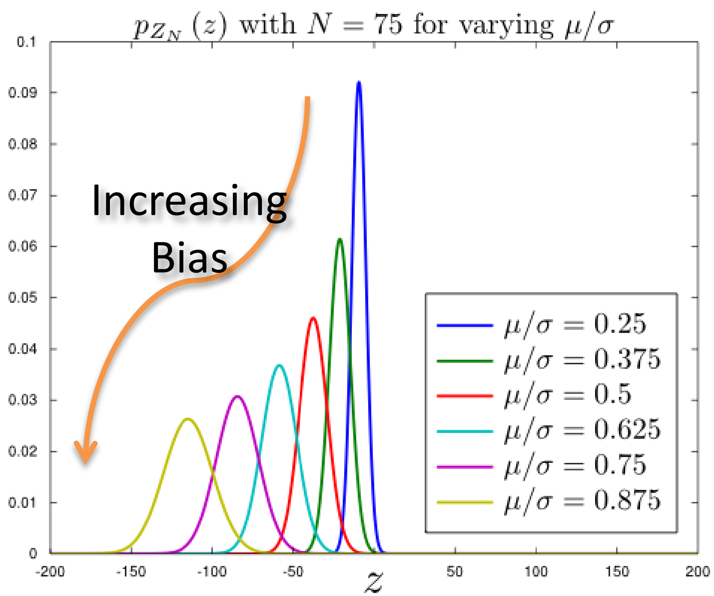
Biased Proposal Distributions

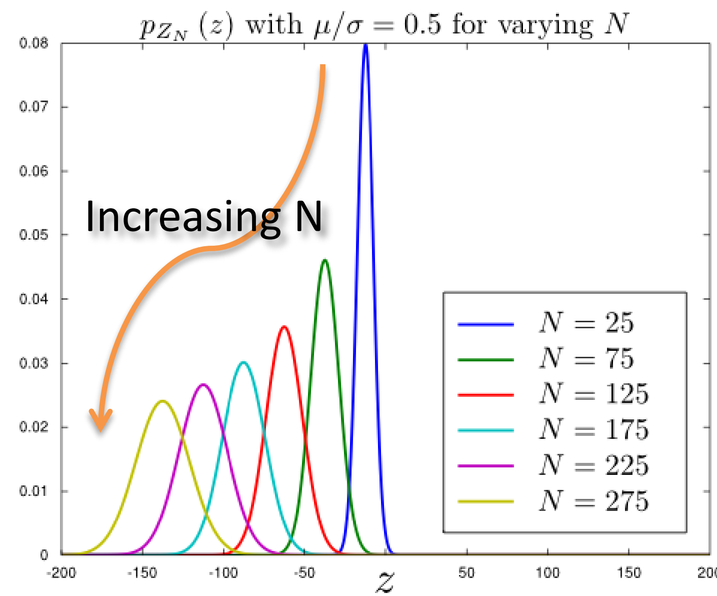
- Assume proposal is generated from N i.i.d. biased Gaussian RVs

$$X_1, X_2, ..., X_N \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

- 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!

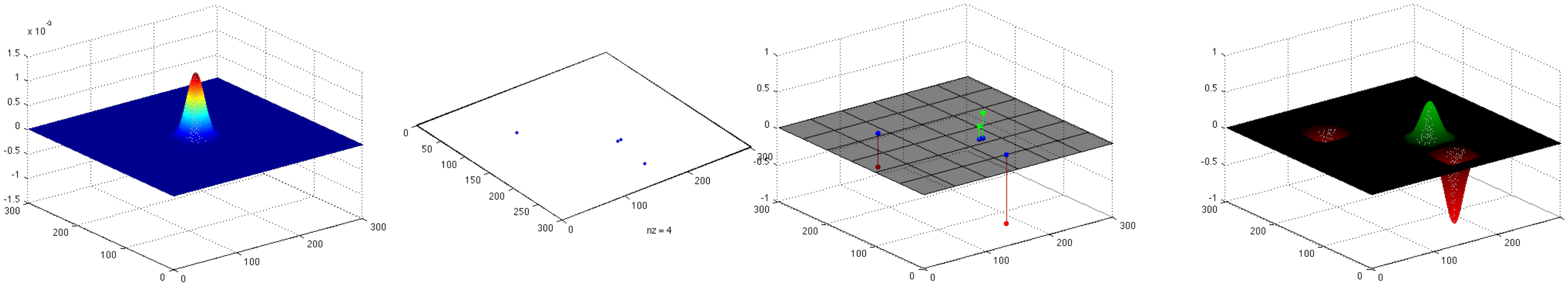
A Quick Recap

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

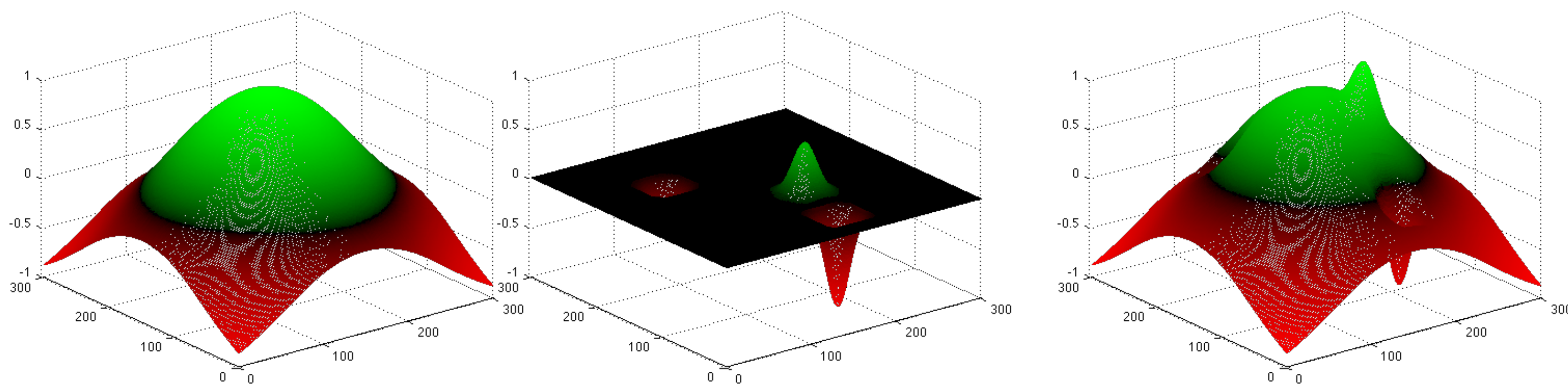
$$\underbrace{\frac{\pi\left(\hat{\varphi}^{(t+1)}\right)}{\pi\left(\varphi^{(t)}\right)}}_{\text{Posterior-Sample Ratio (PSR)}} \cdot \underbrace{\frac{q\left(\varphi^{(t)} \middle| \hat{\varphi}^{(t+1)}\right)}{q\left(\hat{\varphi}^{(t+1)} \middle| \varphi^{(t)}\right)}}_{\text{Forward-Backward Ratio (FBR)}}$$

CSAIL

Our Proposal Distribution



$$\mathbf{h} \quad * \quad \left(\quad \mathbf{c}^{(t)} \quad \circ \quad \mathbf{n}^{(t)} \quad \right) \quad = \quad \mathbf{f}^{(t)}$$



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

Our Proposal Distribution

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

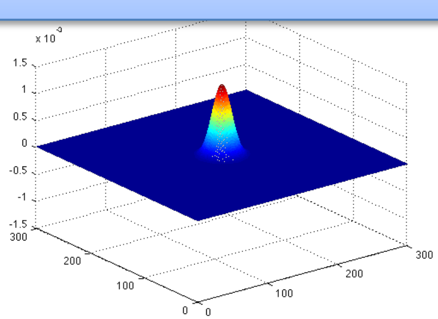
$$\hat{\varphi}^{(t+1)} = \varphi^{(t)} + \mathbf{f}^{(t)}$$

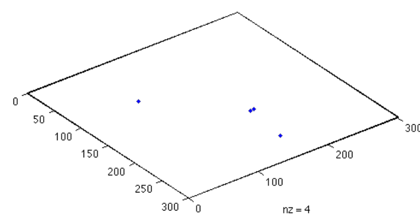
$$\mathbf{f}^{(t)} = \mathbf{h} \;*\; \left(\mathbf{c}^{(t)} \;\circ\; \mathbf{n}^{(t)}\right)$$

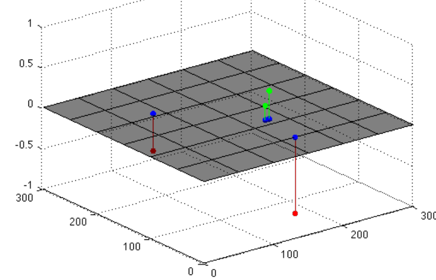
LPF allows sparse points to influence PSR a **lot**

Sparse points only influence the FBR a **little**

Biased noise tends to **increase** the PSR







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

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

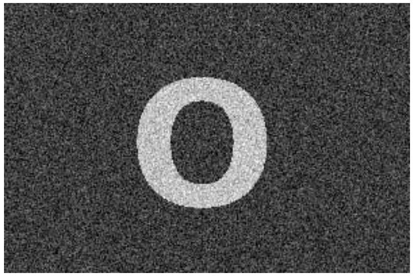
$$v_i \triangleq \text{Gradient Velocity at Pixel } i$$

Results

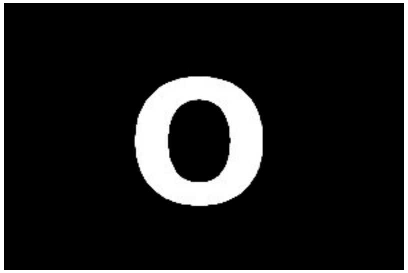
- We show segmentation results in 3 ways:
  - **Histogram image** – A count of times pixels are labeled with the same region across all samples
  - **Probability of Boundary image** – A normalized count of times pixels are labeled on the edge
  - **Segmentation Quantiles** – Thresholding the histogram 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

Topological Changes

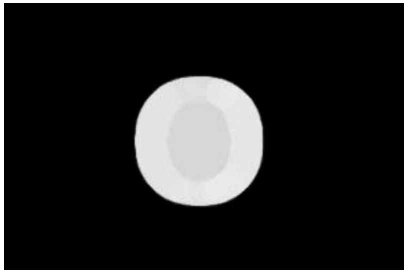
- Other algorithms either catch the inside or outside (depending on initialization), but never both



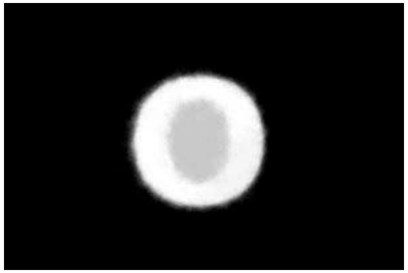
Original



Ours



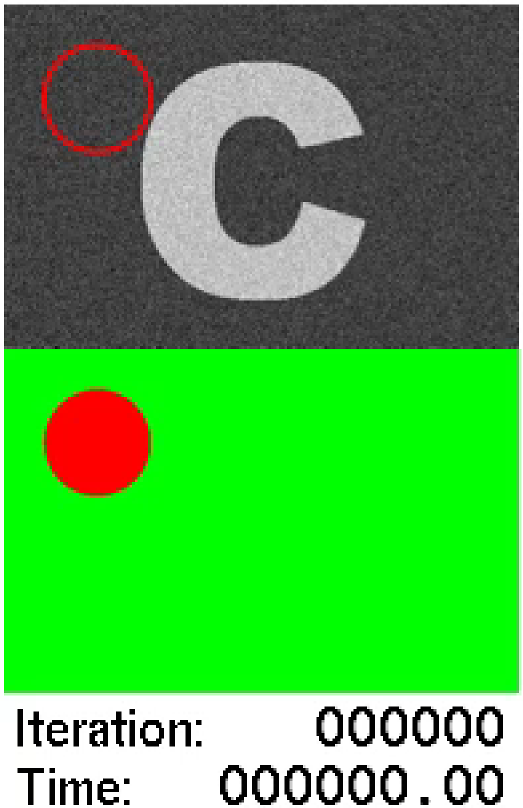
Chen et al. [3]



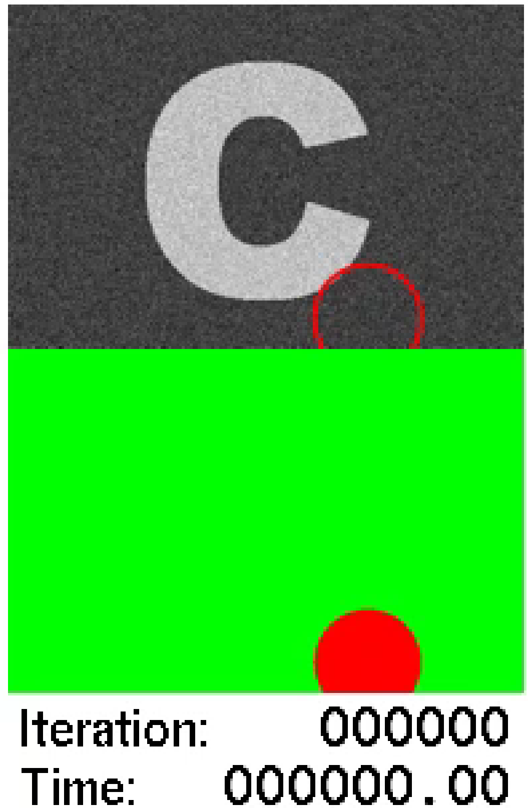
Fan et al. [4]

Comparing Sampling Algorithms

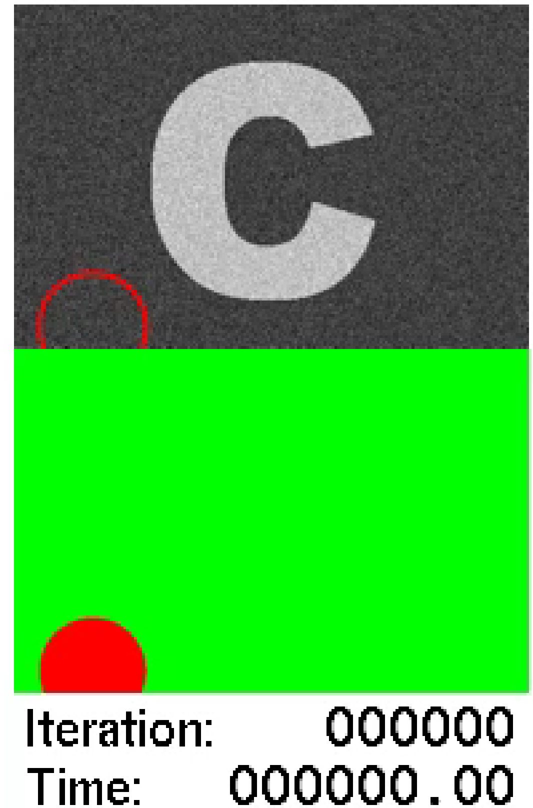
Fan et al. [4]



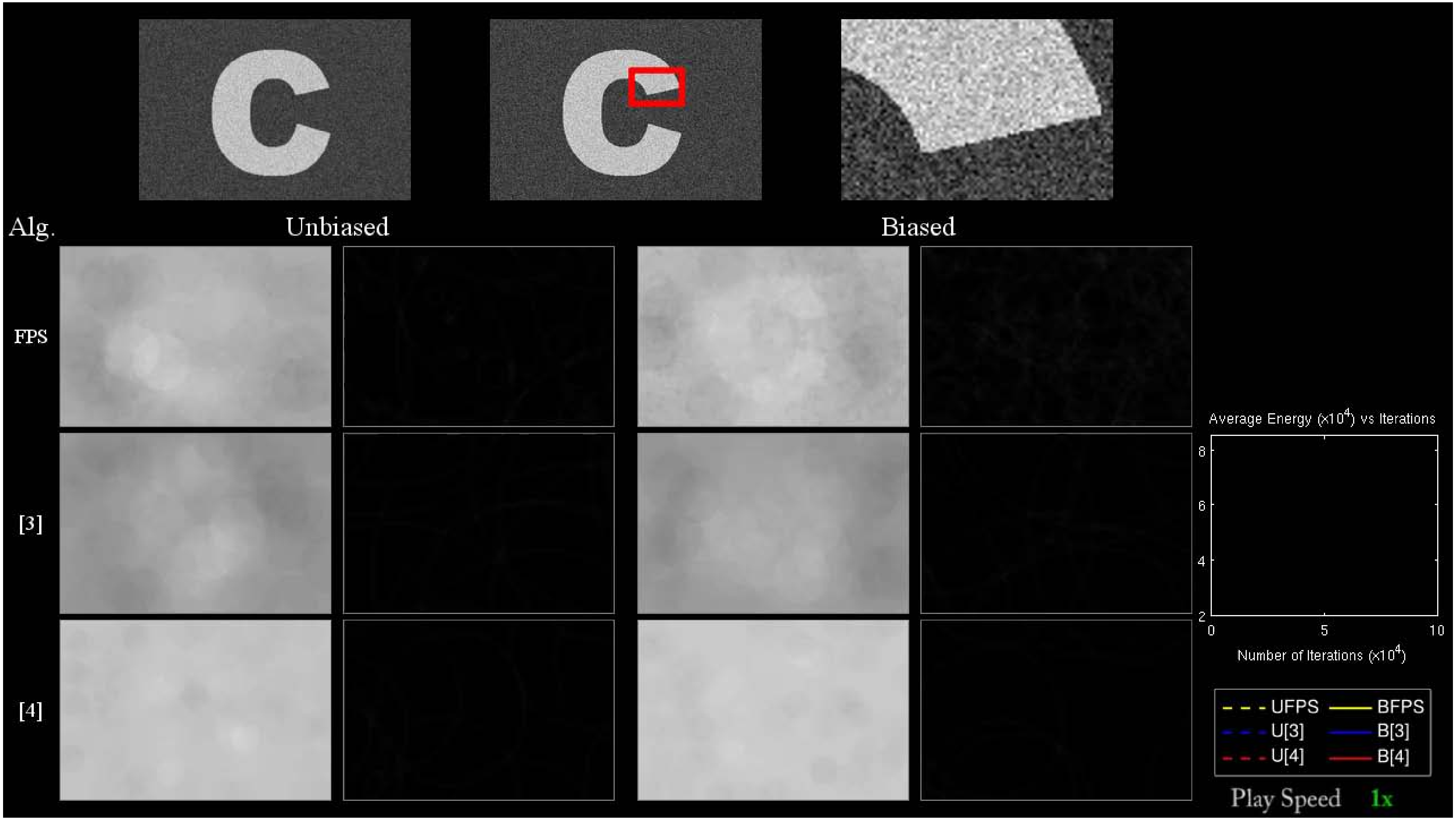
Chen et al. [3]



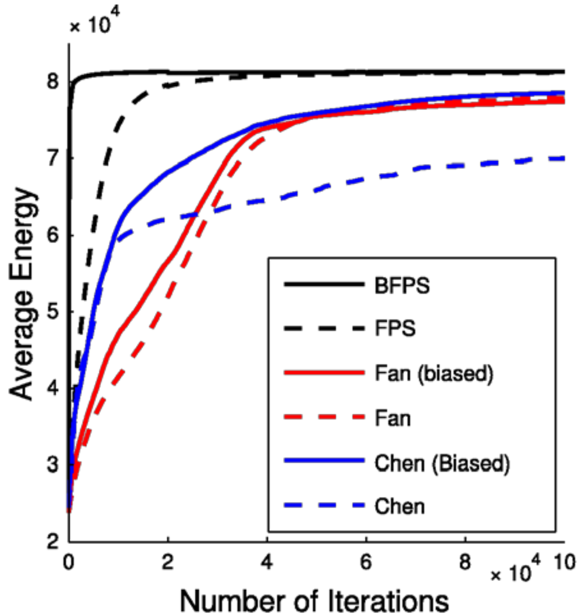
Ours



Computation Time



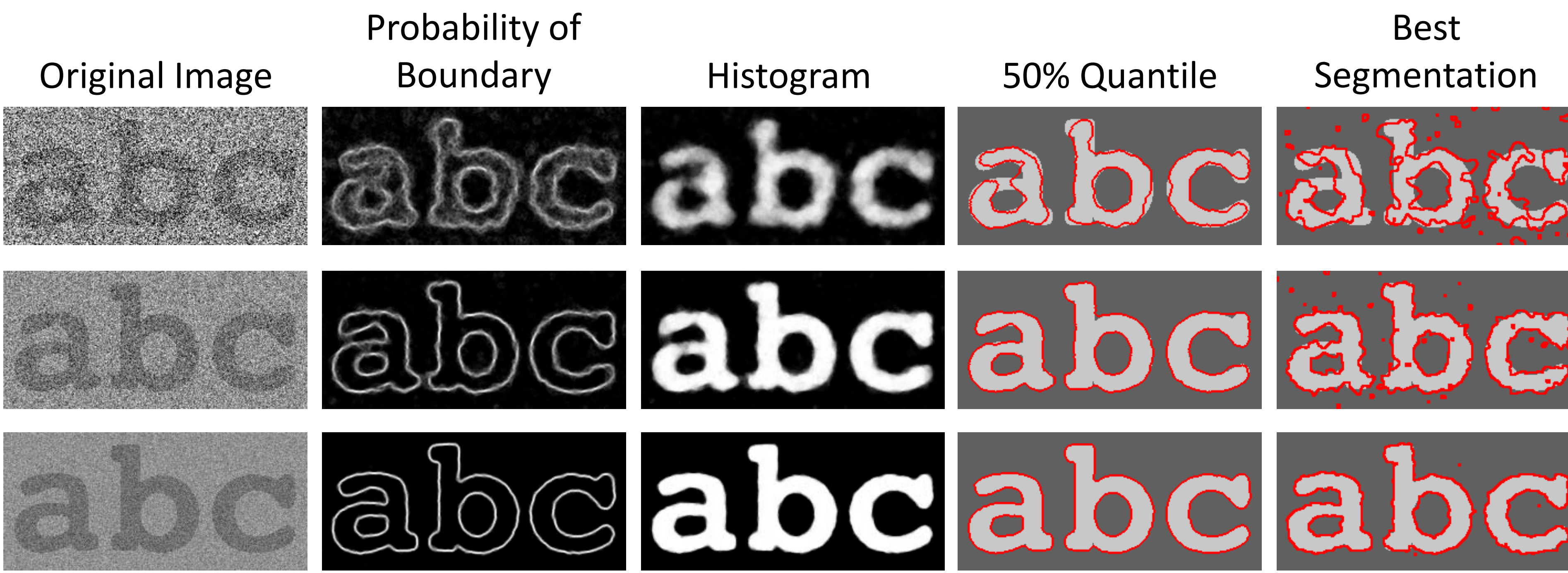
Computation Time



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

Synthetic Results

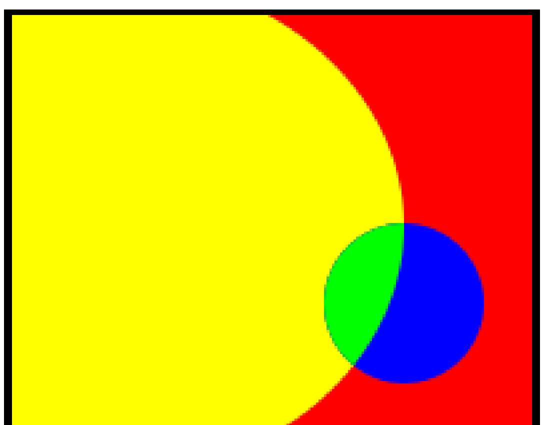
- Synthetic example with varying SNR
- When images have high SNR (i.e. are very separable), sampling makes less of a difference

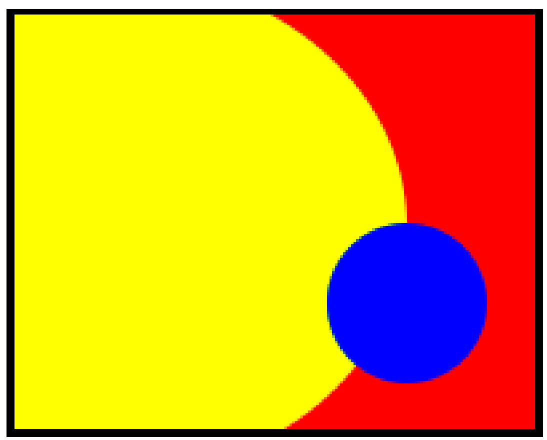


M-Ary Sampling

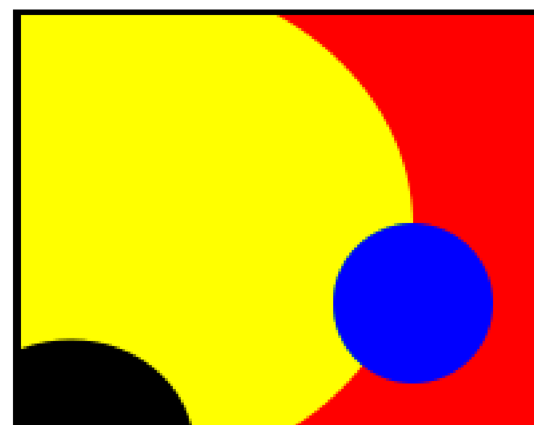
- M-ary segmentation typically achieved with multiple level sets
  - Have to ensure following conditions do not occur
    - Vacuum – pixels are not represented by any region
    - Overlap – pixels are represented by multiple regions

Overlap





Vacuum



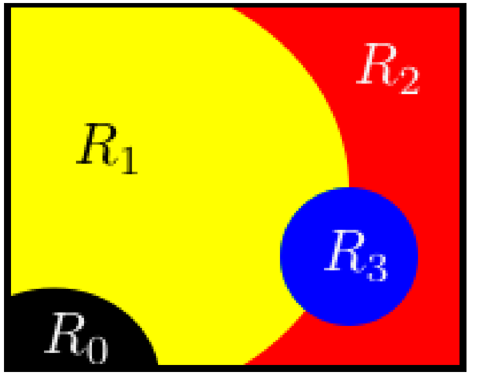
M-Ary Sampling

- 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
- 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



M-Ary Sampling

$$\text{Choose a random level set, } \ell$$

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

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

$$\mathbf{f}_\ell^{(t)} = \left(\mathbf{h}_\ell * \left(\mathbf{c}_\ell^{(t)} \circ \mathbf{n}_\ell^{(t)}\right)\right) \circ \mathbf{1}\left\{R_\ell \cup R_0\right\}$$

M-Ary Sampling

- For a pixel to move from R_ℓ to R_l it must go through R_0
- This must be reflected in our bias

$$\mathbf{v}(\ell, l) \triangleq \text{Gradient velocity between } \varphi_\ell \text{ and } \varphi_l$$

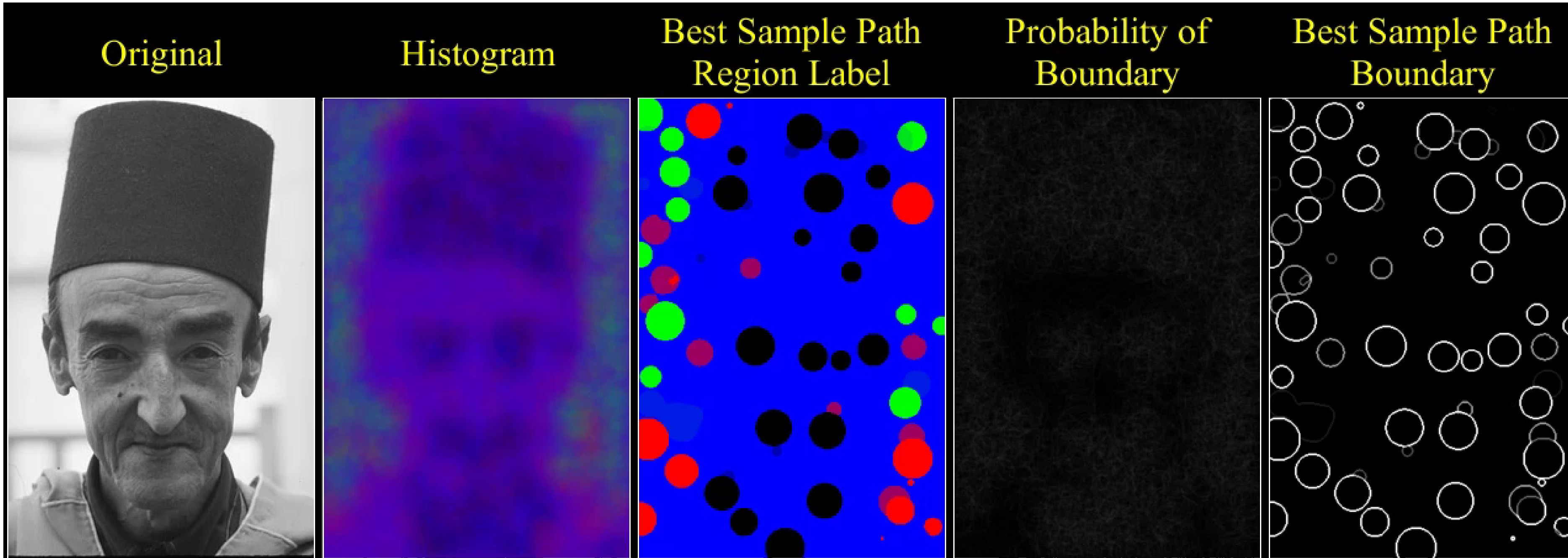
- Proposal only looks at $\mathbf{v}(\ell, 0)$
- Instead of biasing with gradient, bias with **minimal gradient**

$$m_i(\ell) \triangleq \min_{\substack{l \in \{0,1,2,\ldots,M\} \\ l \neq \ell}} v_i(\ell, l)$$

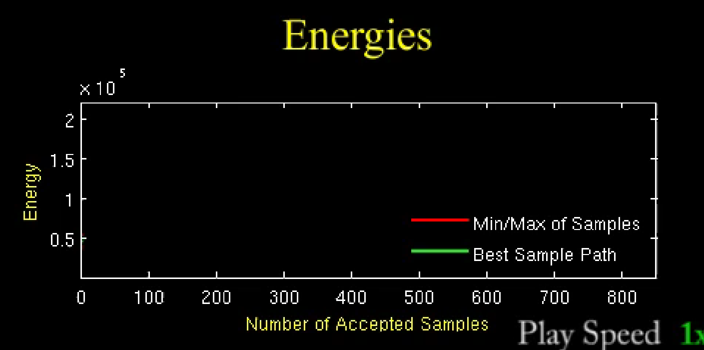
- When using mutual information, the minimal gradient is

$$m_i(\ell) = \log \frac{p_X^\ell(x_i)}{p_X^{\max}(x_i)} \qquad p_X^{\max}(i) = \max_{\substack{l \in \{0,1,2,\ldots,M\} \\ l \neq \ell}} p_X^l(x_i)$$

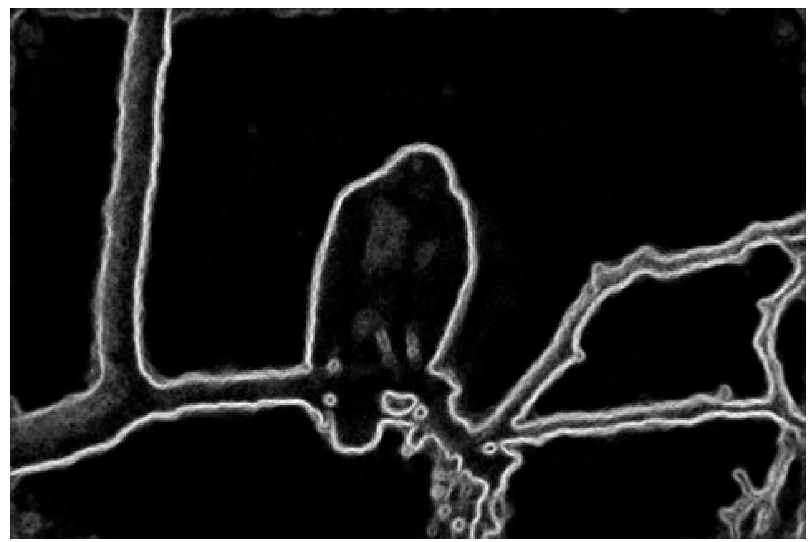
A Natural Image



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.



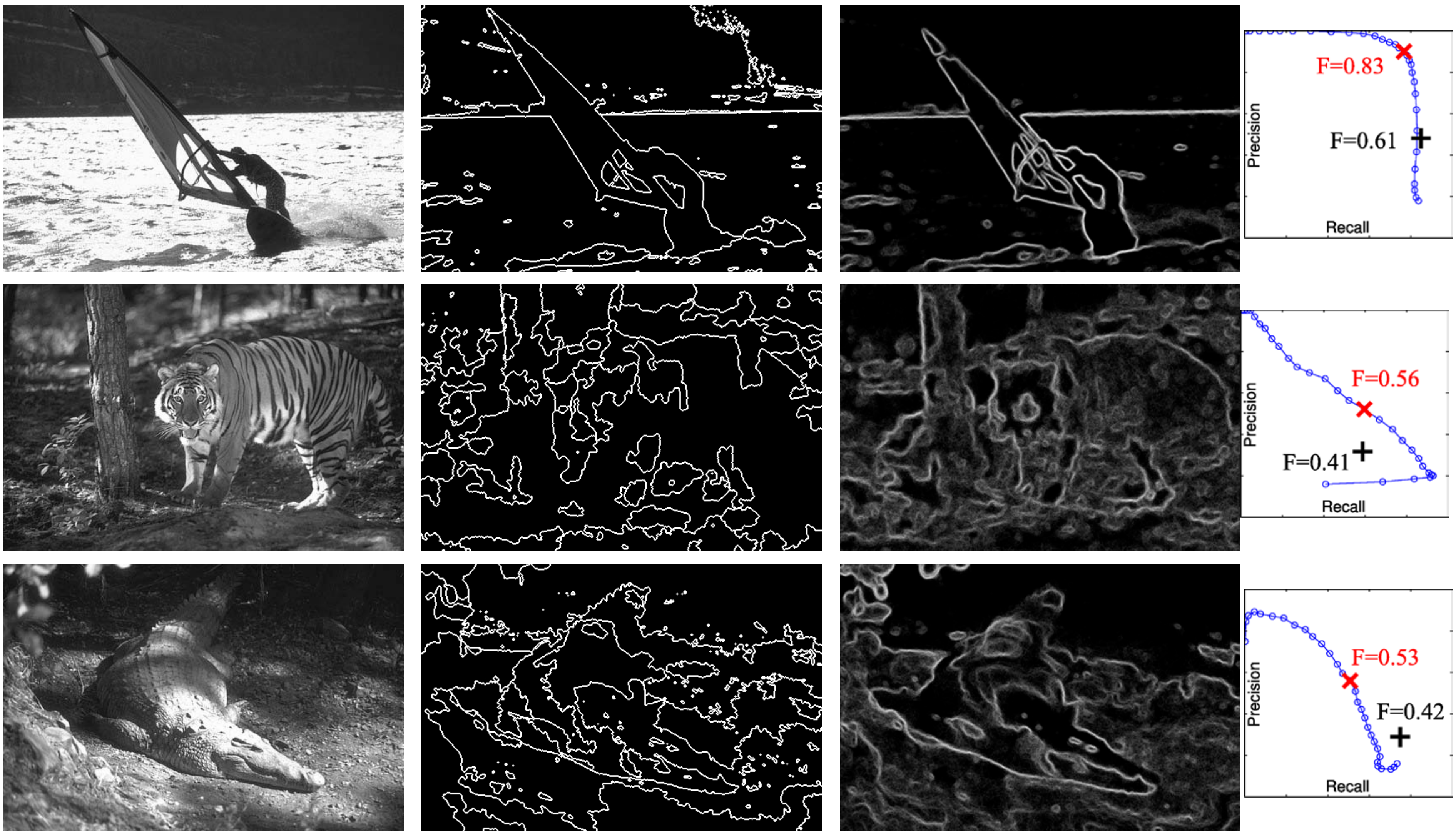
Example Sampling vs. Optimization



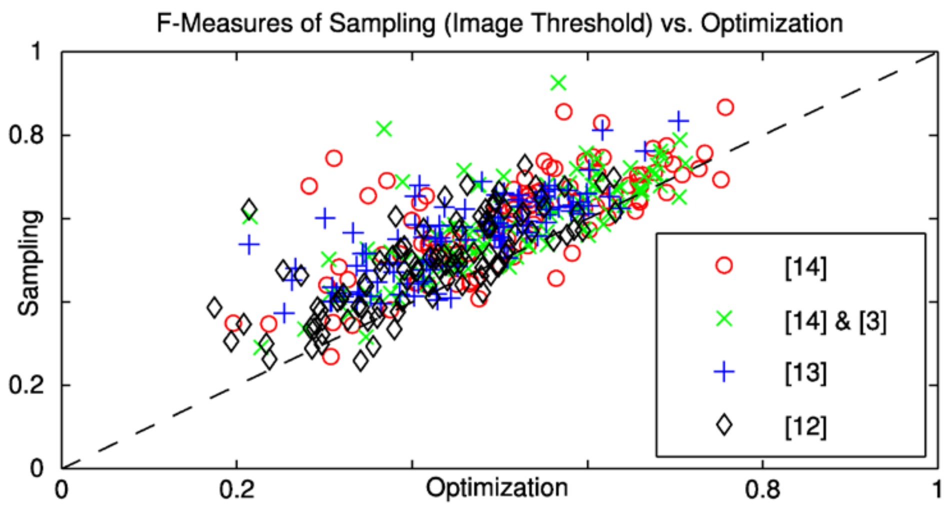
CSAIL

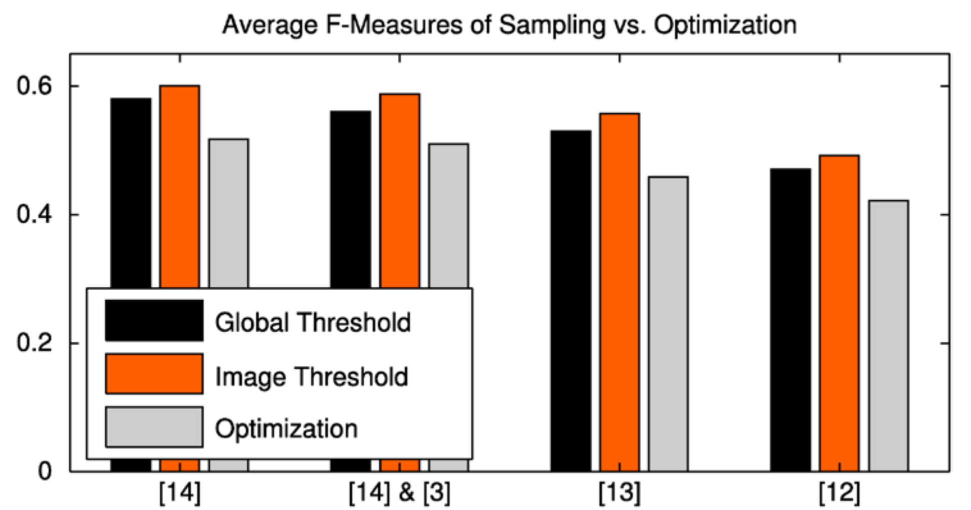
Results on the BSDS

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





[3] Chang, J. and J.W. Fisher III. Analysis of Orientation and Scale in Smoothly Varying Textures. ICCV 2009.

[12] Heiler, M. and C. Schnorr. Natural Image Statistics for Natural Image Segmentation. ICCV 2003.

[13] Houhou, N., Jp.P. Thiran, and X. Bresson. Fast Texture Segmentation Model Based on the Shape Operator and Active Contour. CVPR 2008.

[14] Kim, J., J. W. Fisher II, A. Yezzi,. M. Cetin, and A. Willsky. A nonparametric statistical method for image segmentation using information theory and curve evolution. IEEE Trans. on Image Processing 2005.

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