Image Penalties Based on Neighborhood Statistics (+ Toward Variation)

Ross T. Whitaker

Scientific Computing and Imaging Institute
School of Computing
University of Utah

SCI
INSTITUTE

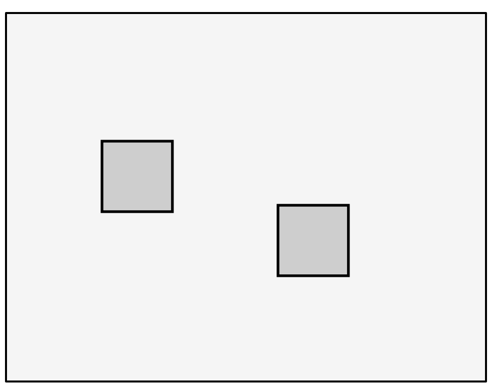
Talk Overview

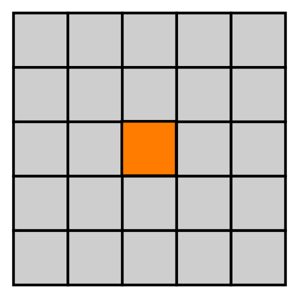
- Algorithm
- NL Means and Optimal Estimators
- Alternate Formulation - Entropy
- Relationships to Other Methods, Extensions
- Toward a Variational Formulation

Denoising Algorithm

$$\hat{f}_i = \frac{\sum_{j \in A} w_{ij} f_j}{\sum_{j \in A} w_{ij}}$$

$$w_{ij} = \exp\left(\frac{\sum_{k \in \mathcal{N}} (f_{i+k} - f_{j+k})^2}{2\sigma^2}\right)$$





Typical Results

Denoising Algorithm

- NL-means: Baudes, Coll, Morel 2005
  - Optimal estimator
  - Single iteration
  - A is large block of pixels
  - σ is related to image noise
- UINTA: Awate, Whitaker 2005
  - Entropy formulation
  - Iterate
  - A is a random subset of image
  - σ optimized through cross validation

SCI INSTITUTE

NL-Means Formulation

- Markov random field
  - Pixels are random variables
  - Conditional distributions depend on neighbors
  - $P(X_u | I \backslash X_u) = P(X_u | Y_u)$ (+consistency)
- Stationarity
- Mixing property

$$\lim_{\|u-v\| \to \infty} |P(X_u, X_v) - P(X_u)P(X_v)| = 0; \forall X_u, X_v \in \mathbf{X}$$

SCI INSTITUTE

Image Model

- Noise model

$$\tilde{G} = G + N$$

- Goal: estimate F from I
- Strategy
  - Construct function of pixel nbhd: $\hat{g}_i \leftarrow f(\tilde{y}_i)$
  - Optimize f

$$\mathrm{argmin_f E}\left[\left(\mathrm{g_i} - \mathrm{f}(\tilde{\mathrm{y}}_\mathrm{i})\right)^2\right]$$

$$f(\tilde{y}_i) = E\left[X|\tilde{Y} = \tilde{y}_i\right] = E\left[\tilde{X}|\tilde{Y} = \tilde{y}_i\right]$$

Estimating Conditional Expectation

- Nonparametric regression (Nadaraya-Watson)

$$E\left[X|Y=y\right] = \frac{1}{n}\frac{\sum_{i=1}^{n} x_i K(y-y_i)}{\sum_{i=1}^{n} K(y-y_i)}$$

- (x_i, y_i) independent samples from joint distribution
- Asymptotic convergence (pointwise) n->infty

Estimation From Image Neighborhoods

- MRF -> the image pixels are samples from the conditional
- Independence: strictly, not so
  - Mixing property (Levina 1998) => asymp. unbiased as n->infty
- So…
  - Can estimate $f(\tilde{y}_i)$

  with lots of image samples
- NL-means
  - Each pixel becomes a weighted average of pixels with similar nieghborhoods

SCI INSTITUTE

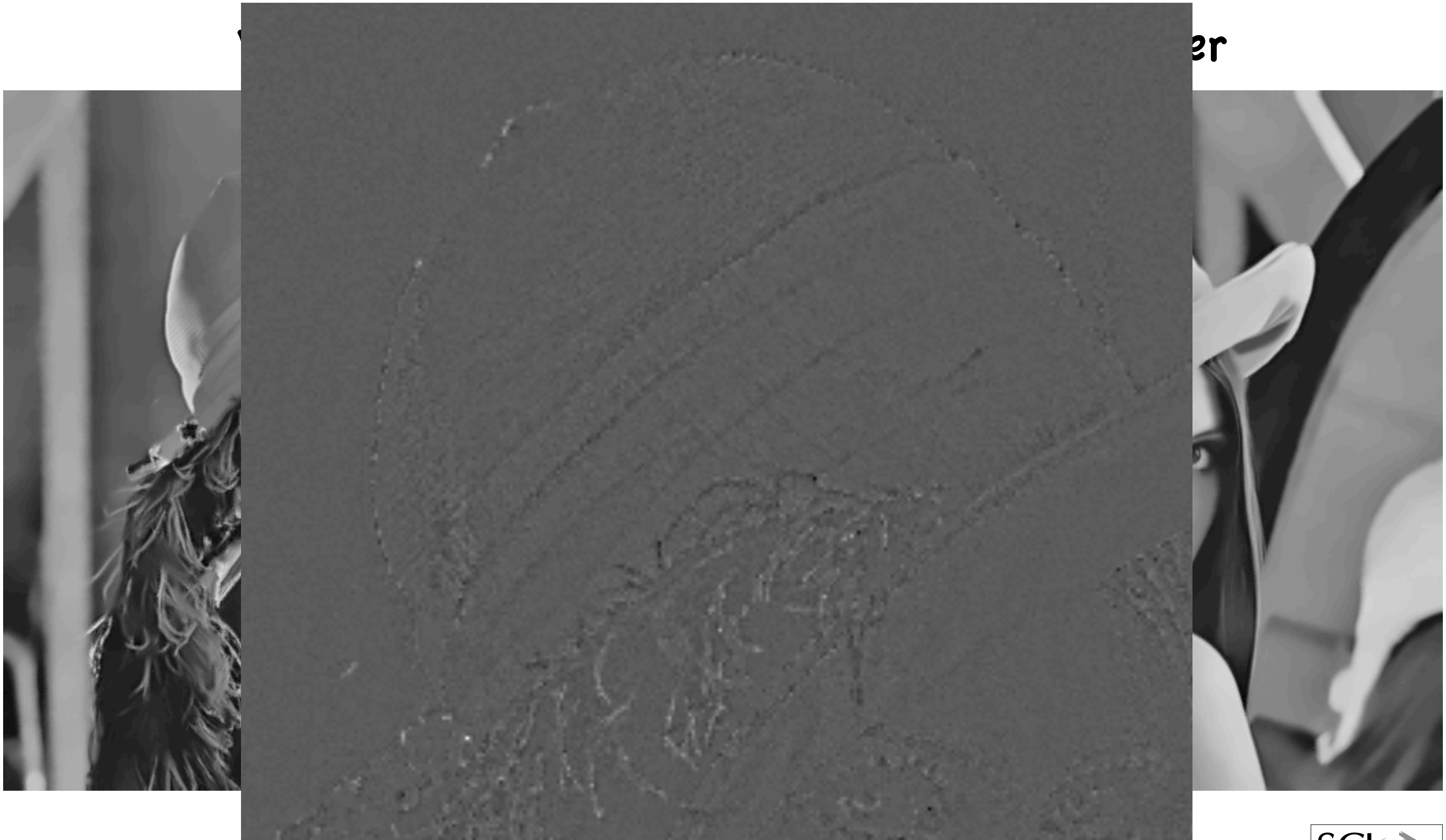
Practical Considerations

- Set A need not be random
  - Large block of pixels
- Images are not fully stationary
  - Nearby statistics dominate
  - Choose A to be near the pixel in question
- Bandwidth of kernel (Gaussian) is important
  - Too small -> estimates are noisy
  - Too large -> estimates are biased
- Computation time is significant

What Does the NL-Means Theory Tell Us?

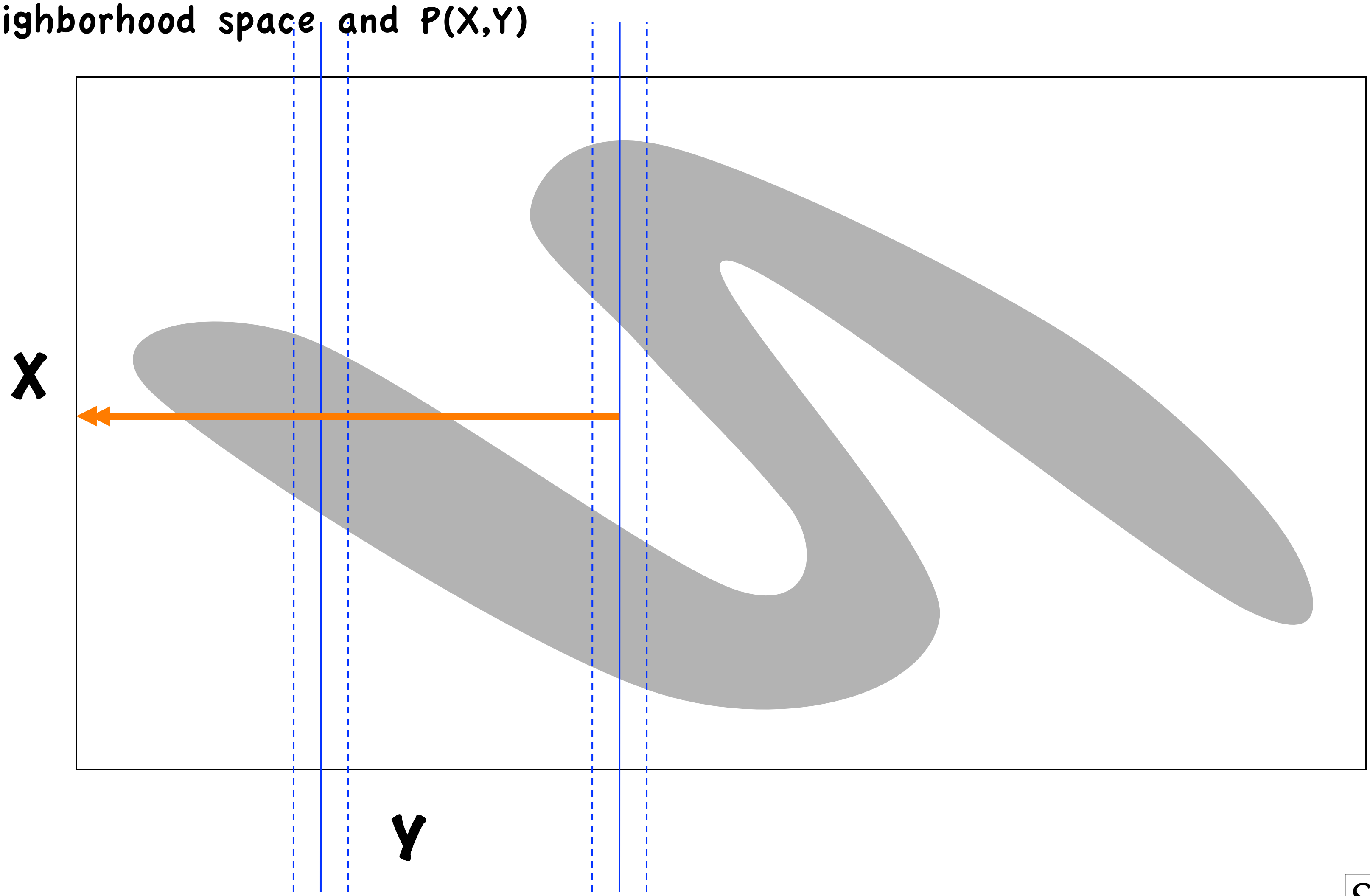
- One pass is enough
  - "optimal"
  - Yet, in practice, many people iterate
- What about the center pixel?
  - In theory, Y excludes the center pixel
    - You cannot condition a random variable on itself
  - In practice, everyone includes the center pixel in the nbhd comparisons
    - What better indicator is there of a pixels value?

Affects of Center Pixel



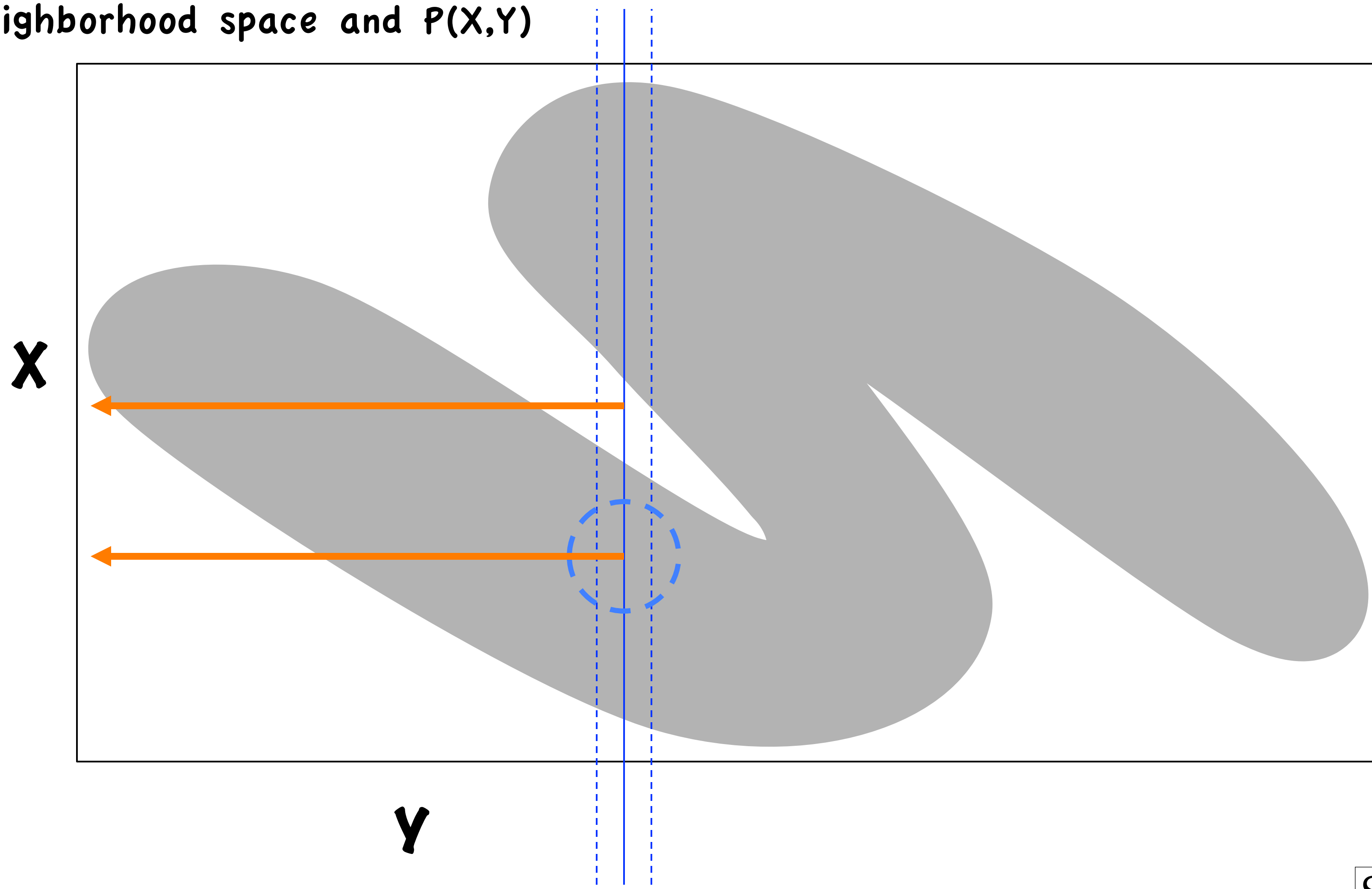
Visualizing NL-Means - No Center Pixel

Neighborhood space and P(X,Y)



Visualizing NL-Means - with Center Pixel

Neighborhood space and P(X,Y)



Center Pixel

- It makes "sense" to use the center pixel in comparing neighborhoods
  - Results are quantifiably better
- Is there a theory that explains why the center pixel helps?

SCI INSTITUTE

Iterations

- Some researchers have noticed iterating helps <u>in some cases</u>

One iteration

One iteration – smaller σ

Multiples iterations – smaller σ

Conditional Distributions of the Observed Image

- Optimal for $f(\tilde{y})$
  - But the neighborhoods are noisy
- W.r.t. the ideal image

$$E\left[X|\tilde{Y} = \tilde{y}_i\right]$$

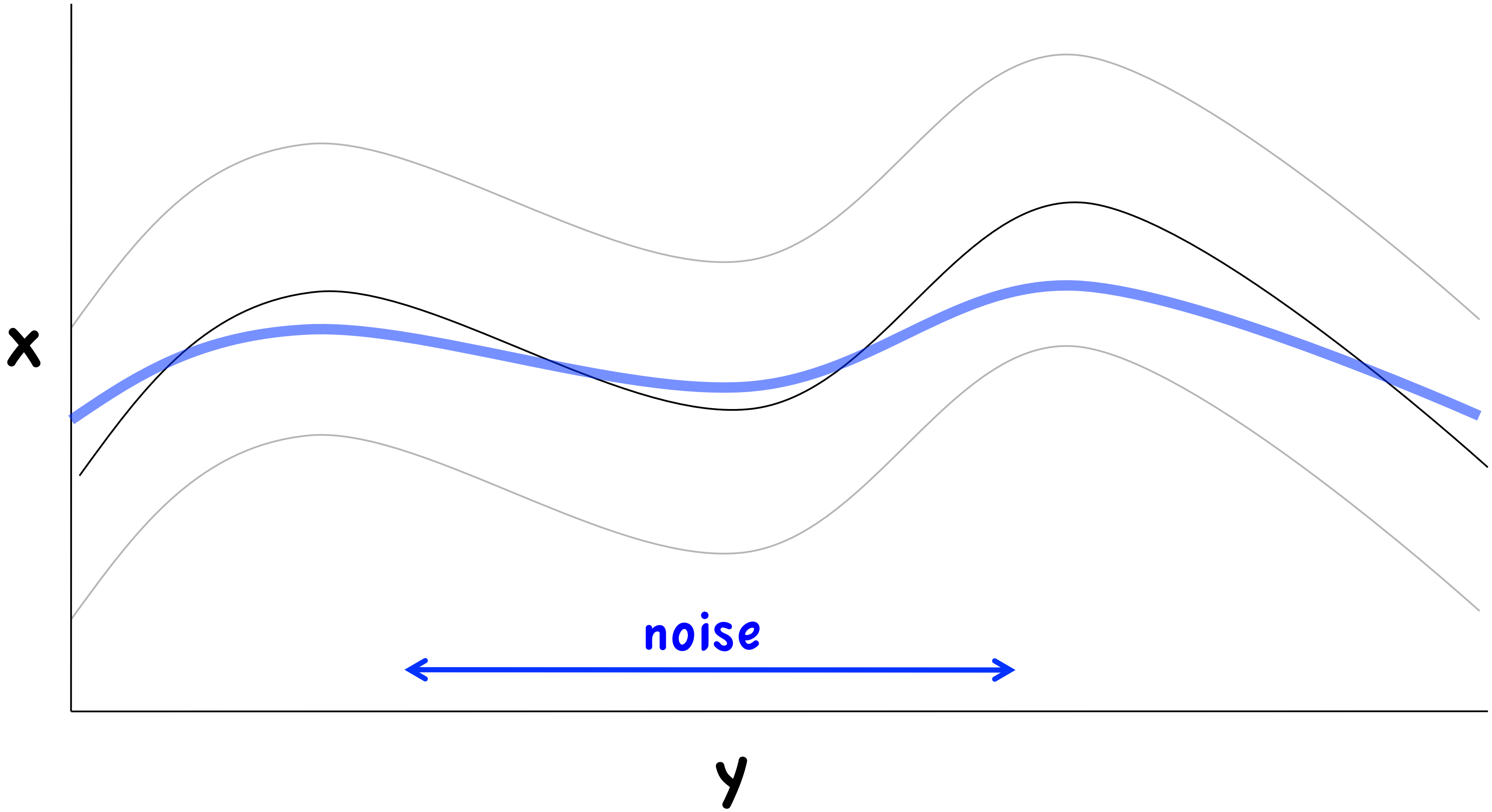
is biased (error beyond variability of G)

- Ideally we would like

$$E\left[X|Y = y\right]$$

SCI INSTITUTE

Measurement Noise in Regression



Optimal Nonparametric Regression with Measurement Noise

- Difficult, open problem
- Some work in statistics
- Known noise -> deconvolution problem

An Alternate Formulation

- Consider the joint distribution of pixels and their neighborhoods

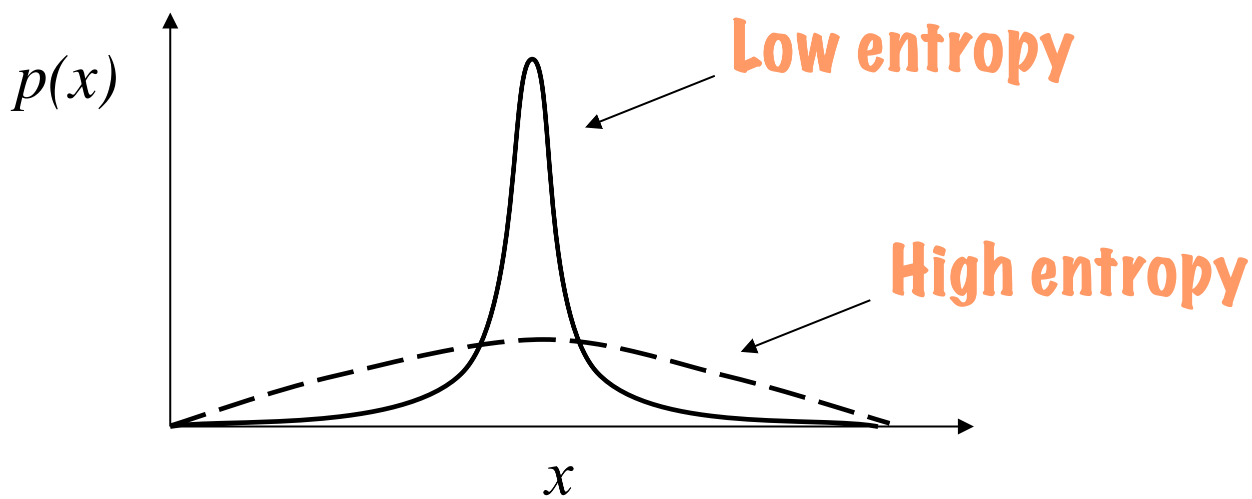
$$P(X,Y) = P(Z)$$

- The entropy of this distribution describes:
  - inherent stochasticity in image
  - the repeatability of image neighborhoods
  - the degree to which the image "looks like itself"
  - the amount of noise in the image

Entropy
(Shannon 1948)

- Entropy of a random variable *X* (instance x)
  - Measure of *uncertainty* – information content of a

$$h(X) = -\int p(x) \log p(x) dx = -E_p\left[\log p(X)\right]$$



UINTA Strategy
Awate & Whitaker 2005

- Treat the entropy of the image as a measure of "regularity" or "goodness"
  - An alternative geometric quantities such as TV
- Estimate the entropy with nonparametric density estimation
- Use an iterative strategy to reduce entropy
  - Combine with other terms, noise models, etc.

Estimating Entropy Nonparametrically

- Expectation of log(P) via sample mean

$$h(X) \approx \frac{1}{n}\sum_{i=1}^{n}\Big(-\log P(x_i)\Big)$$

- Estimate P for neighborhoods (Z) using Parzen windowing

$$P(\mathbf{z}_t) \approx \frac{1}{|\mathcal{A}_t|}\sum_{u\in\mathcal{A}_t} G_d(\mathbf{z}_t - \mathbf{z}_u; \Psi\ ), \text{ where}$$

$$t \notin \mathcal{A}_t.$$

G() - Gaussian kernel

Ψ Covariance/bandwidth

SCI INSTITUTE

Estimating Entropy Nonparametrically

- Expectation of log(P) via sample mean

$$h(X) \approx \frac{1}{n}\sum_{i=1}^{n}\Big(-\log P(x_i)\Big)$$

- Estimate P for neighborhoods (Z) using Parzen windowing

$$P(\mathbf{z}_t) \approx \frac{1}{|\mathcal{A}_t|}\sum_{u\in\mathcal{A}_t} G_d(\mathbf{z}_t - \mathbf{z}_u; \Psi\ ),\ \text{where}$$

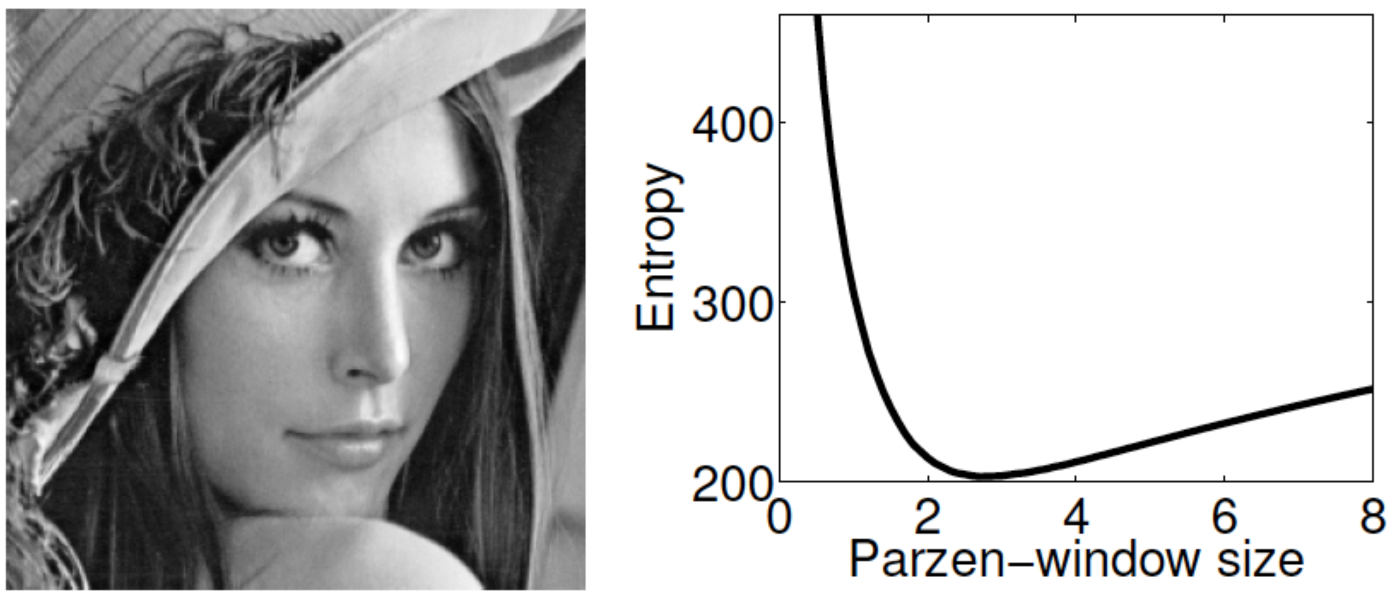
$$t \notin \mathcal{A}_t.$$

G() - Gaussian kernel

Ψ Covariance/bandwidth

Computations on Entropy

- Select kernel bandwidth to minimize entropy
  - Maximum likelihood with cross validation

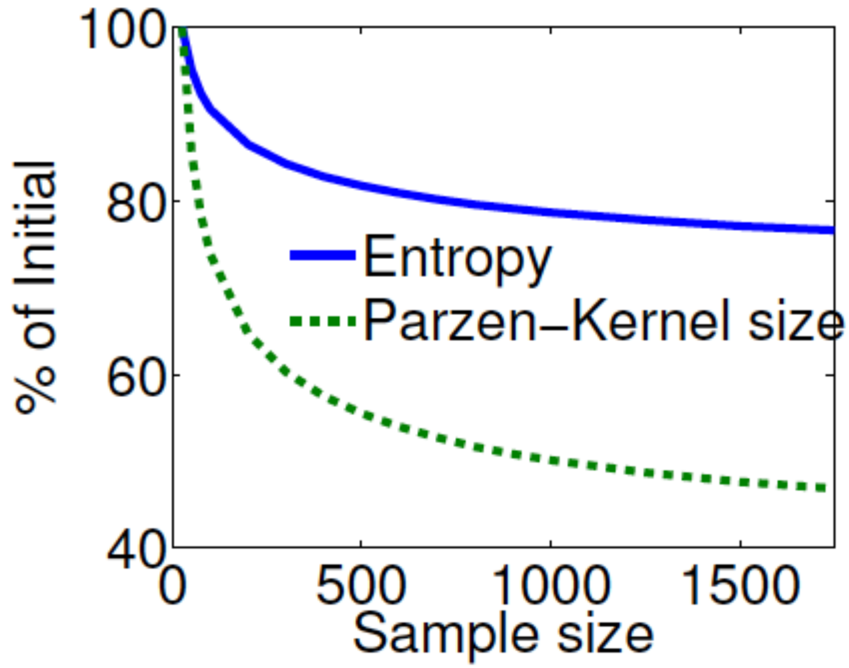


(a) (b)

Figure 3.1. Optimal kernel bandwidth. (a) The *Lena* image. (b) The entropy estimate for the *Lena* image as a function of Parzen-window kernel σ.

Samples and Bandwidth

(a)



(b)

Entropy Minimization

- Entropy as sample mean

$$
\begin{aligned}
h(Z) &= -E_p[\log p(Z)] \\
&\approx \frac{1}{|B|}\sum_{i\in B} \log p\left(z_i\right) \\
&\approx \frac{1}{|B|}\sum_{i\in B} \log\left(\frac{1}{|A|}\sum_{j\in A} G(z_i - z_j, \psi)\right)
\end{aligned}
$$

  - Set *B*: all pixels in image
  - Set *A*: a small *random* selection of pixels
  - z_i shorthand for $z(s_i)$
- Stochastic approximation

Entropy Minimization

- Stochastic approximation
  - Reduce $O(|B|^2)$ to $O(|A||B|)$
  - Efficient optimization
- Stochastic-gradient descent

$$
\begin{aligned}
\Delta x &= -\lambda \frac{\partial h(X|Y=y)}{\partial x} \\
&\approx \frac{\lambda \psi^{-1}}{|B|} \left[\sum_{j \in A} \frac{G(z_j - z, \Psi)}{\sum_{k \in A} G(z_k - z, \Psi)} x_j - x \right]
\end{aligned}
$$

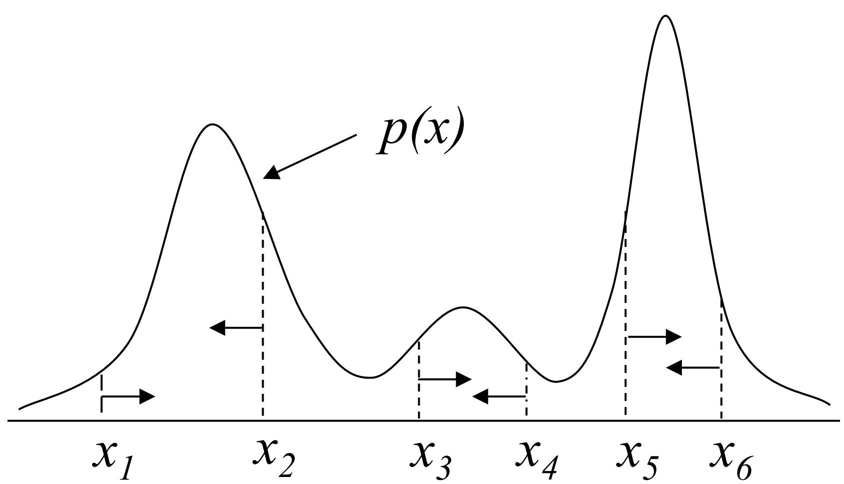
SCI INSTITUTE

Mean-Shift Procedure (Fukunaga et al. 1975)

- Fixed point: derivative=0, weights lag <-> mean shift

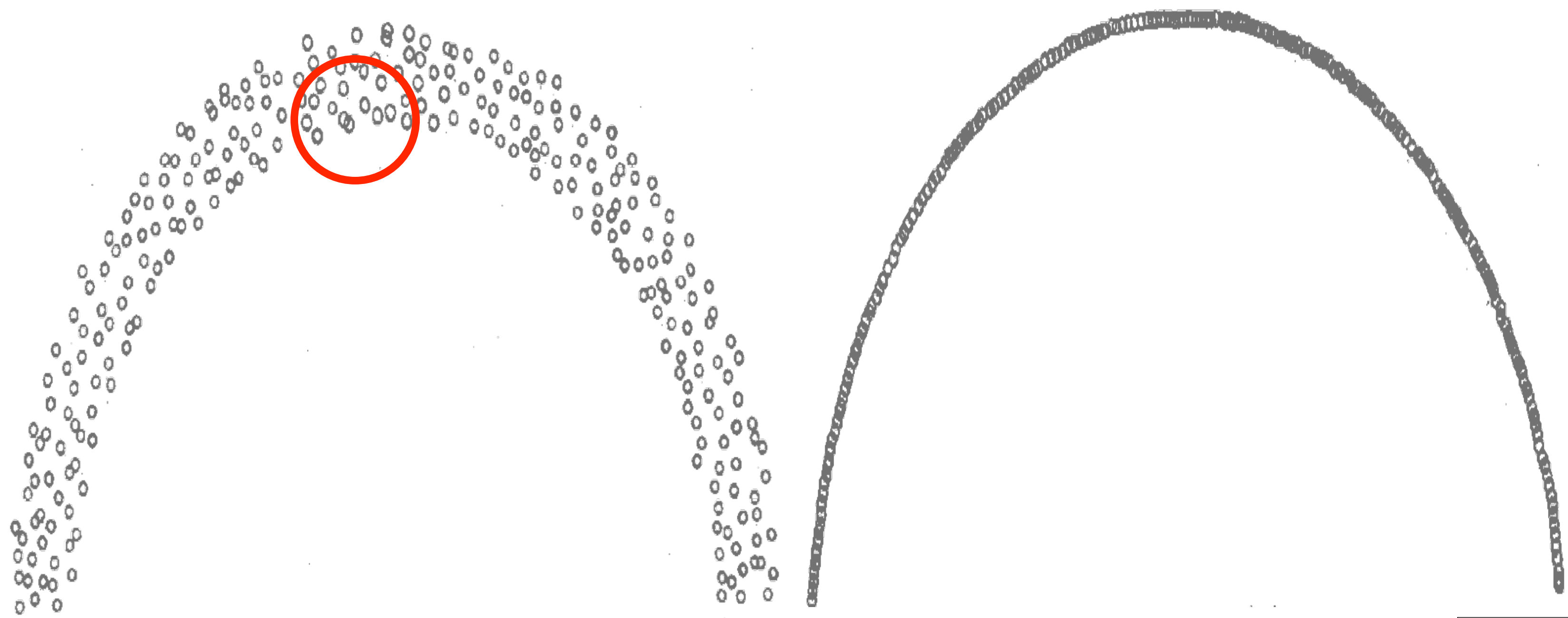
$$x_i \longleftarrow \sum_j w_{ij} x_j$$

- Mean-shift – a mode seeking procedure



Mean-Shift Procedure (Fukunaga et al. 1975)

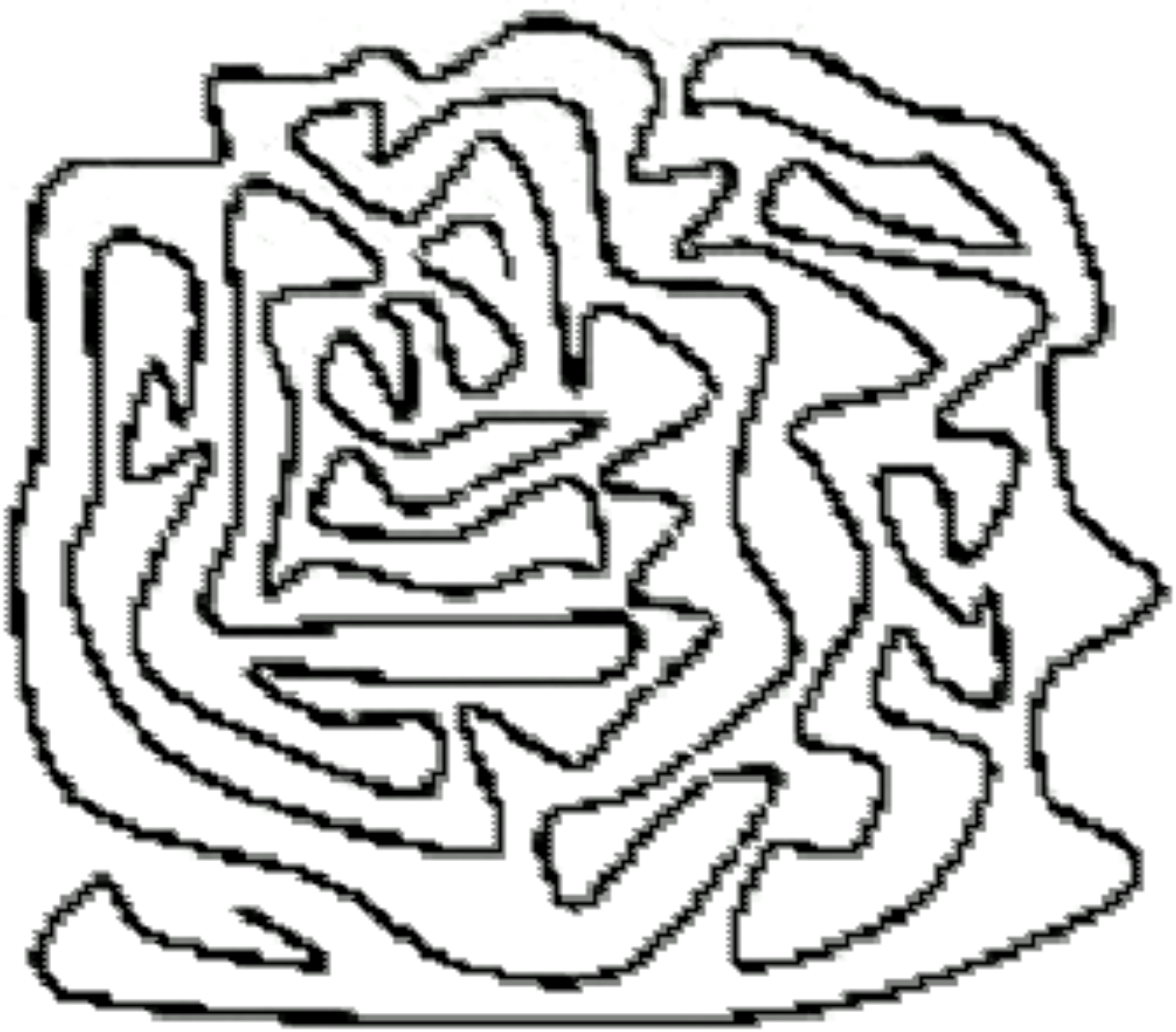
- Data filtering to reduce noise
  - Hand tuned parameters



Relationships to Other Image Filters

- Bilateral filter
  - Lowpass blurring (averaging locally in image space)
  - B.L. filter (averaging locally in space+intensity)
- UINTA/NL-means
  - Averaging locally in the space if image neighborhoods

Entropy Scale Space?



SCI INSTITUTE

Adding a Noise Model

$$E(u) = \lambda \int (u-g)^2 dx + \int |\nabla u|^2 dx$$ **Dirichlet**

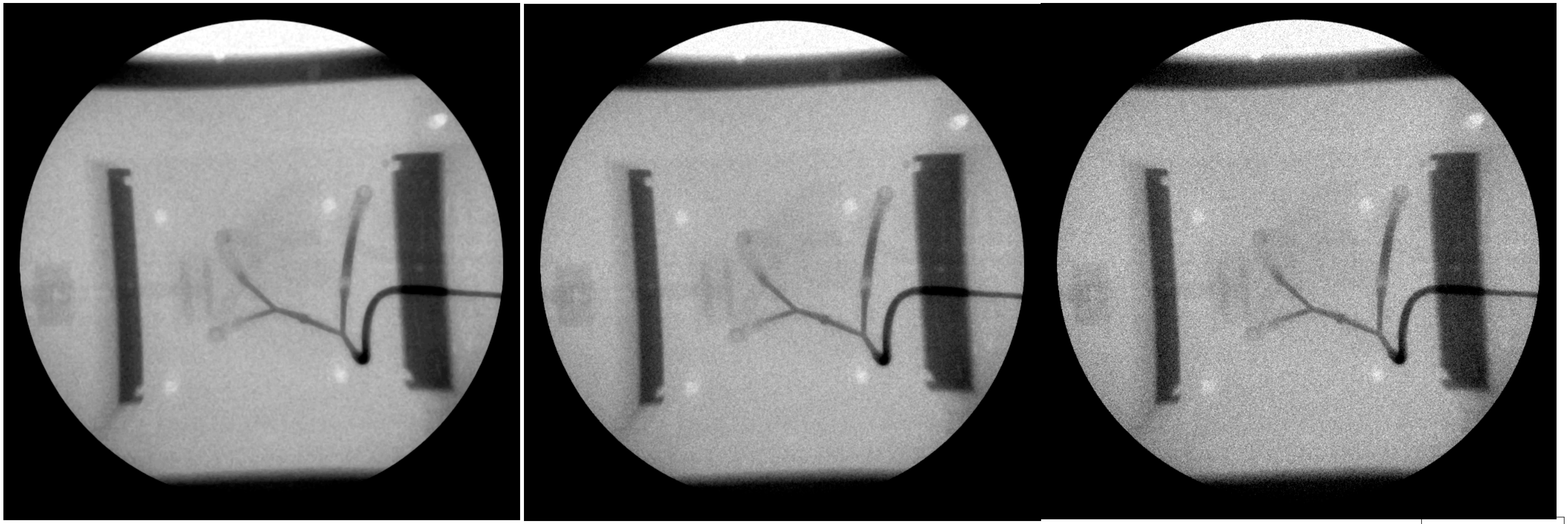
$$E(u) = \lambda \int (u-g)^2 dx + \int |\nabla u| dx$$ **Total Variation**

$$E(u) = \lambda \int (u-g)^2 dx + H(Z_u)$$ **Entropy**

Fixed Point Algorithm

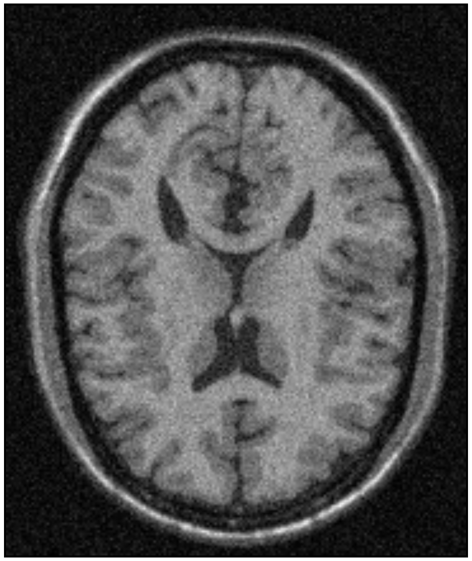
- Control the effects of the input data

$$u_i^{k+1} = \frac{\lambda g_i + \sum_j w_{i,j} u_j}{\lambda + \sum_j w_{i,j}}$$

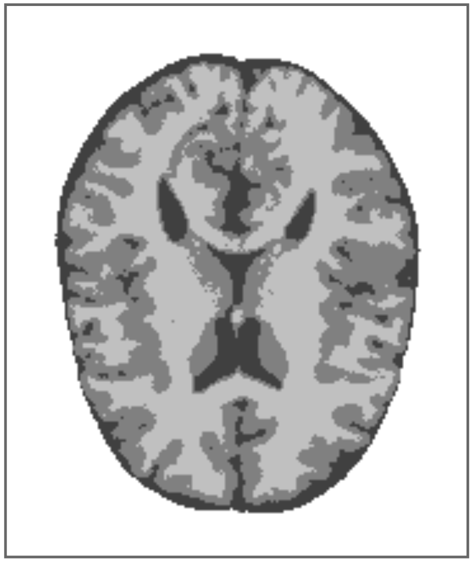


MRI Tissue Classification

- Algorithm: 1) initialize with atlas, 2) iteratively relabel to reduce tissue-wise nhd entropy

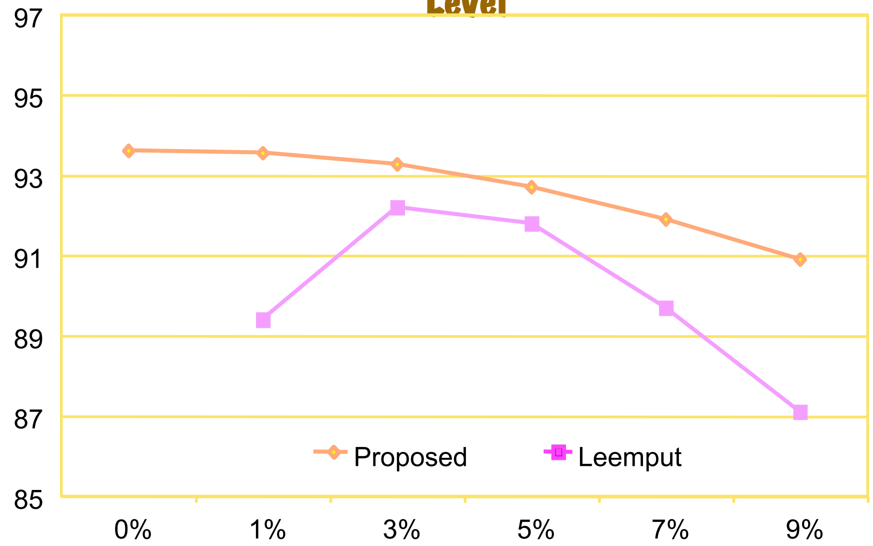


MRI Input



GM, WM, CSF Seg.

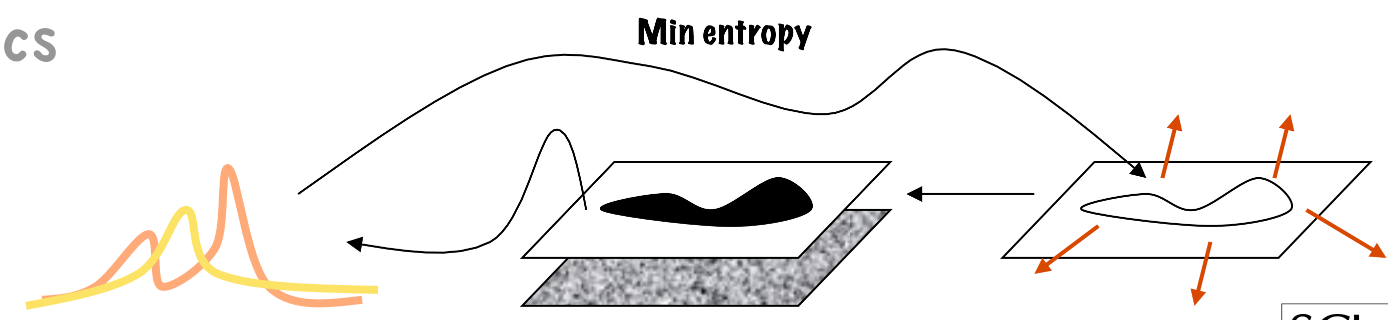
GM Classification Performance vs Noise Level



Comparison: SOTA-EM w/MRFs & Atlas (Leemput et al.)

Texture Segmentation

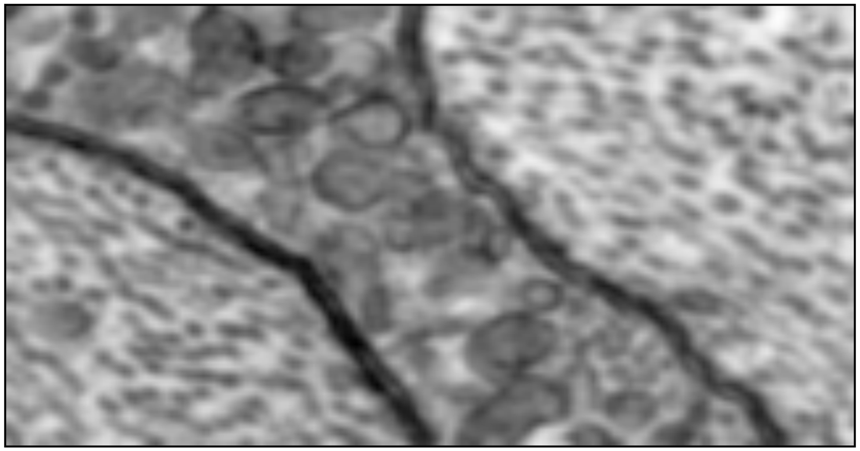
- Reassign class labels to reduce in-class entropy
  - Deformable model to keep spatial coherence
- Recompute pdfs from new class labels
  - Random samples + nonparametric nhd statistics
- Iterate



Texture Segmentation
Awate et al., 2005

- Initialization -> checkerboard
- Deformable model -> level sets (Tsai and Seglmi, 2004)






Is There A Variational Formulation?

Continuous Formulation

1. Convert updates to PDE
2. Entropy formulation directly to images/ functions

$$u(x),\ u : \Re^2 \mapsto \Re$$ **image**

$$n(x),\ n : \Re^2 \mapsto \Re$$ **Neighborhood mask**

Differences of Neighborhoods

$$K : \Re^2 \times \Re^2 \mapsto \Re$$

$$k(x,y) = K\left(\int (u(x+\alpha) - u(y+\alpha))^2 n^2(\alpha) d\alpha\right)$$

$$P(z_x) \approx C \int k(x,y) dy$$

C is a normalization- something to do with supp(u)

$$H(Z) = \int P(z) \lg P(z) dz$$

SCI INSTITUTE

Entropy

$$H(Z) = \int P(z) \lg P(z) dz$$

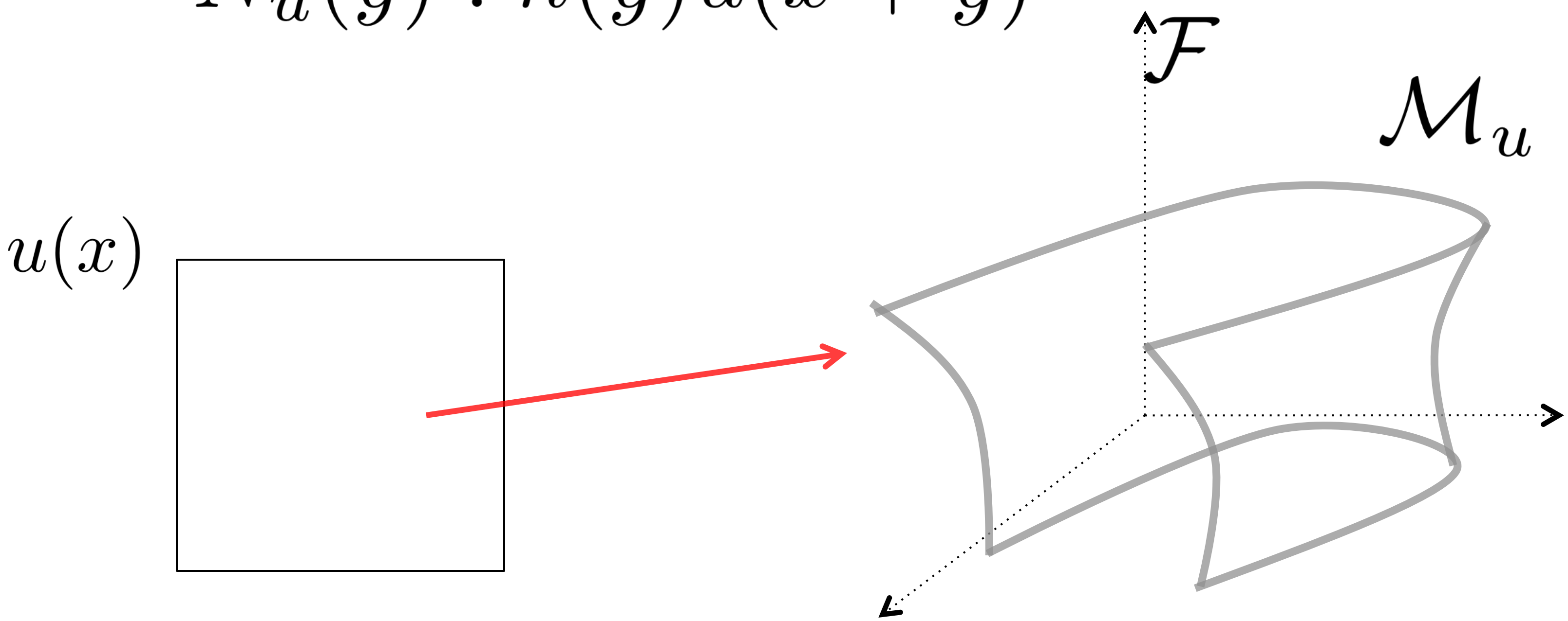
$$H(Z_u) \approx \int \lg \left[\int k(x,y) dy \right] dx + C$$

$$H(u) = \int \lg \left[\int K \left(\int (u(x+\alpha) - u(y+\alpha))^2 n^2(\alpha) d\alpha \right) dy \right] dx$$

How Do We Make Sense of This?
(Hand Waving and Speculation)

$$N_u(y) : \Re^2 \mapsto \mathcal{F}$$

$$N_u(y) : n(y)u(x+y)$$



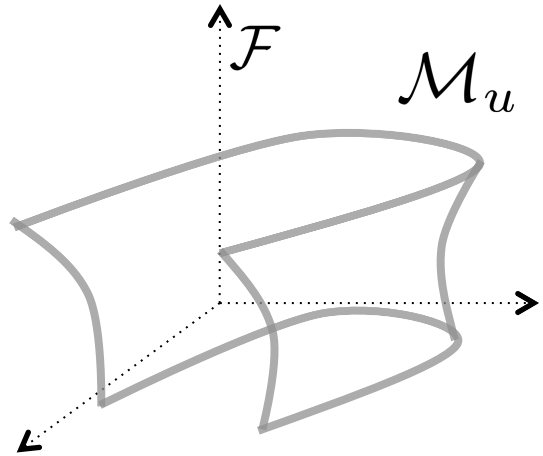
More Speculation on Continuous Case

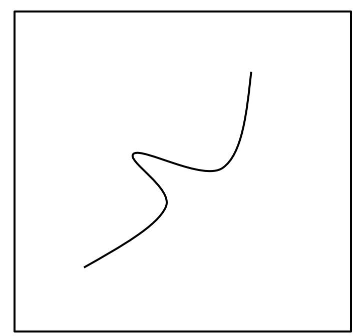
- Think of M_u as a measure
  - density dM_u from $u(x)$
- For z in F
  - M induces a potential field

$$U(z) = \lg\left[\mathcal{M} \otimes \mathcal{K}\right]$$

  - U(z) is related to distance to a smoothed version of M

$$f \in \mathcal{F}, \quad \mathcal{K}(f) = K(||f||_{L2})$$





SCI INSTITUTE

Continous Continued...

- The entropy is an integral over M
  - accounting for local density

$$H(u) = \int_{\mathcal{M}} U(f) df$$

$$\frac{dH}{du} = \frac{dH}{d\mathcal{M}} \cdot \frac{d\mathcal{M}}{du}$$

????