Image Processing with Nonparametric Neighborhood Statistics

Ross T. Whitaker

Scientific Computing and Imaging Institute
School of Computing
University of Utah

PhD - Suyash P. Awate
University of Pennsylvania, Dept of Radiology

See also: http://www.cs.utah.edu/~suyash/pubs/uinta.html

SCI INSTITUTE

Papers

- Suyash P. Awate and Ross T. Whitaker, "Higher-Order Image Statistics for Unsupervised, Information-Theoretic, Adaptive Image Filtering UINTA", IEEE Computer Vision and Pattern Recognition (CVPR) 2005, v 2, pp 44-51
- Suyash P. Awate and Ross T. Whitaker, "Nonparametric Neighborhood Statistics for MRI Denoising", Information Processing in Medical Imaging (IPMI) 2005, pp 677-688
- Tolga Tasdizen, Suyash P. Awate, Ross T. Whitaker, Norman Foster, "MRI Tissue Classification with Neighborhood Statistics: A Nonparametric, Entropy-Minimizing Approach", Medical Image Computing and Computer Assisted Intervention (MICCAI) 2005, v 2, pp 517-525
- Suyash P. Awate and Ross T. Whitaker, "Unsupervised, Information-Theoretic, Adaptive Image Filtering with Applications to Image Restoration UINTA", IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI) 2006, 28(3):364-376
- Suyash P. Awate, Tolga Tasdizen, Ross T. Whitaker, "Unsupervised Texture Segmentation with Nonparametric Neighborhood Statistics", European Conference on Computer Vision (ECCV) 2006
- Suyash P. Awate, Tolga Tasdizen, Norman Foster, Ross T. Whitaker, "Adaptive, Nonparametric Markov Modeling for Unsupervised, MRI Brain-Tissue Classification", Medical Image Analysis (MEDIA) 2006, 10(5):726-739
- Suyash P. Awate and Ross T. Whitaker, "Feature-Preserving MRI Denoising using a Nonparametric, Empirical-Bayes Approach", IEEE Trans. Medical Imaging (TMI) 2007 (To Appear)

SCI INSTITUTE

Talk Overview

- Motivation
- Image denoising
- Density estimation
- UINTA filtering strategy overview
- Entropy minimization
- Implementation issues: statistics, image processing
- Other applications
- Final thoughts

Images

Denoising Vs Reconstruction

- Any geometric/statistical penalty can be applied in two ways:
  1. Gradient descent as filter (choose # iterations)
  2. With data (fidelity) term to steady state
     - Variational
     - Noise/measurement models, optimality, etc.

SCI INSTITUTE

Variational Methods
E.g Anisotropic Diffusion

- Perona&Malik (1990)

$$\frac{\partial f}{\partial t} = \nabla \cdot c(|\nabla f|)\nabla f$$

- Penalty:
  - Quadratic on grad-mag with outliers (discontinuities)
    - Nordstrom 1990; Black et. al 1998
  - Favors piecewise const. Images

SCI INSTITUTE

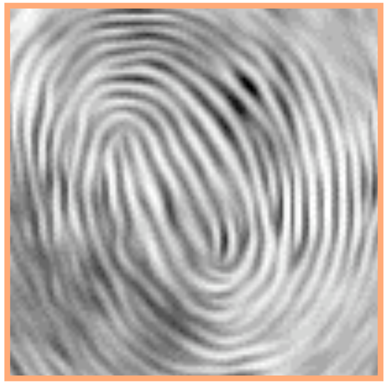
Other Flattening Approaches

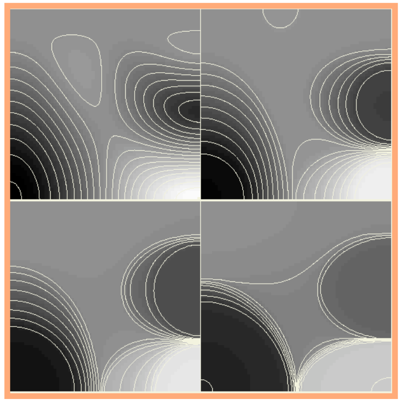
- Total variation
  - Rudin et. al (1992)
- Mumford-Shah (1989) related
  - Explicit model of edges
  - Cartoon model
- Level sets to model edges
  - Chan & Vese (2000)
  - Tsai, Yezzi, Willsky (2000)
- Model textures + boundaries
  - Meyer (2000)
  - Vese & Osher (2002)

SCI INSTITUTE

PDE Methods
Other Examples

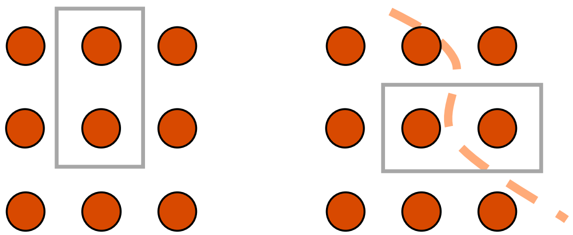
- Weickert (1998)
  - Coherence enhancing
- Tasdizen et. al (2001)
  - Piecewise-flat normals
- Wilmore flows
  - Minimize curvature





Markov Random Fields
E.g. Geman and Geman (1984)

- Gibbs energies on cliques
  - Quantify image preferences
    - Discrete geometric configurations
  - Given a priori
  - Hidden variables/processes to capture features



Issues

- Prioritize geometric configurations a priori
  - Works well of the model fits, otherwise...
- Free parameters
  - Thresholds -> determine when to apply different models (e.g. "preserve edge or smooth")
- Generality
  - Cartoon-like simplifications are disastrous in many applications
- Increasing the geometric complexity
  - Is there a better way?

Examples

Lena

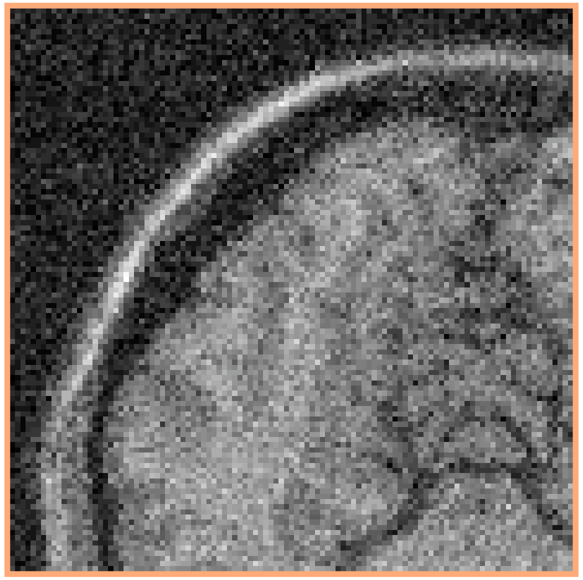
Anisotropic Diffusion

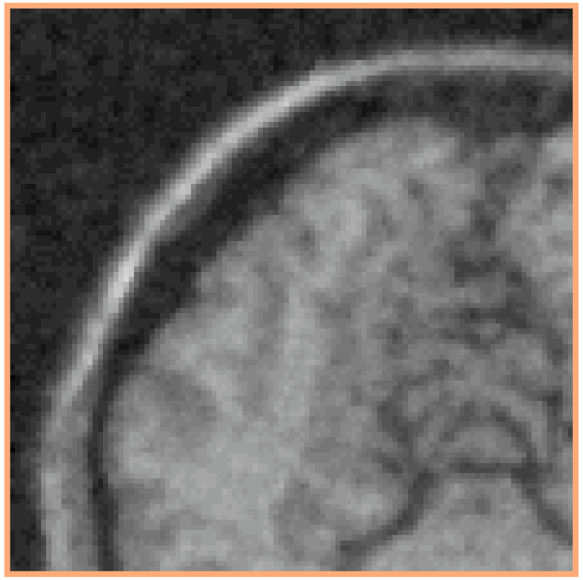
Coherence Enhancing

Curvature Flow

Examples

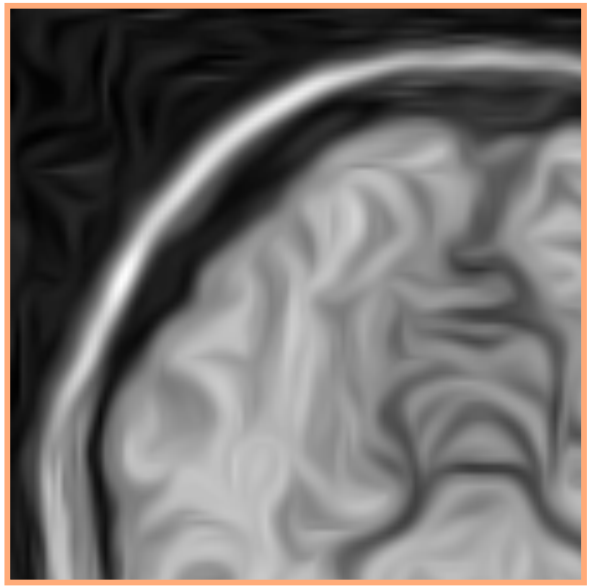
MRI (Simulated noise)

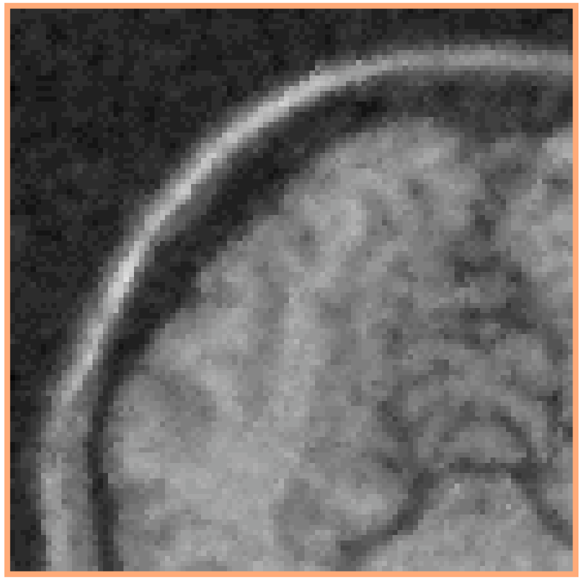




Bilateral Filtering

Coherence Enhancing

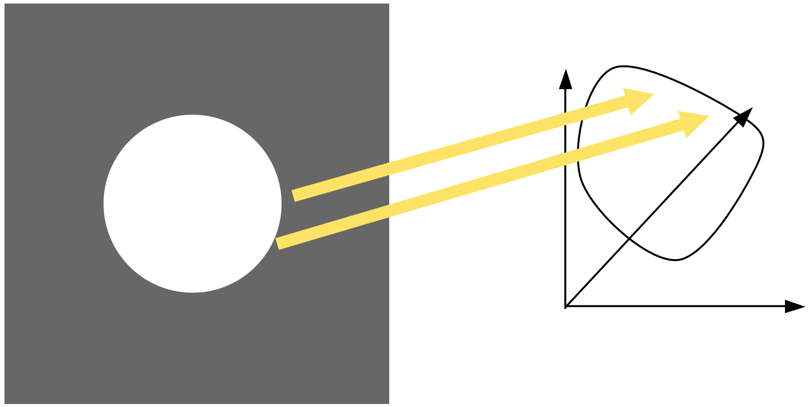


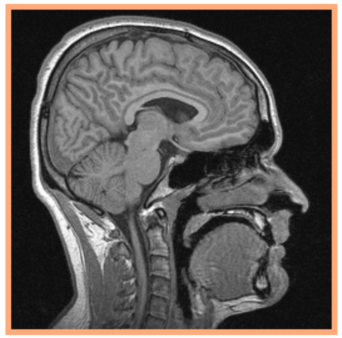


Anisotropic Diffusion

Observations About Images

- Statistics of *natural* images are not *so* random
  - Huang & Mumford (1999)
- But not so simple
  - Manifolds in high-dimensional spaces
  - de Silva & Carlsson (2003)





SCI INSTITUTE

Proposed Strategy

- Infer the appropriate Markovian relationships from the data
  - Images neighborhoods (nhds) as random processes
  - Move away from geometric formulations
- Increase redundancy (functional dependency) of image nhds
  - Information content
  - Entropy
- Optimal posteriori estimates (noise model)

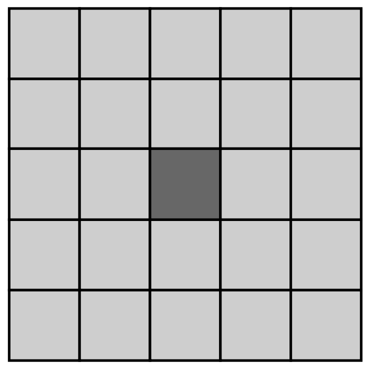
Related Work

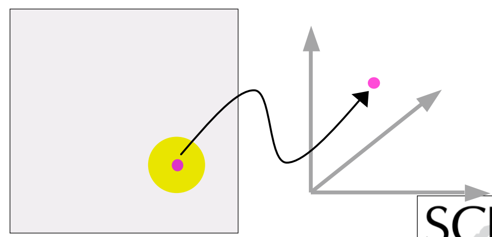
- DUDE algortihm–Weissman et. al (2003)
  - Discrete channels + noise model
  - MLE estimation
- Texture synthesis
  - Efros & Leung (1999)
  - Wei & Levoy (2002)
- NL-means, Baudes et al. (CVPR 2005)
  - Independent, simultaneously presented
  - More later…
- Sparsity in image neighborhoods
  - Roth and Black 2005
  - Elad and Aharon 2006

SCI INSTITUTE

Image Model

- Pixels and neighborhoods Z = (X, Y)
  - P(Z), P(X|Y)
- Scenario
  - Corrupted image -> noise model
  - Prior knowledge P(X|Y)
  - Theorems:
    - Can produce most likely image x′ using P(X|Y = y′)
    - Iterate to produce optimal estimate





Modeling P(Z)

- Set of image neighborhoods
  - Large, complex, high-dimensions
- Approach
  - Represent complexity through examples
  - Nonparametric density estimation

Nonparametric, Multivariate Density Estimation

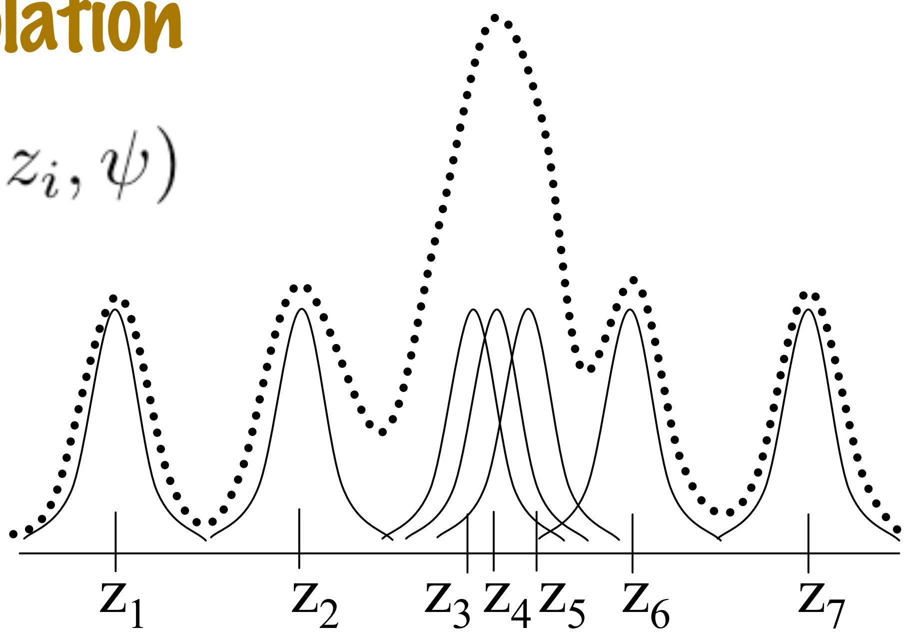
- Nonparametric estimation
  - No prior knowledge of densities
  - Can model *real* densities
- Statistics in higher dimensions
  - Curse of dimensionality (volume of n-sphere -> 0)
  + However, empirically more optimistic
  + Z has identical marginal distributions
  + Lower dimensional manifolds in feature space

SCI INSTITUTE

Parzen Windows (Parzen 1962)

- Scattered-data interpolation

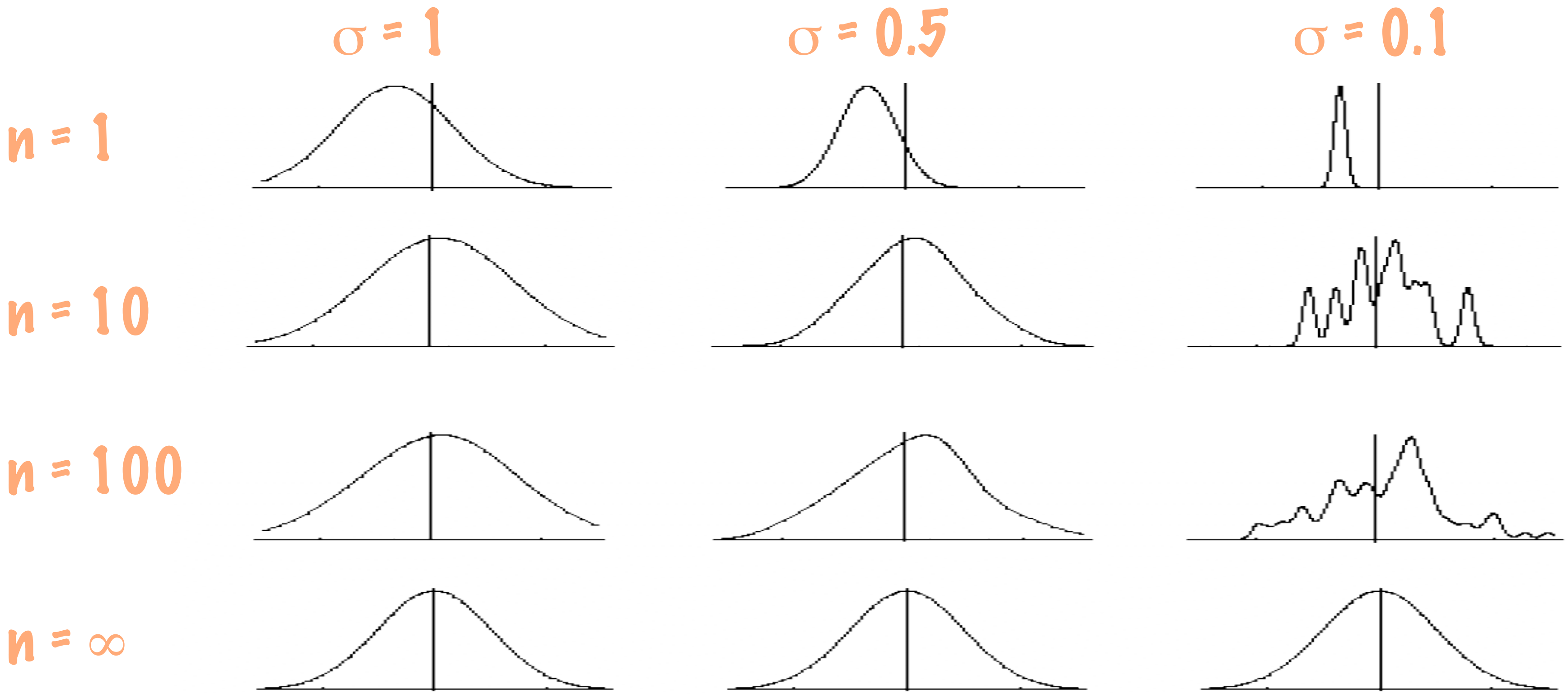
$$p(z) \approx \frac{1}{|A|} \sum_{z_i \in A} G(z - z_i, \psi)$$



- Window function
  - $G \equiv$ Gaussian
  - Covariance matrix: $\psi = \sigma^2 I$

Parzen Windows (Parzen 1962)

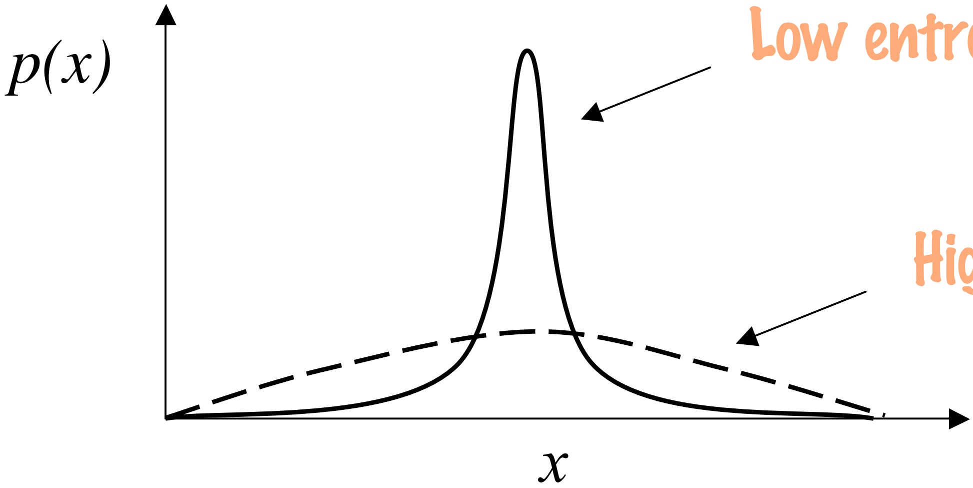
- Effects of finite sampling (Duda & Hart)



Entropy (Shannon 1948)

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

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



SCI INSTITUTE

UINTA Strategy
Awate & Whitaker CVPR 2005, PAMI 2006

- Iterative algorithm
- Progressively minimizes the entropy of image nhds Z = (X, Y)
  - Pixel entropies (X) conditioned on nhd values (Y)
  - Gradient descent (time steps -> mean shift)
- Nonparametric density estimation
  - Stochastic gradient descent

SCI INSTITUTE

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

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

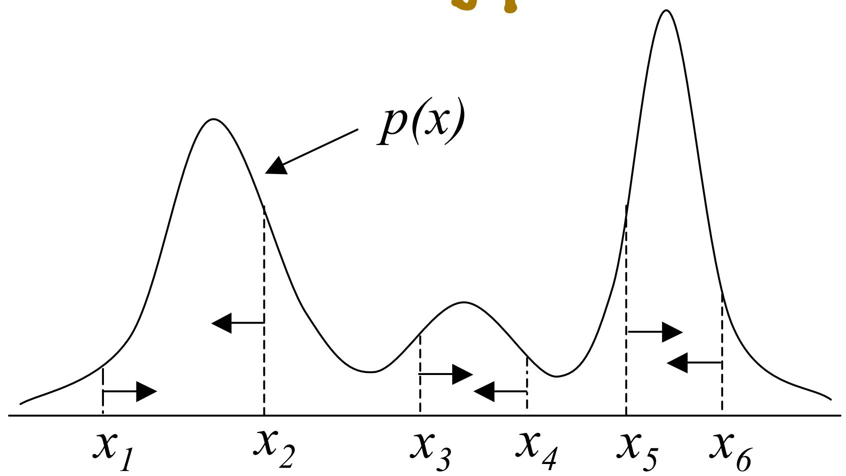
SCI INSTITUTE

Mean-Shift Procedure (Fukunaga et al. 1975)

- Entropy minization <-> mean shift

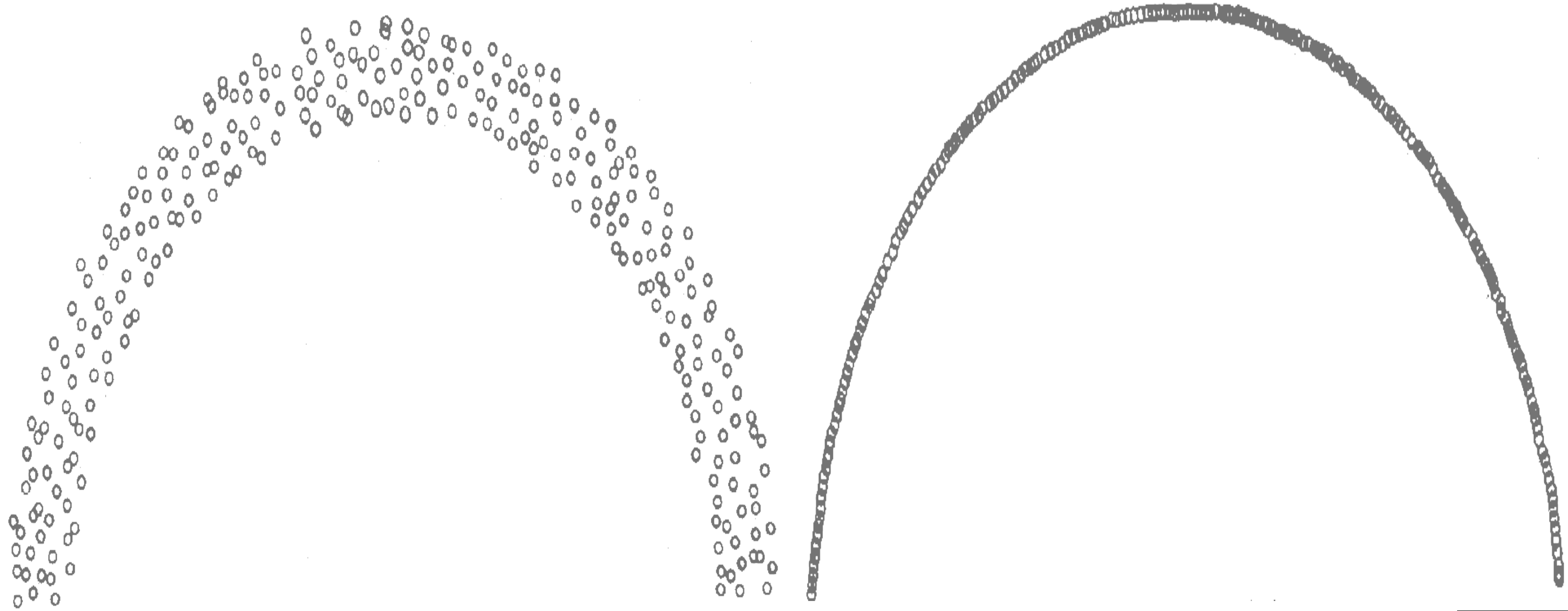
$$\lambda = \Psi|B| \quad x \longleftarrow \sum_j w_j x_j$$

- Mean-shift – a mode seeking procedure



Mean-Shift Procedure (Fukunaga et al. 1975)

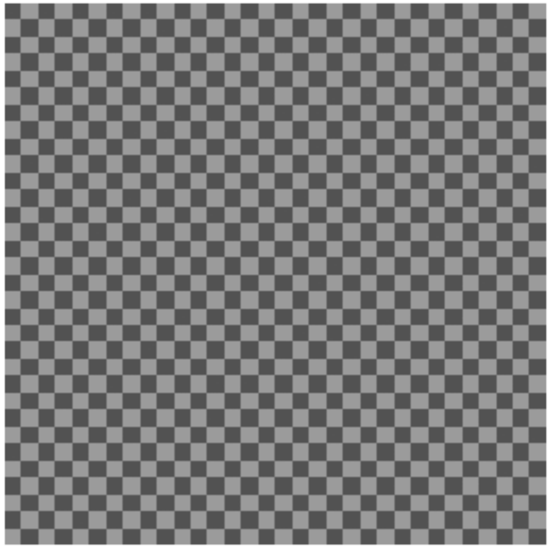
- Data filtering to reduce noise
  - Hand tuned parameters



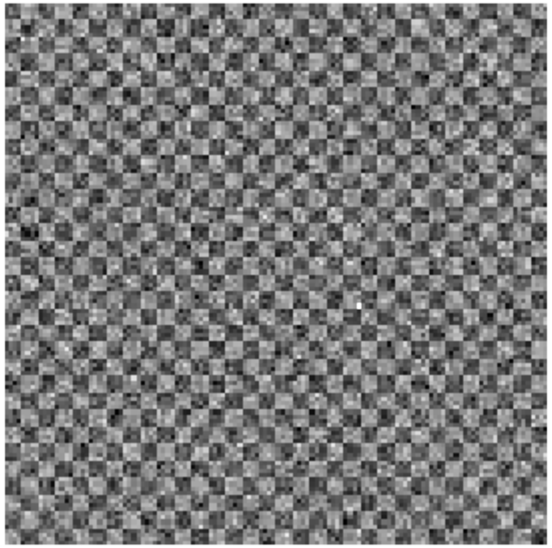
Implementation Issues

- Scale selection for Parzen windowing
  - Automatic – min entropy with cross validation
- Rotational invariance
- Boundary neighborhoods
- Random sample selection – nonstationary image statistics
- Stopping criteria

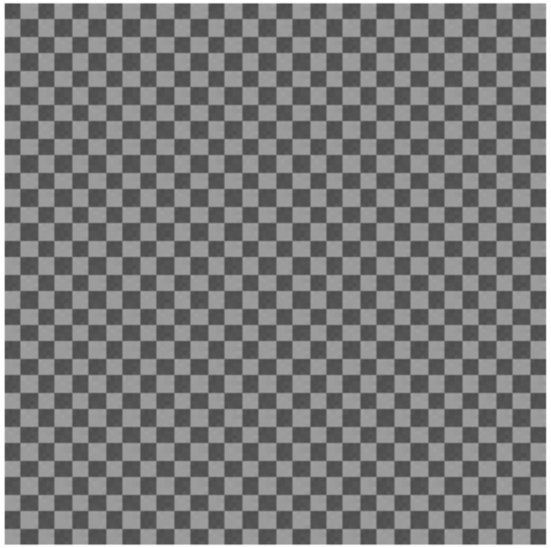
Results



Original



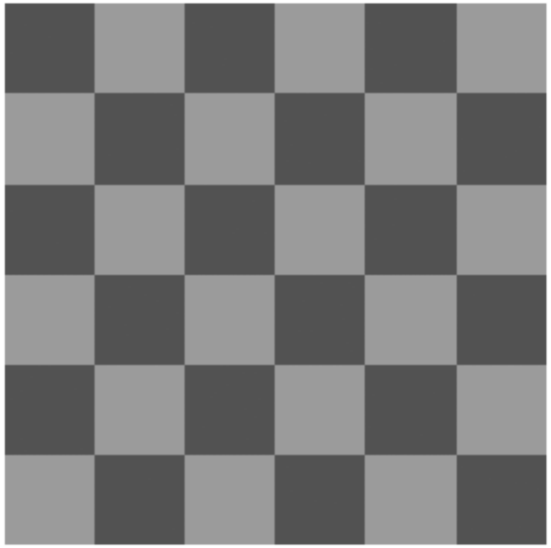
Noisy



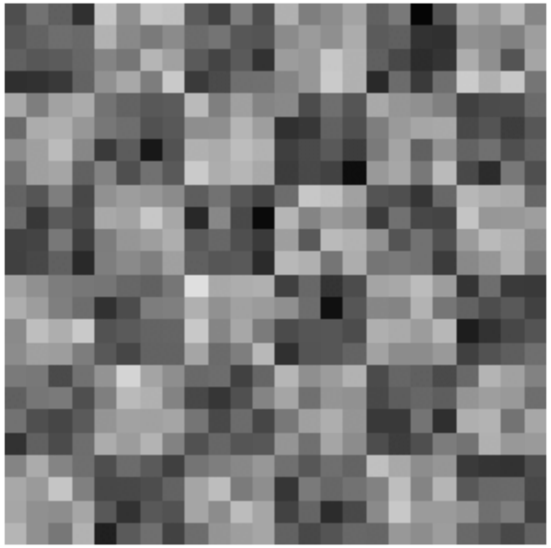
Filtered

SCI INSTITUTE

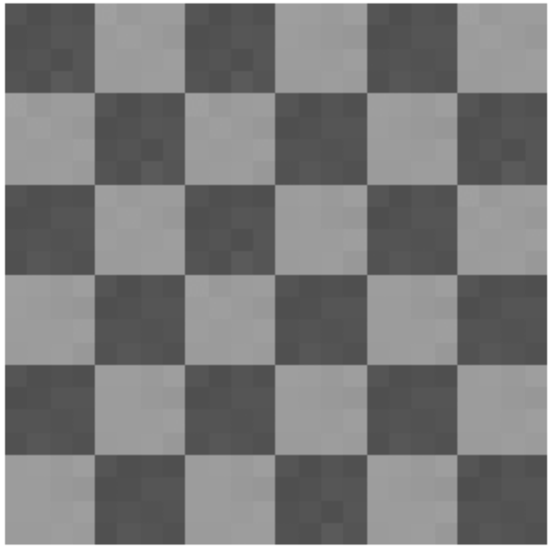
Checkerboard With Noise



Original



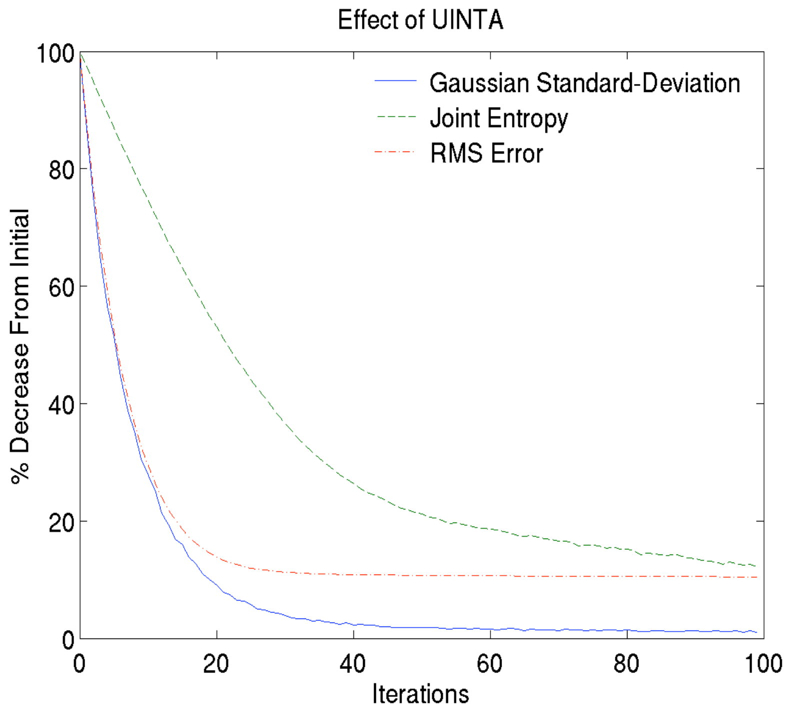
Noisy

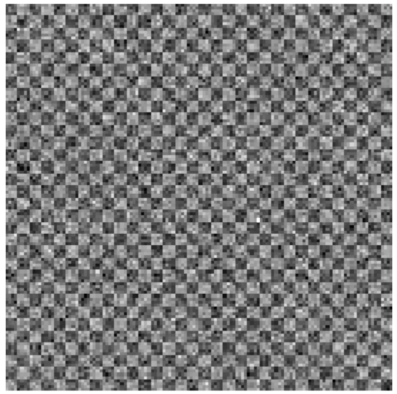


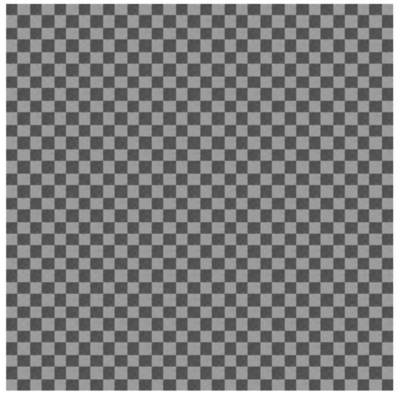
Filtered

Quality of Denoising

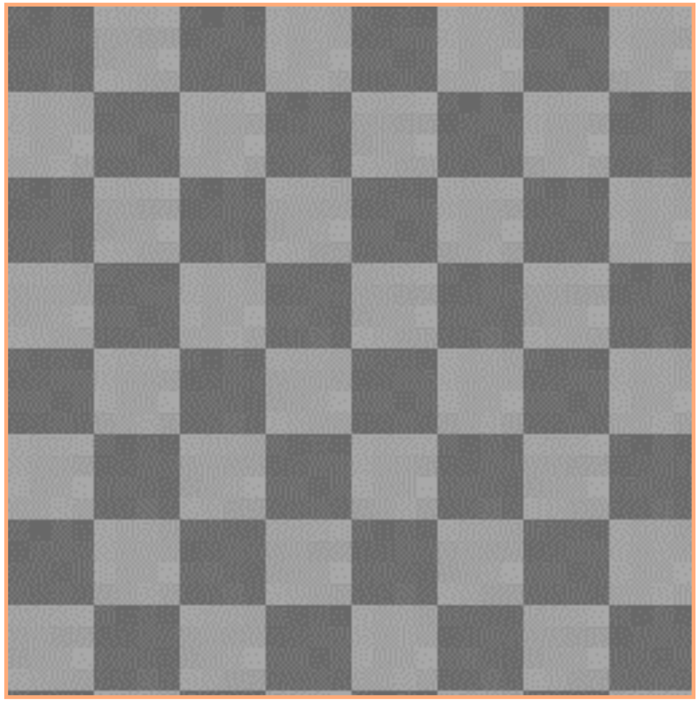
- σ, joint entropy, and RMS- error vs. number of iterations

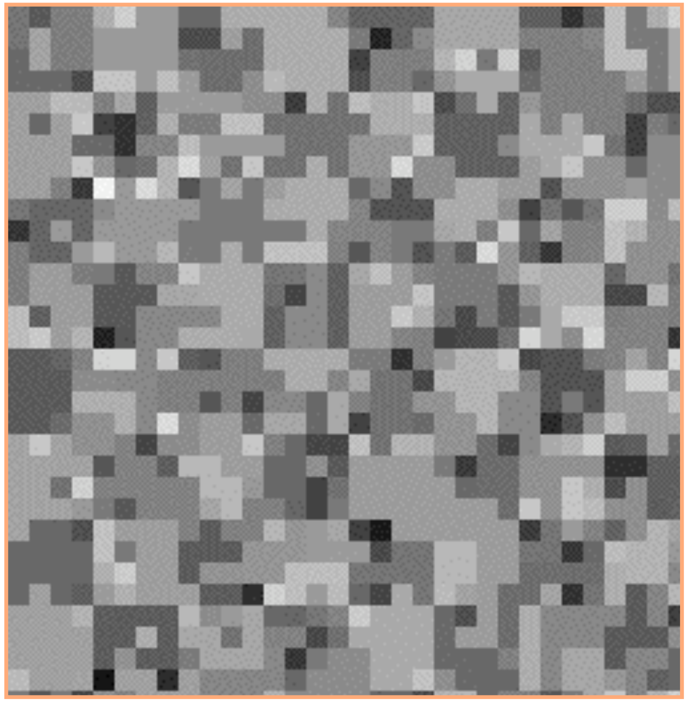




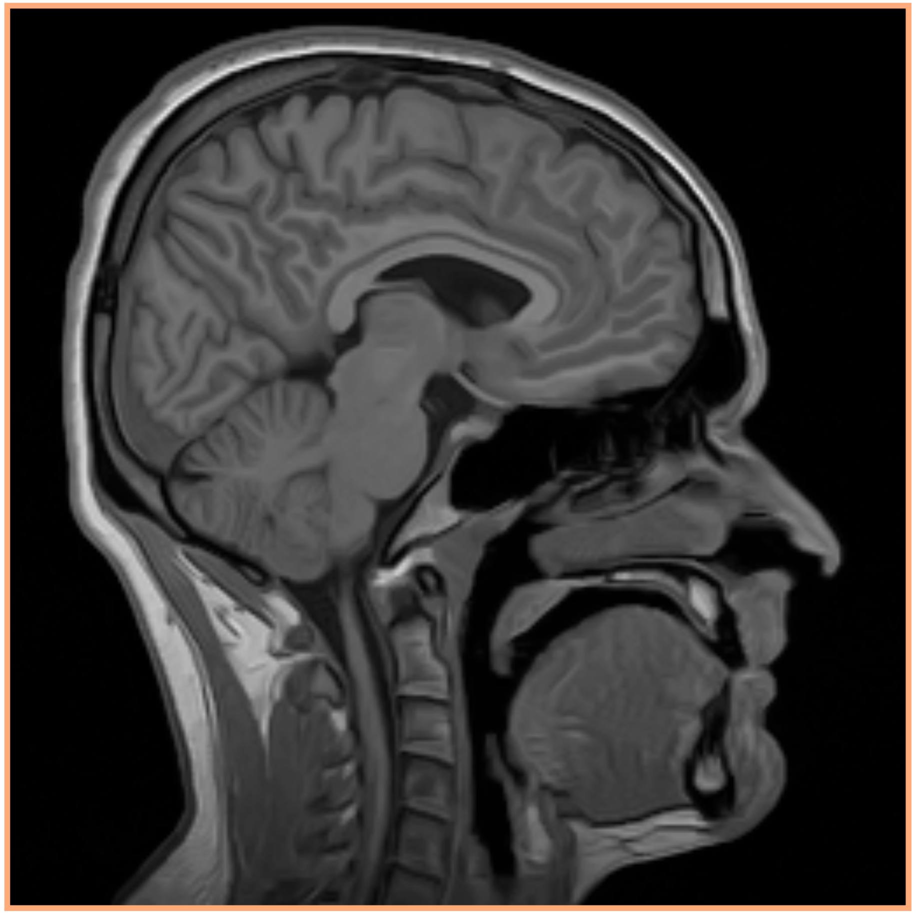


Vs Perona Malik



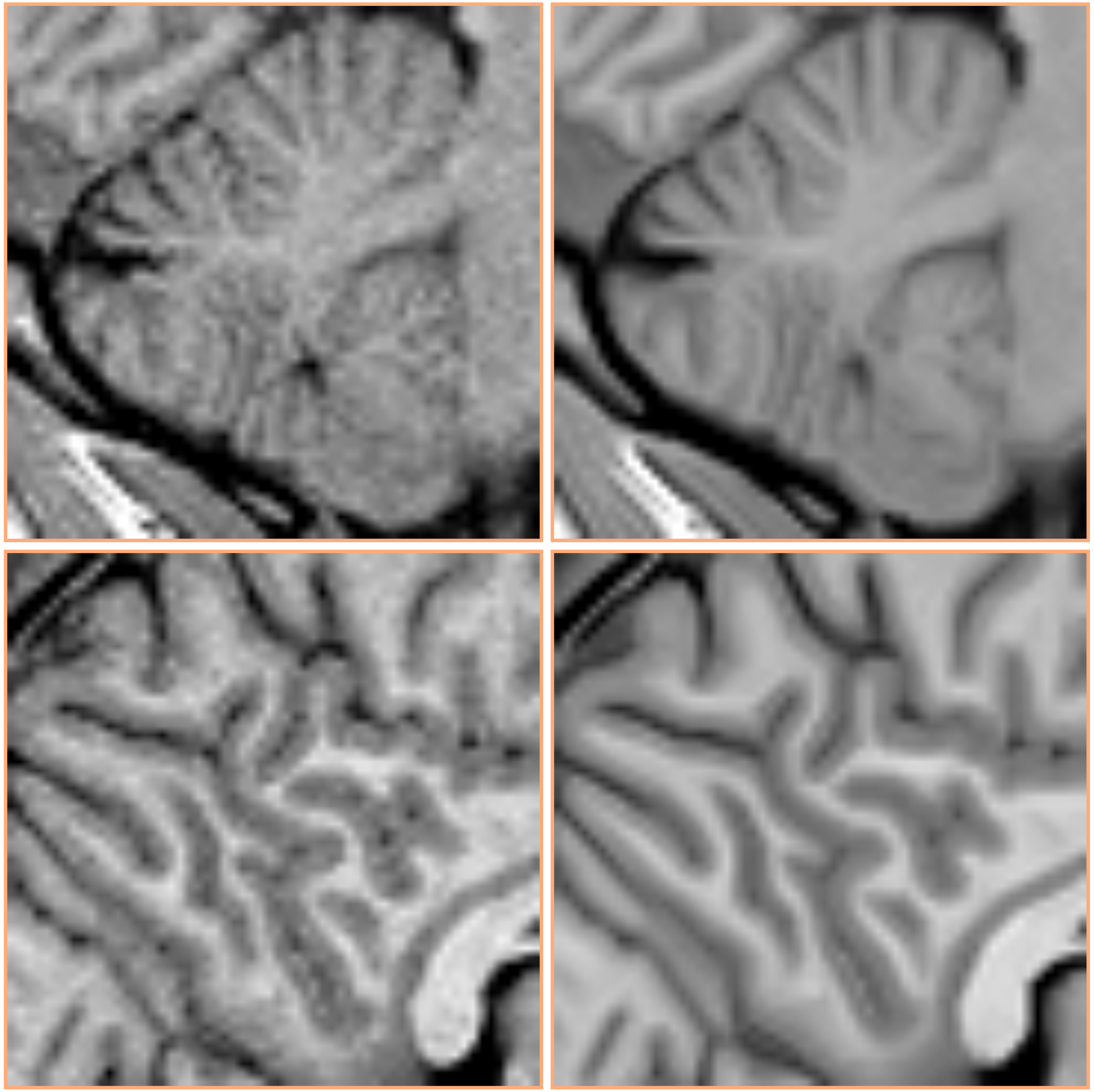


MRI Head

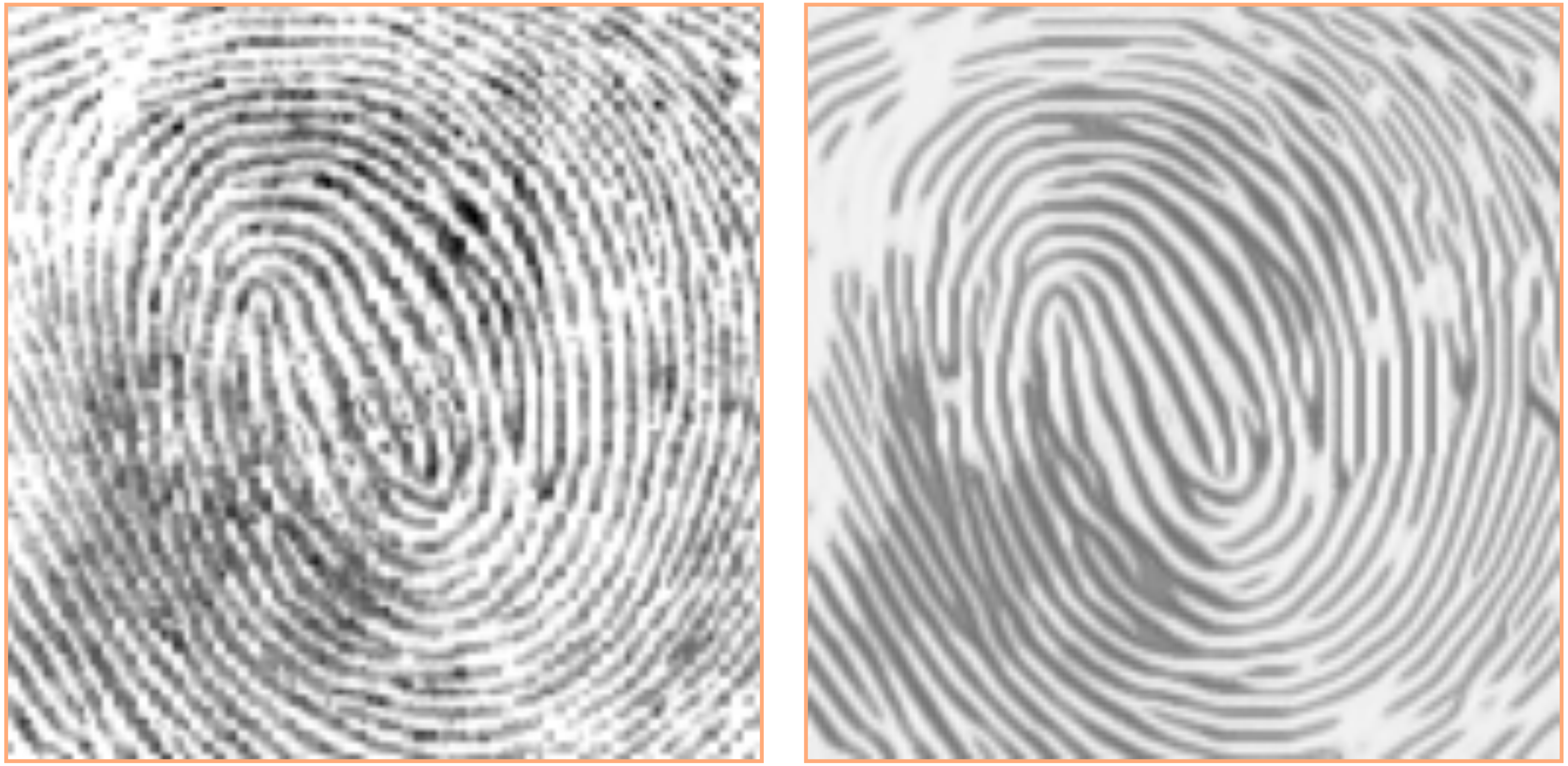


SCI INSTITUTE

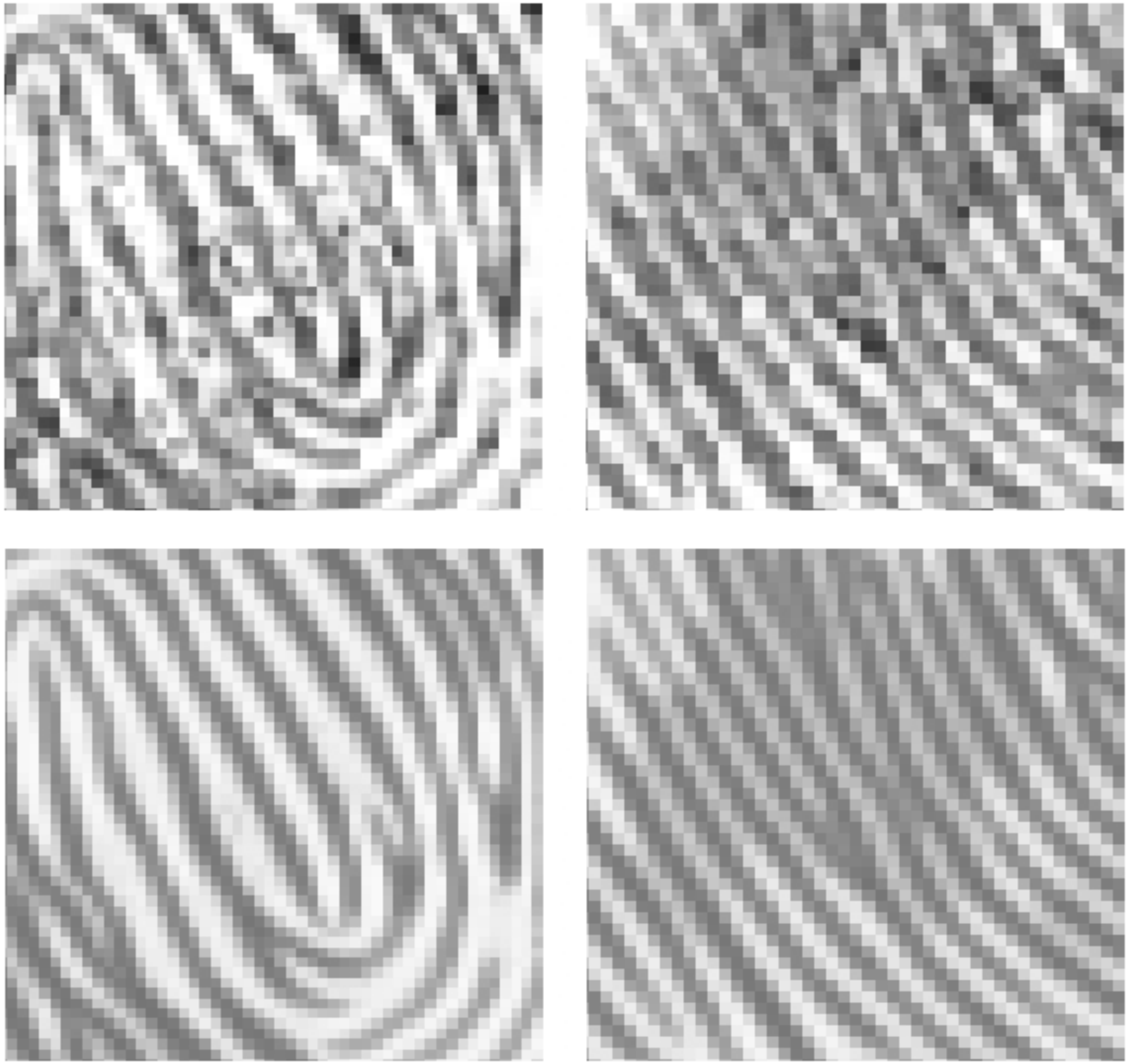
MRI Head



Fingerprint

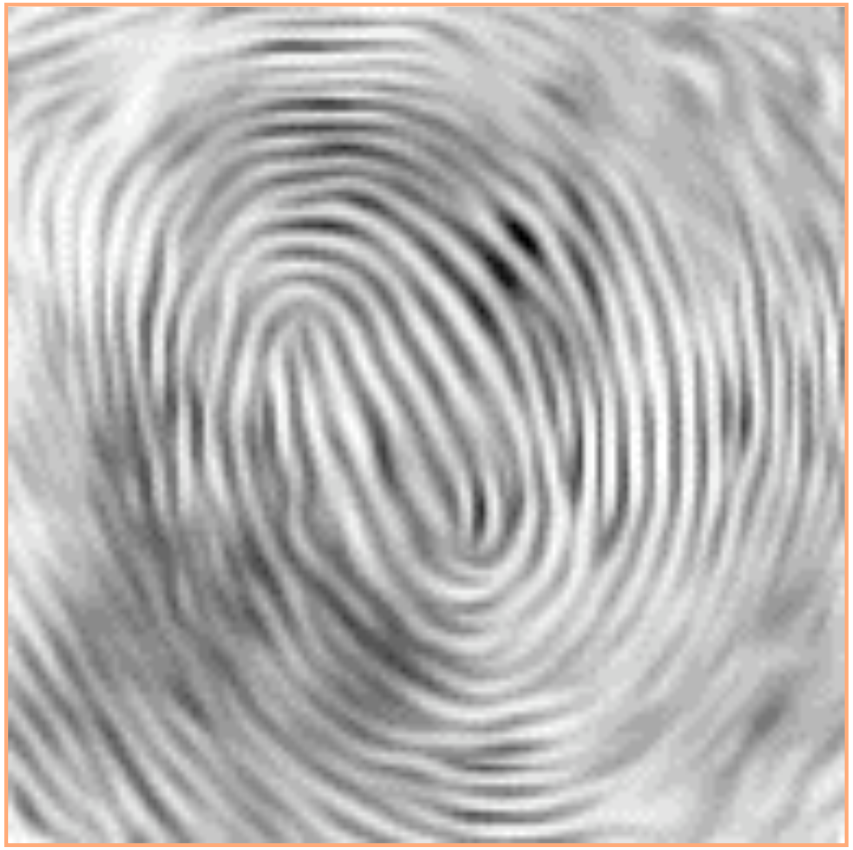


Fingerprint



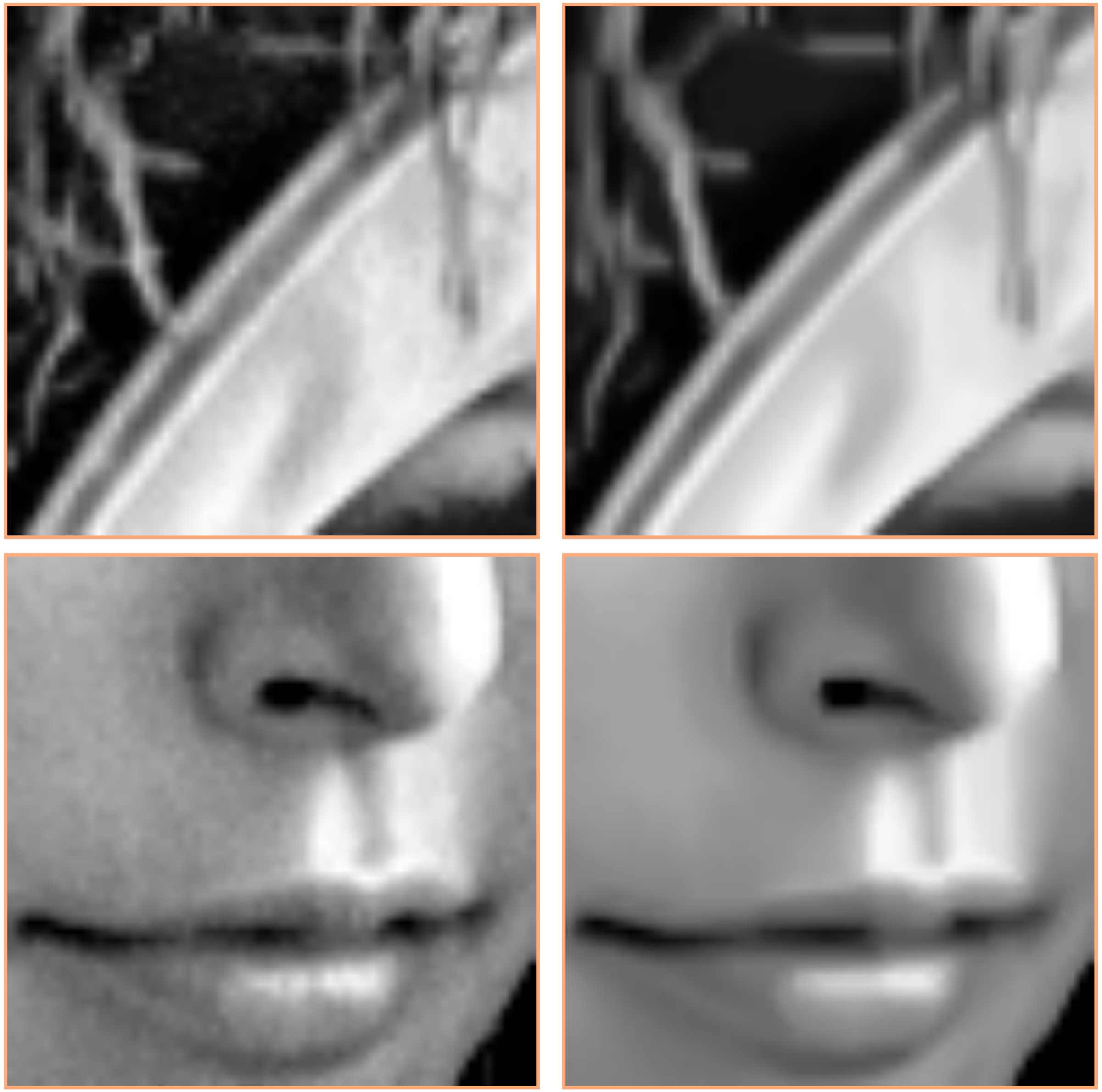
Vs Perona Malik

Vs Coherence Enhancing

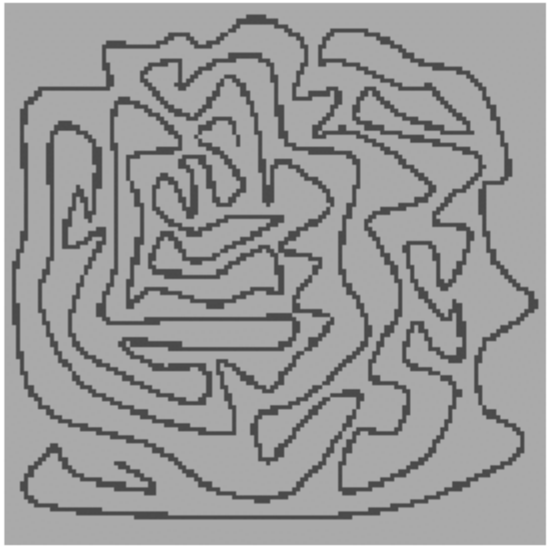


Lena

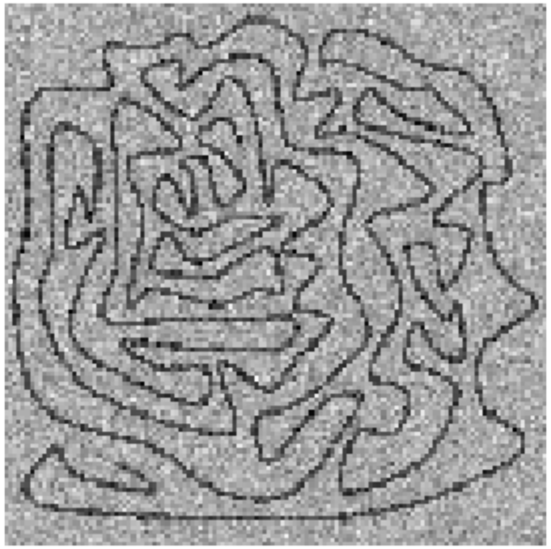
Lena



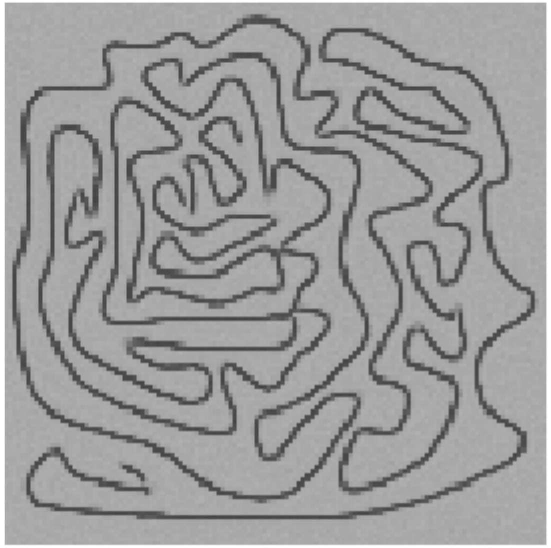
Results



Original



Noisy

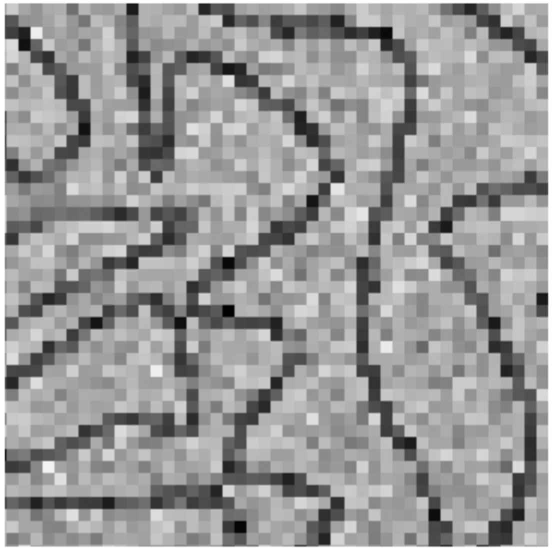


Filtered

SCI INSTITUTE

Results

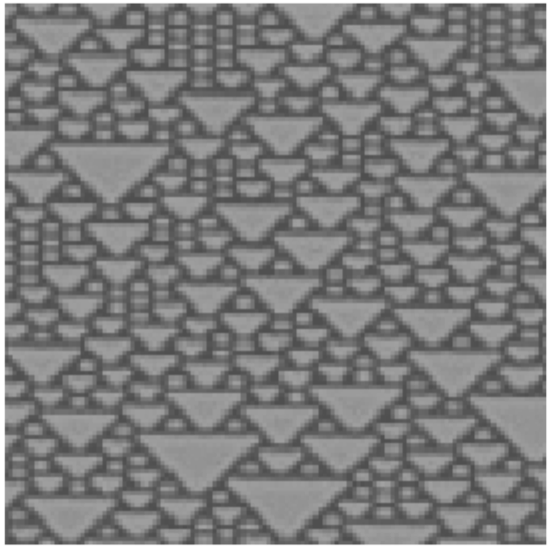
Original



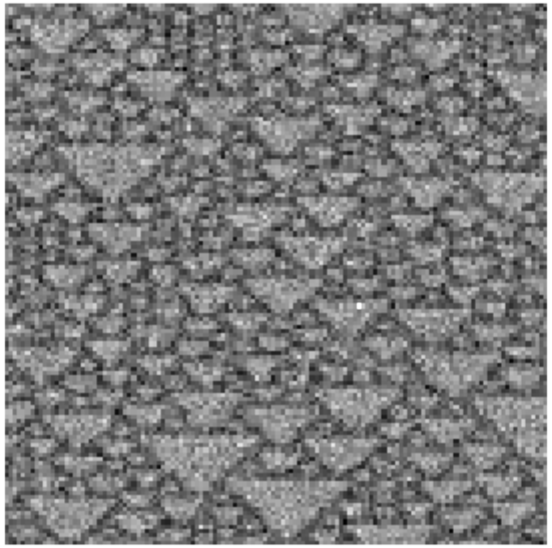
Noisy

Filtered

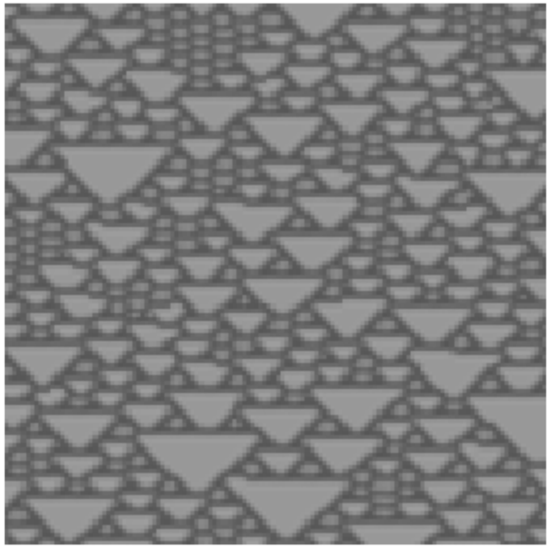
Results



Original

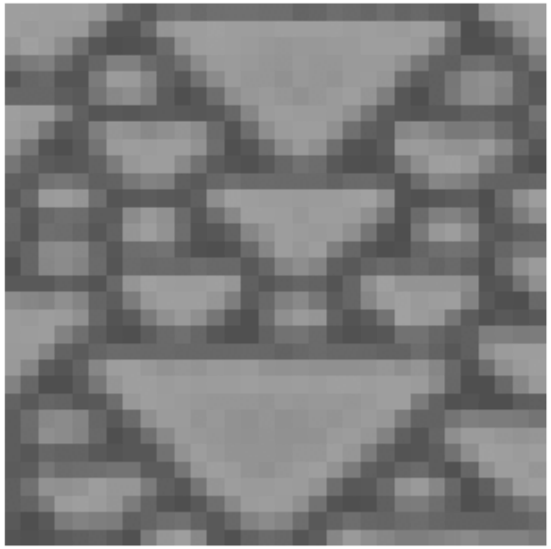


Noisy

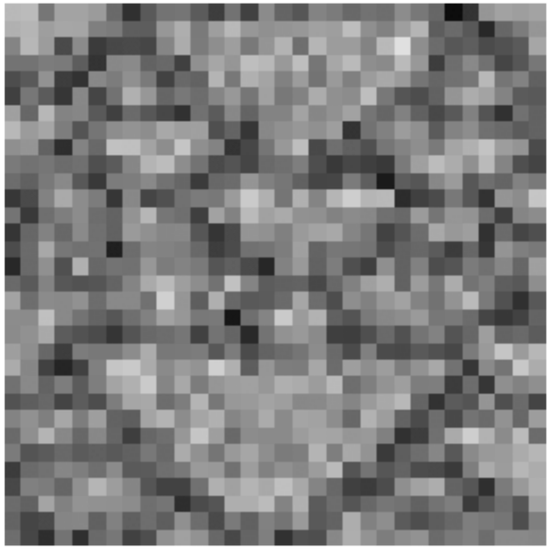


Filtered

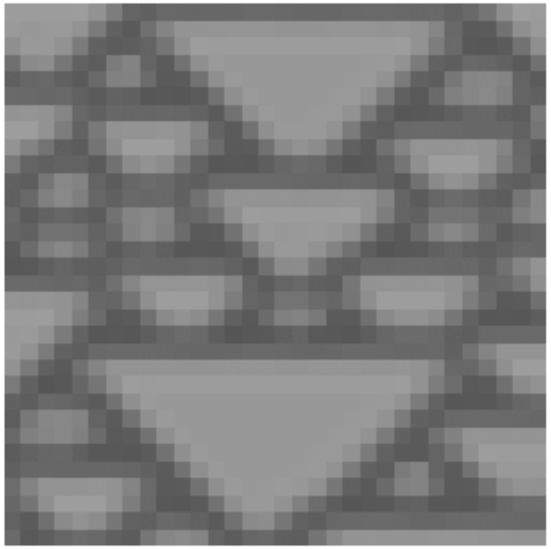
Fractal



Original

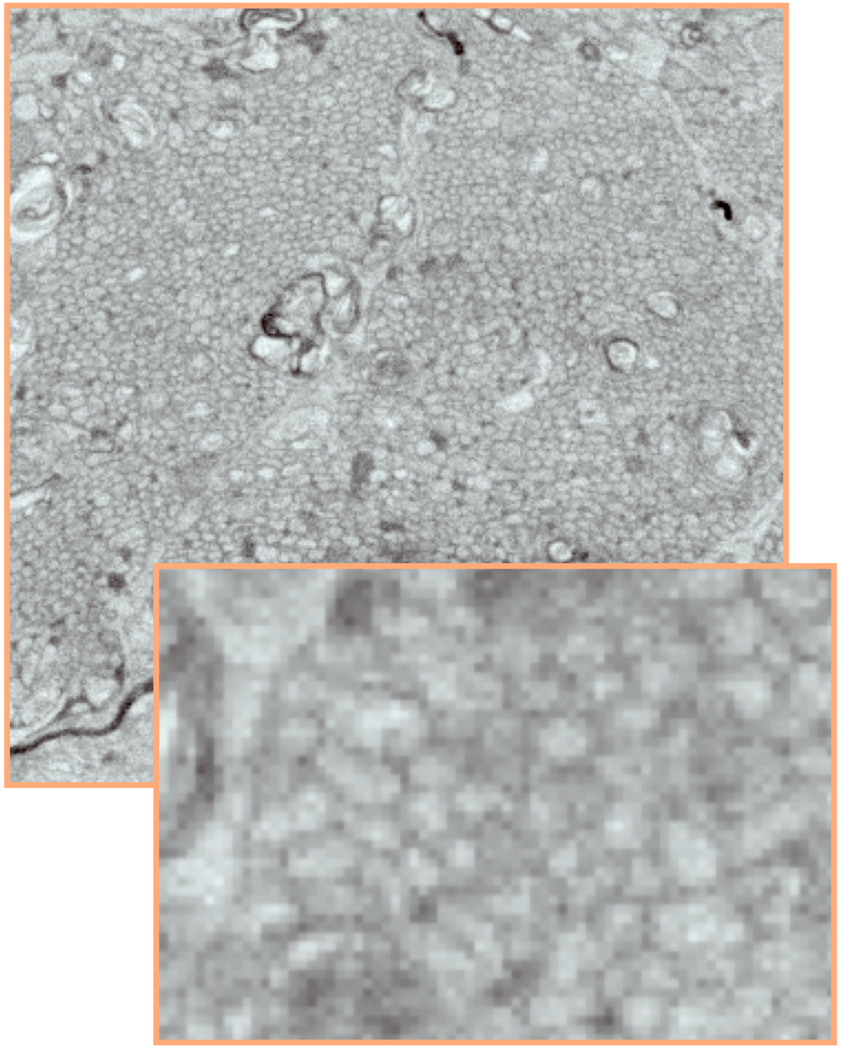


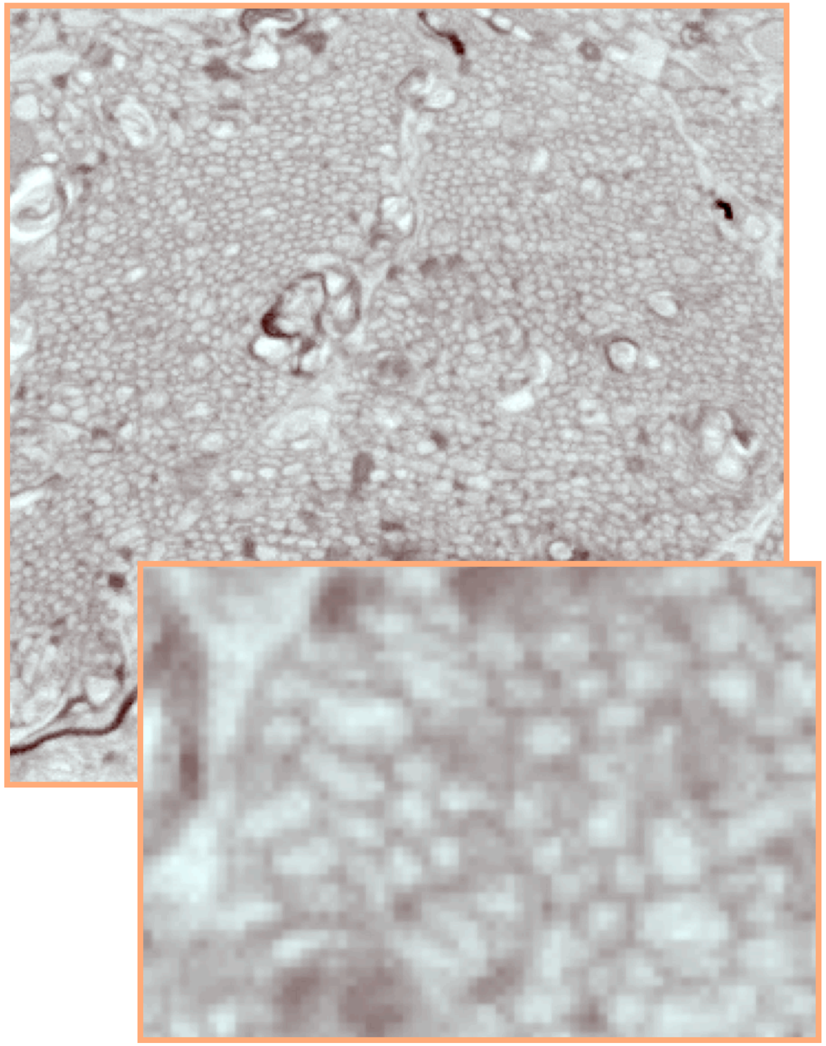
Noisy



Filtered

Microscopy

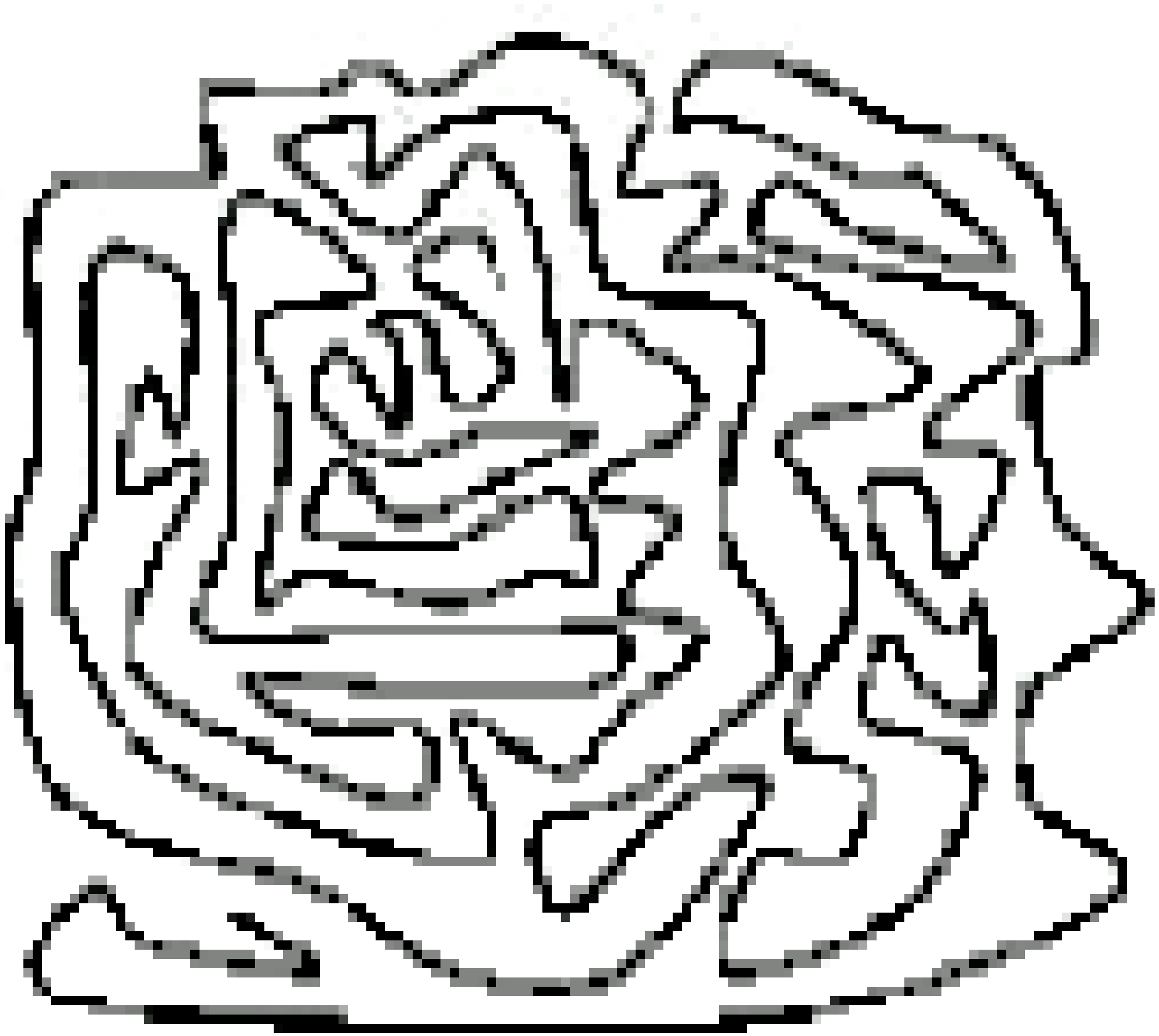




Quantitative Results

- Generalizes well
  - Relatively insensitive to a few parameters (e.g. nhd size)
- Compares favorably with s.o.t.a. wavelet denoisers
  - Close but worse for standard images (photographs)
  - Better for less typical images (defy wavelet shrinkage assumptions)
- Spectral data -> gets even better

Entropy Scale Space?

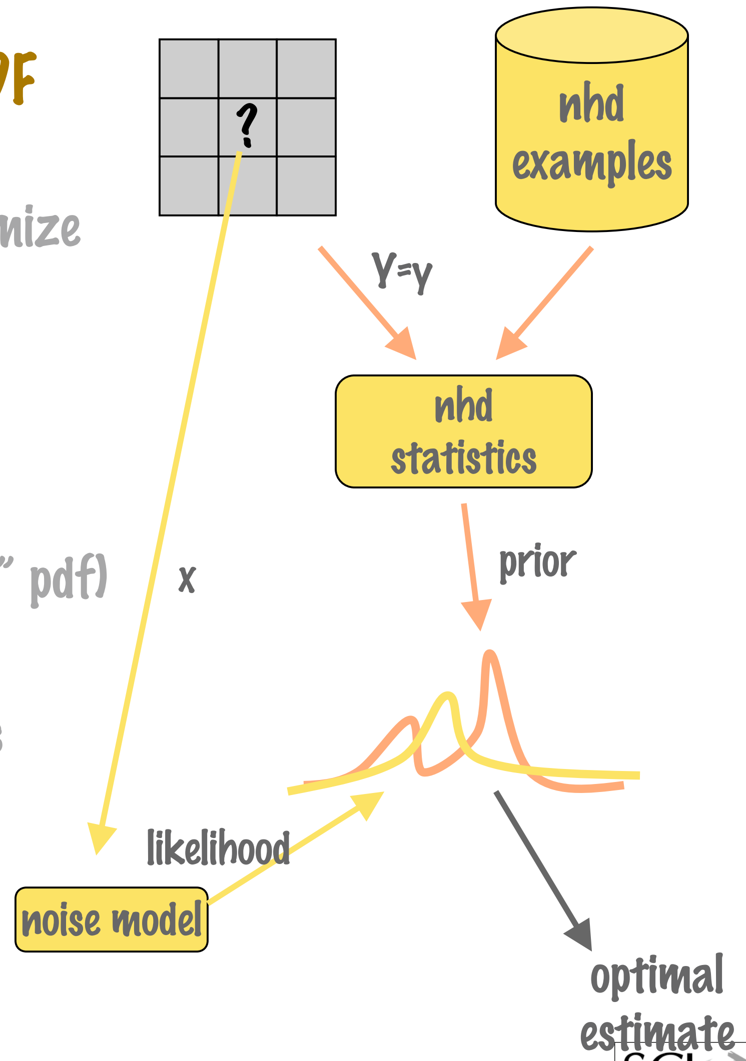


Other Applications

- Optimal reconstruction
  - Noise model
  - Awate&Whitaker, IPMI, 2005
- MRI head segmentation
  - Iterative tissue classification
  - Tasdizen et al., MICCAI, 2005
- Texture segmentation
  - Awate, Tasdizen, Whitaker, ECCV 2005

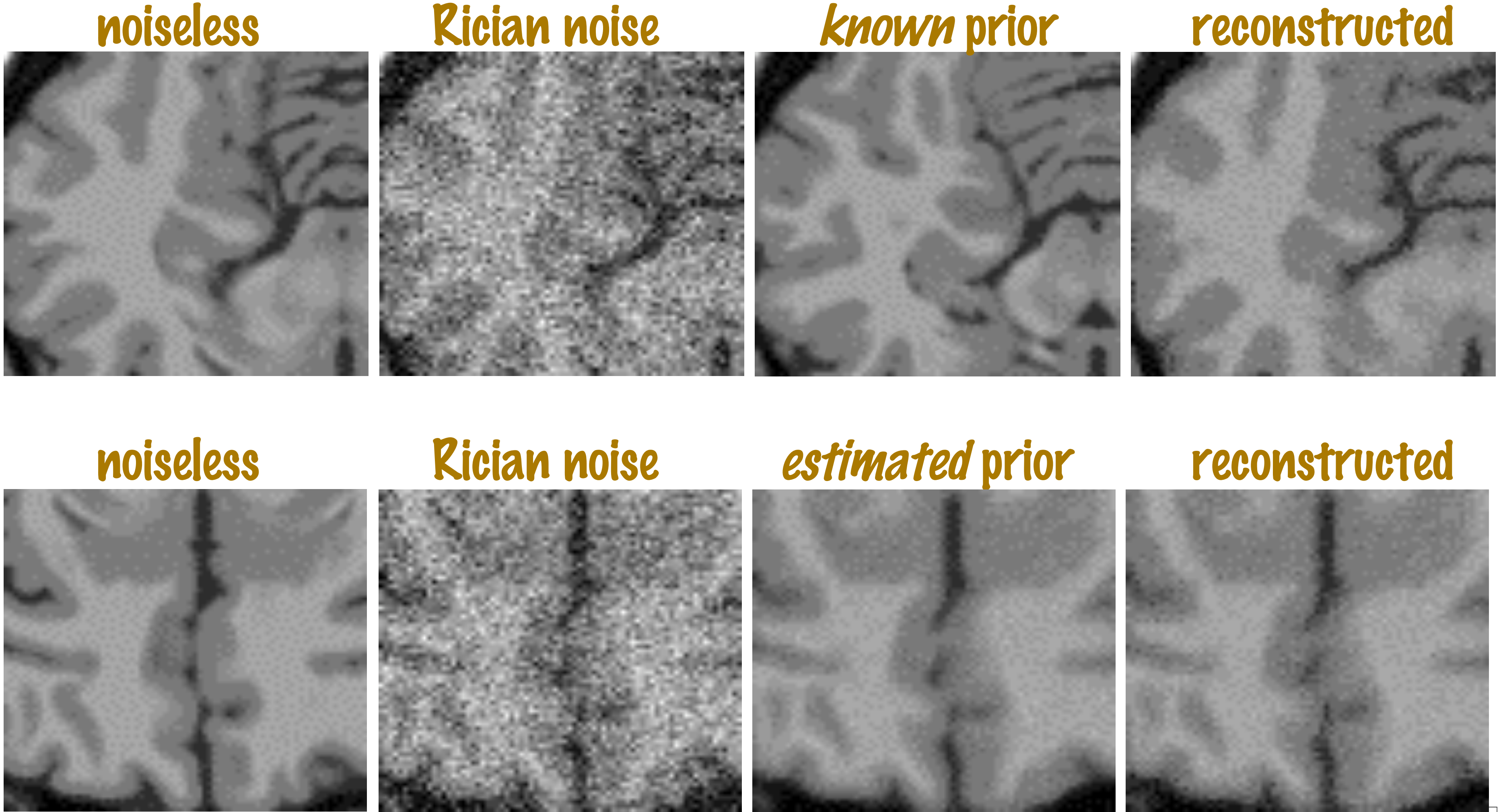
Optimal Reconstruction

- What if we had a noise model and a PDF conditioned on image nhds?
  - -> "Optimal" estimate for each pixel (minimize expected error)
- Image statistics (each nhd forms a "lookup")
  - Database of "perfect" image nhds
  - Bootstrap from the noisy images ("denoise" pdf)
- Noisy neighborhoods
  - Iterate on sequence of improving estimates



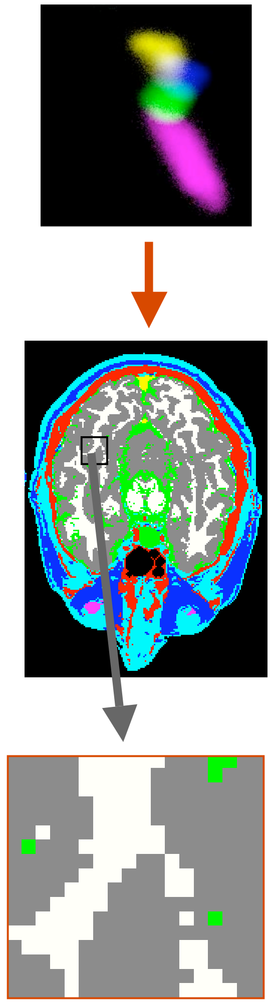
SCI INSTITUTE

Optimal Estimation (MRI)



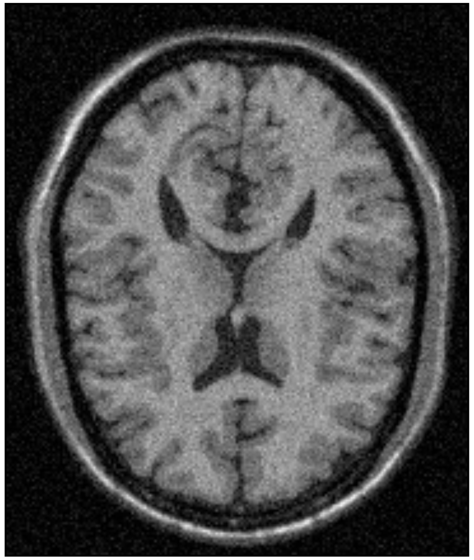
MRI Tissue Classification

- Classify pixels based on spectral (multi-modal) MRI measurements
- Overlapping (noisy) clusters
  - ambiguity and misclassified pixels
- Use spatial data to influence decision
  - Large-scale (absolute) relationships <-> statistical atlases
  - Local (nhd) relationships <-> Markov random fields (smooth configurations)
- Idea: *learn* nhd relationships from data
  - Classify (iterative) to reduce in-class nhd entropy

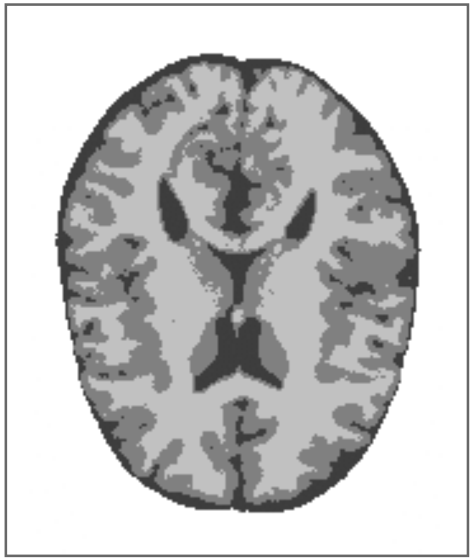


MRI Tissue Classification

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

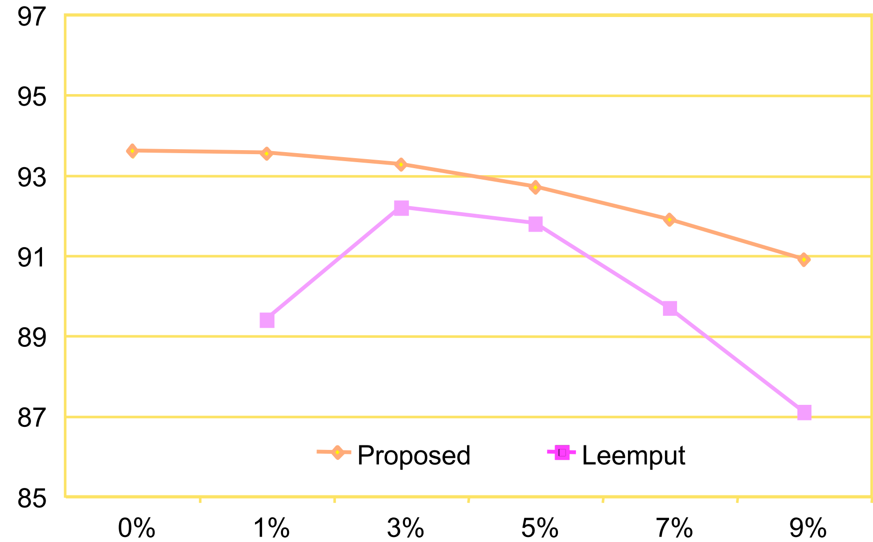


MRI Input



GM, WM, CSF Seg.

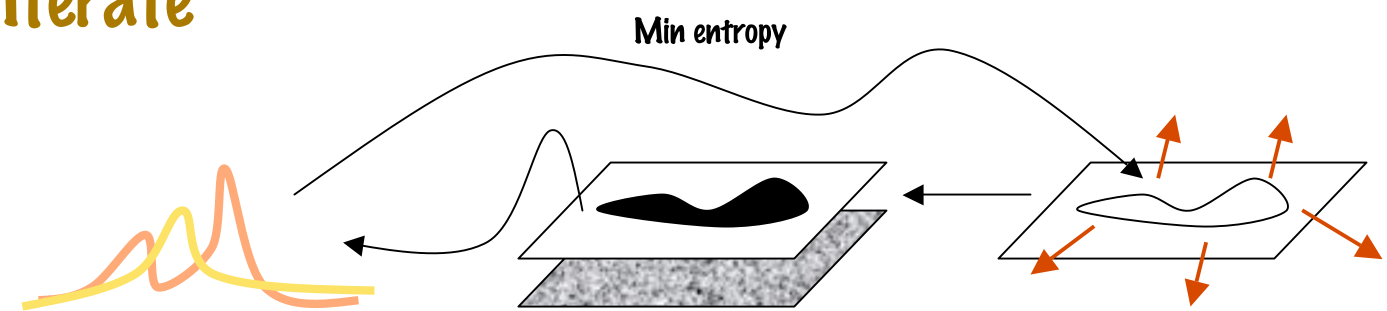
GM Classification Performance vs Noise Level



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

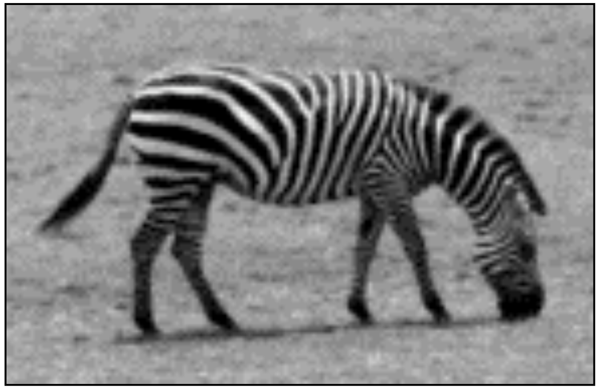
Texture Segmentation

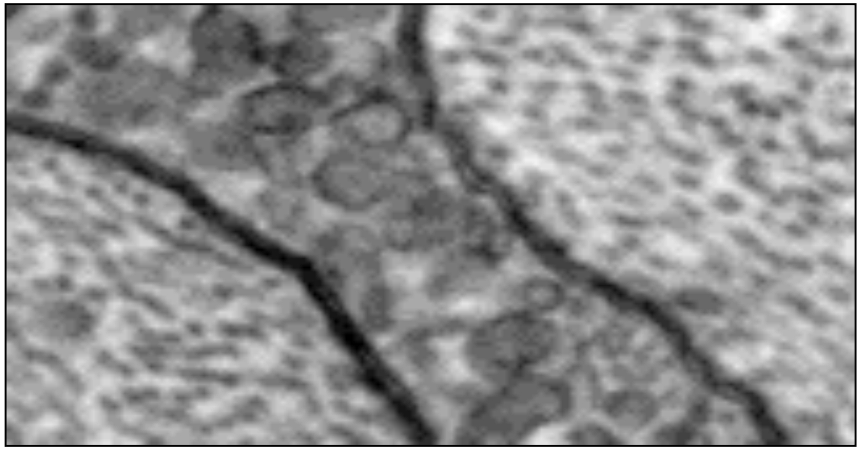
- Reassign class labels to reduce in-class entropy
  - Deformable model to keep spatial coherence
- Recompute pdfs from new class labels
  - Random samples + nonparametric nhd statistics
- Iterate



Texture Segmentation
Awate et al., 2005

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





Thanks

- Sponsors (NSF, NIH)
- Team: S. Awate, T. Tasdizen, N. Foster

SCI INSTITUTE

Thoughts