Markov Random Fields and Stochastic Image Models

Charles A. Bouman
School of Electrical and Computer Engineering
Purdue University
Phone: (317) 494-0340
Fax: (317) 494-3358
email bouman@ecn.purdue.edu
Available from: http://dynamo.ecn.purdue.edu/~bouman/

Tutorial Presented at:
1995 IEEE International Conference
on Image Processing
23-26 October 1995
Washington, D.C.

Special thanks to:

Ken Sauer
Department of Electrical Engineering
University of Notre Dame

Suhail Saquib
School of Electrical and Computer Engineering
Purdue University

Overview of Topics

1. Introduction
2. The Bayesian Approach
3. Discrete Models
   (a) Markov Chains
   (b) Markov Random Fields (MRF)
   (c) Simulation
   (d) Parameter estimation
4. Application of MRF's to Segmentation
   (a) The Model
   (b) Bayesian Estimation
   (c) MAP Optimization
   (d) Parameter Estimation
   (e) Other Approaches
5. Continuous Models
   (a) Gaussian Random Process Models
      i. Autoregressive (AR) models
      ii. Simultaneous AR (SAR) models
      iii. Gaussian MRF's
      iv. Generalization to 2-D
   (b) Non-Gaussian MRF's
      i. Quadratic functions
      ii. Non-Convex functions
      iii. Continuous MAP estimation
      iv. Convex functions
   (c) Parameter Estimation
      i. Estimation of σ
      ii. Estimation of T and p parameters
6. Application to Tomography
   (a) Tomographic system and data models
   (b) MAP Optimization
   (c) Parameter estimation
7. Multiscale Stochastic Models
   (a) Continuous models
   (b) Discrete models
8. High Level Image Models

References in Statistical Image Modeling

1. Overview references [100, 89, 50, 54, 162, 4, 44]
2. Type of Random Field Model
   (a) Discrete Models
      i. Hidden Markov models [134, 135]
      ii. Markov Chains [41, 42, 156, 132]
      iii. Ising model [127, 126, 122, 130, 100, 131]
      iv. Discrete MRF [13, 14, 160, 48, 161, 47, 16, 169, 36, 51, 49, 116, 167, 99, 50, 72, 104, 157, 55, 181, 121, 123, 23, 91, 176, 92, 37, 125, 128, 140, 168, 97, 119, 11, 39, 77, 172, 93]
      v. MRF with Line Processes[68, 53, 177, 175, 178, 173, 171]
   (b) Continuous Models
      i. AR and Simultaneous AR [95, 94, 115]
      ii. Gaussian MRF [18, 15, 87, 95, 94, 33, 114, 153, 38, 106, 147]
      iii. Nonconvex potential functions [70, 71, 21, 81, 107, 66, 32, 143]
      iv. Convex potential functions [17, 75, 107, 108, 155, 24, 146, 90, 25, 27, 32, 149, 26, 148, 150]
3. Regularization approaches
   (a) Quadratic [165, 158, 137, 102, 103, 98, 138, 60]
   (b) Nonconvex [139, 85, 88, 159, 19, 20]
   (c) Convex [155, 3]
4. Simulation and Stochastic Optimization Methods [118, 80, 129, 100, 68, 141, 61, 76, 62, 63]
5. Computational Methods used with MRF Models
   (a) Simulation based estimators [116, 157, 55, 39, 26]
   (b) Discrete optimization
      i. Simulated annealing [68, 167, 55, 181]
      ii. Recursive optimization [48, 49, 169, 156, 91, 172, 173, 93]
      iii. Greedy optimization [160, 16, 161, 36, 51, 104, 157, 55, 125, 92]
      iv. Multiscale optimization [22, 72, 23, 128, 97, 105, 110]
      v. Mean field theory [176, 177, 175, 178, 171]
   (c) Continuous optimization
      i. Simulated annealing [153]
      ii. Gradient ascent [87, 149, 150]
      iii. Conjugate gradient [10]
      iv. EM [70, 71, 81, 107, 75, 82]
      v. ICM/Gauss-Seidel/ICD [24, 146, 147, 25, 27]
      vi. Continuation methods [19, 20, 153, 143]
6. Parameter Estimation
   (a) For MRF
      i. Discrete MRF
         A. Maximum likelihood [130, 64, 71, 131, 121, 108]

B. Coding/maximum pseudolikelihood [15, 16, 18, 69, 104]
C. Least squares [49, 77]

ii. Continuous MRF
A. Gaussian [95, 94, 33, 114, 38, 115, 106]
B. Non-Gaussian [124, 148, 26, 133, 145, 144]

iii. EM based [71, 176, 177, 39, 180, 26, 178, 133, 145, 144]

(b) For other models
i. EM algorithm for HMM's and mixture models [9, 8, 46, 170, 136, 1]
ii. Order identification [2, 94, 142, 37, 179, 180]

7. Application

(a) Texture classification [95, 33, 38, 115]
(b) Texture modeling [56, 94]
(c) Segmentation remotely sensed imagery [160, 161, 99, 181, 140, 28, 92, 29]
(d) Segmentation of documents [157, 55]
(e) Segmentation (nonspecific) [48, 16, 47, 36, 49, 51, 167, 116, 96, 104, 114, 23, 37, 115, 125, 168, 97, 120, 11, 39, 110, 180, 172, 93]
(f) Boundary and edge detection [41, 42, 57, 156, 65, 175]
(g) Image restoration [87, 68, 169, 96, 153, 91, 90, 177, 82, 150]
(h) Image interpolation [149]
(i) Optical flow estimation [88, 101, 83, 111, 143, 178]
(j) Texture modeling [95, 94, 44, 33, 38, 123, 115, 112, 56, 113]
(k) Tomography [79, 70, 71, 81, 75, 107, 108, 24, 146, 25, 147, 32, 26, 27]
(l) Crystallography [53]
(m) Template matching [166, 154]
(n) Image interpretation [119]

8. Multiscale Bayesian Models

(a) Discrete model [28, 29, 40, 154]
(b) Continuous model [12, 34, 5, 6, 7, 112, 35, 52, 111, 113, 166]
(c) Parameter estimation [34, 29, 166, 154]

9. Multigrid techniques [78, 30, 31, 117, 58]

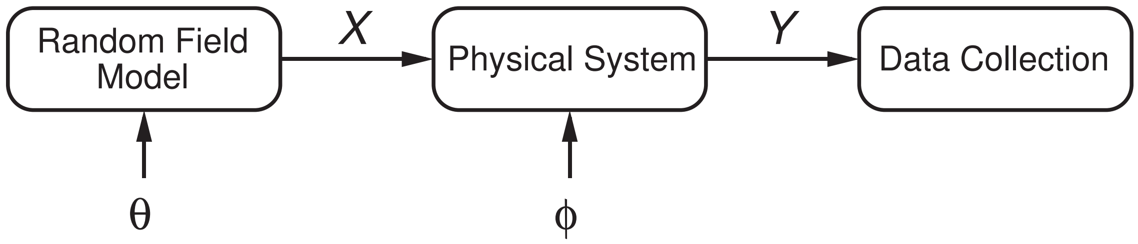
The Bayesian Approach

θ - Random field model parameters

X - Unknown image

ϕ - Physical system model parameters

Y - Observed data

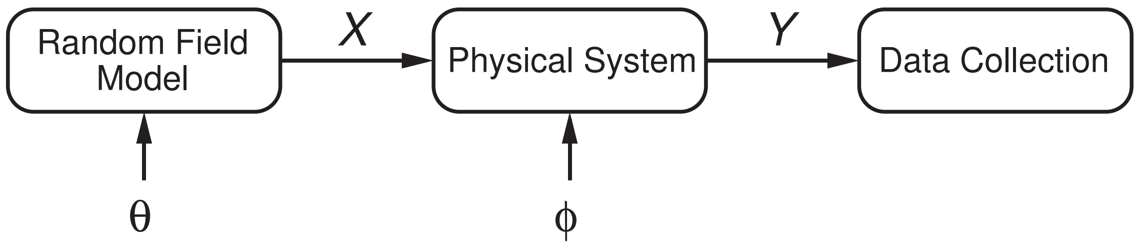


- Random field may model:
  - Achromatic/color/multispectral image
  - Image of discrete pixel classifications
  - Model of object cross-section
- Physical system may model:
  - Optics of image scanner
  - Spectral reflectivity of ground covers (remote sensing)
  - Tomographic data collection

Bayesian Versus Frequentist?

- How does the Bayesian approach differ?
  - Bayesian makes assumptions about prior behavior.
  - Bayesian requires that you choose a model.
  - A **good** prior model can improve accuracy.
  - But model mismatch can impair accuracy
- When should you use the frequentist approach?
  - When (# of data samples)>>(# of unknowns).
  - When an accurate prior model does not exist.
  - When prior model is not needed.
- When should you use the Bayesian approach?
  - When (# of data samples)≈(# of unknowns).
  - When model mismatch is tolerable.
  - When accuracy without prior is poor.

Examples of Bayesian Versus Frequentist?

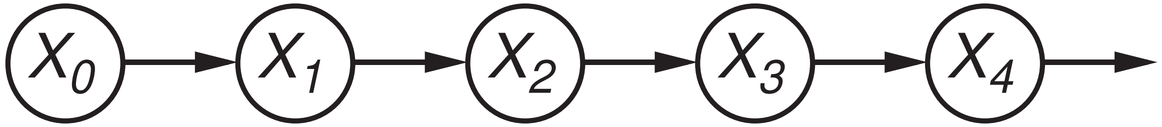


- Bayesian model of image X
  - (# of image points)≈(# of data points.)
  - Images have unique behaviors which may be modeled.
  - Maximum likelihood estimation works poorly.
  - Reduce model mismatch by estimating parameter θ.
- Frequentist model for θ and ϕ
  - (# of model parameters)<<(# of data points.)
  - Parameters are difficult to model.
  - Maximum likelihood estimation works well.

Markov Chains

- Topics to be covered:
  - 1-D properties
  - Parameter estimation
  - 2-D Markov Chains
- **Notation:** Upper case $\Rightarrow$ Random variable

Markov Chains



- Definition of (homogeneous) Markov chains

$$p(x_n|x_i \; i < n) = p(x_n|x_{n-1})$$

- Therefore, we may show that the probability of a sequence is given by

$$p(x) = p(x_0) \prod_{n=1}^{N} p(x_n|x_{n-1})$$

- Notice: X_n is **not** independent of X_{n+1}

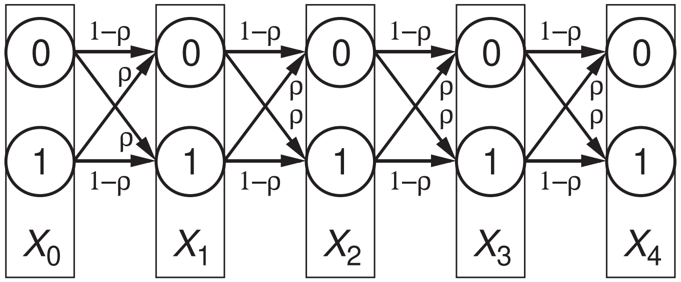
$$p(x_n|x_i \; i \neq n) = p(x_n|x_{n-1}, x_{n+1})$$

Parameters of Markov Chain

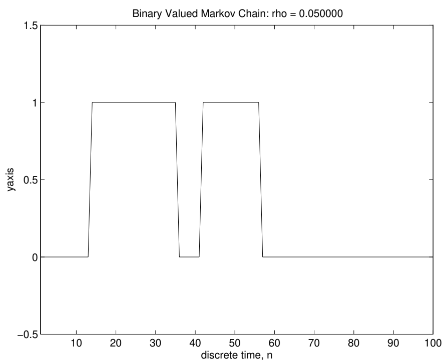
- Transition parameters are:

$$\theta_{j,i} = p(x_n = i | x_{n-1} = j)$$

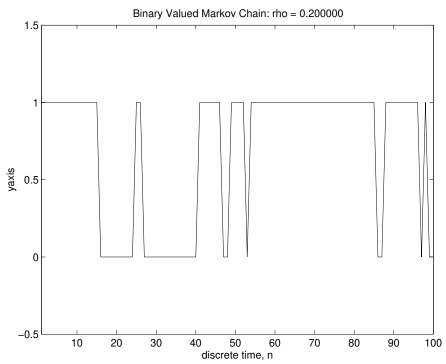
- Example: $\theta = \begin{bmatrix} 1-\rho & \rho \\ \rho & 1-\rho \end{bmatrix}$



- ρ is the probability of changing state.



$\rho = 0.05$



$\rho = 0.2$

Parameter Estimation for Markov Chains

- Maximum likelihood (ML) parameter estimation

$$\hat{\theta} = \arg\max_{\theta} p(x|\theta)$$

- For Markov chain

$$\hat{\theta}_{j,i} = \frac{h_{j,i}}{\sum_k h_{j,k}}$$

where $h_{j,i}$ is the histogram of transitions

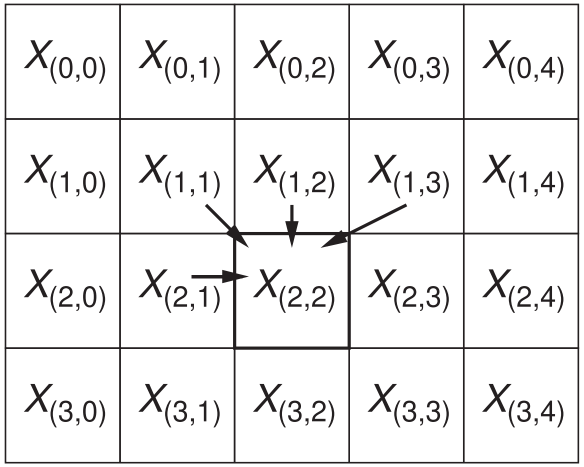
$$h_{j,i} = \sum_n \delta(x_n = i \,\&\, x_{n-1} = j)$$

- Example

$$x_n = 0, 0, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1$$

$$\theta = \begin{bmatrix} h_{0,0} & h_{0,1} \\ h_{1,0} & h_{1,1} \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 1 & 6 \end{bmatrix}$$

2-D Markov Chains



- Advantages:
  - Simple expressions for probability
  - Simple parameter estimation
- Disadvantages:
  - No natural ordering of pixels in image
  - Anisotropic model behavior

Discrete State Markov Random Fields

- Topics to be covered:
  - Definitions and theorems
  - 1-D MRF's
  - Ising model
  - M-Level model
  - Line process model

Markov Random Fields

- Noncausal model
- Advantages of MRF's
  - Isotropic behavior
  - Only local dependencies
- Disadvantages of MRF's
  - Computing probability is difficult
  - Parameter estimation is difficult
- Key theoretical result: Hammersley-Clifford theorem

Definition of Neighborhood System and Clique

- Define

  S - set of lattice points

  s - a lattice point, $s \in S$

  X_s - the value of X at s

  ∂s - the neighboring points of s

- A neighborhood system ∂s must be symmetric

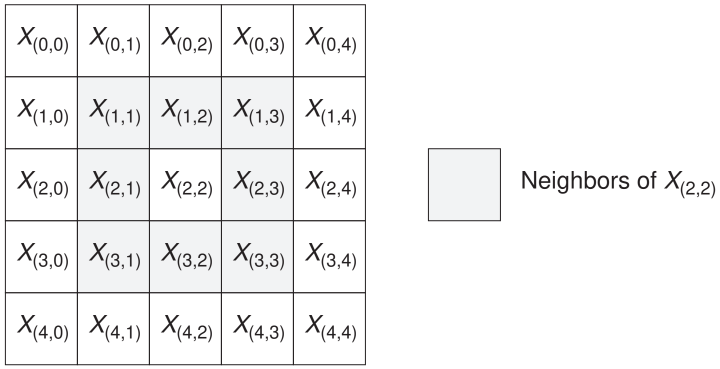
$$r \in \partial s \Rightarrow s \in \partial r \ \text{ also } s \notin \partial s$$

- A clique is a set of points, c, which are all neighbors of each other

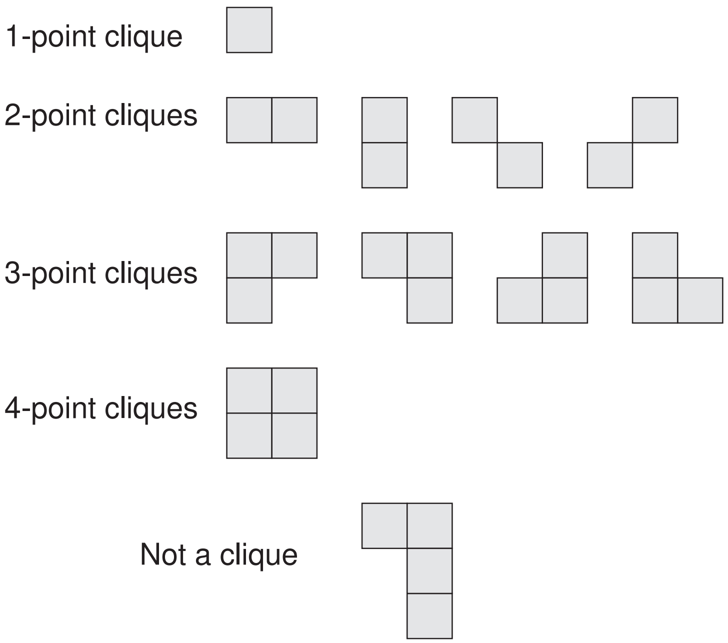
$$\forall s, r \in c, \; r \in \partial s$$

Example of Neighborhood System and Clique

- Example of 8 point neighborhood



- Example of cliques for 8 point neighborhood



Gibbs Distribution

x_c - The value of X at the points in clique c.

$V_c(x_c)$ - A potential function is any function of x_c.

- A (discrete) density is a Gibbs distribution if

$$p(x) = \frac{1}{Z} \exp\left\{ -\sum_{c \in \mathcal{C}} V_c(x_c) \right\}$$

  $\mathcal{C}$ is the set of all cliques

  Z is the normalizing constant for the density.

- Z is known as the **partition function**.
- $U(x) = \sum_{c \in \mathcal{C}} V_c(x_c)$ is known as the **energy function**.

Markov Random Field

- Definition: A random object X on the lattice S with neighborhood system ∂s is said to be a Markov random field if for all $s \in S$

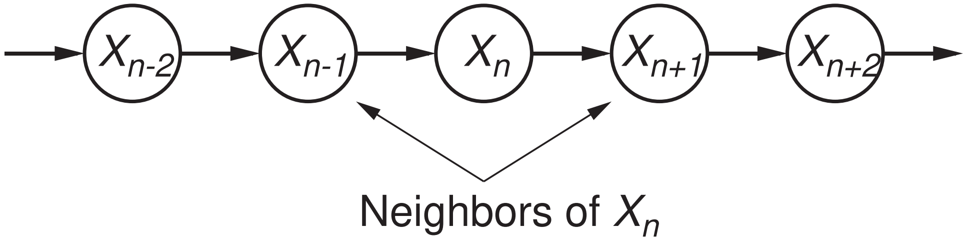
$$p(x_s | x_r \text{ for } r \neq s) = p(x_s | x_{\partial r})$$

Hammersley-Clifford Theorem[14]

$$\left(\begin{array}{c} X \text{ is a Markov random field} \\ \& \\ \forall x,\ P\{X = x\} > 0 \end{array} \right) \Longleftrightarrow \left(\begin{array}{c} P\{X = x\} \text{ has the form} \\ \text{of a Gibbs distribution} \end{array} \right)$$

- Gives you a method for writing the density for a MRF
- Does not give the value of Z, the partition function.
- Positivity, $P\{X = x\} > 0$, is a technical condition which we will generally assume.

Markov Chains are MRF's



- Neighbors of n are $\partial n = \{n-1, n+1\}$
- Cliques have the form $c = \{n-1, n\}$
- Density has the form

$$
\begin{aligned}
p(x) &= p(x_0) \prod_{n=1}^{N} p(x_n | x_{n-1}) \\
&= p(x_0) \exp\left\{ \sum_{n=1}^{N} \log p(x_n | x_{n-1}) \right\}
\end{aligned}
$$

- The potential functions have the form

$$
V(x_n, x_{n-1}) = \log p(x_n | x_{n-1})
$$

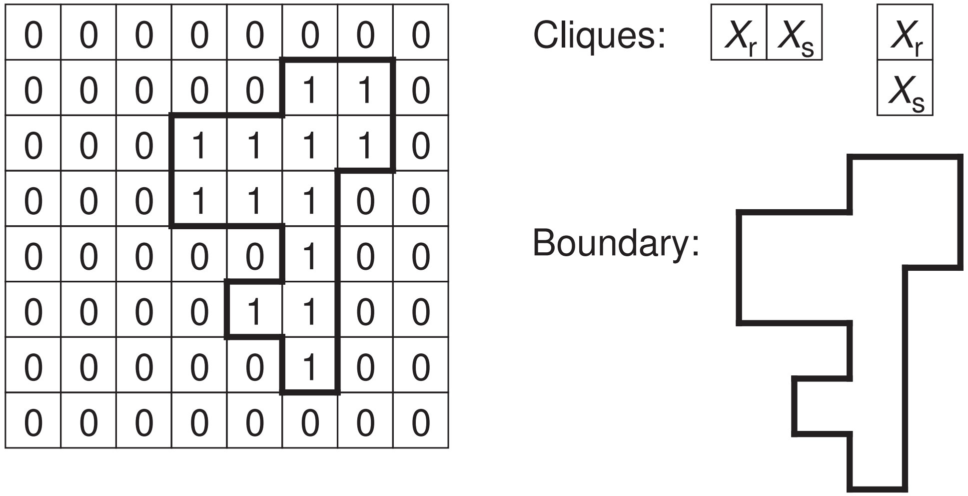
1-D MRF's are Markov Chains

- Let X_n be a 1-D MRF with $\partial n = \{n-1, n+1\}$
- The discrete density has the form of a Gibbs distribution

$$p(x) = p(x_0) \exp\left\{ \sum_{n=1}^{N} V(x_n, x_{n-1}) \right\}$$

- It may be shown that this is a Markov Chain.
- Transition probabilities may be difficult to compute.

The Ising Model: A 2-D MRF[100]



- Potential functions are given by

$$V(x_r, x_s) = \beta\delta(x_r \neq x_s)$$

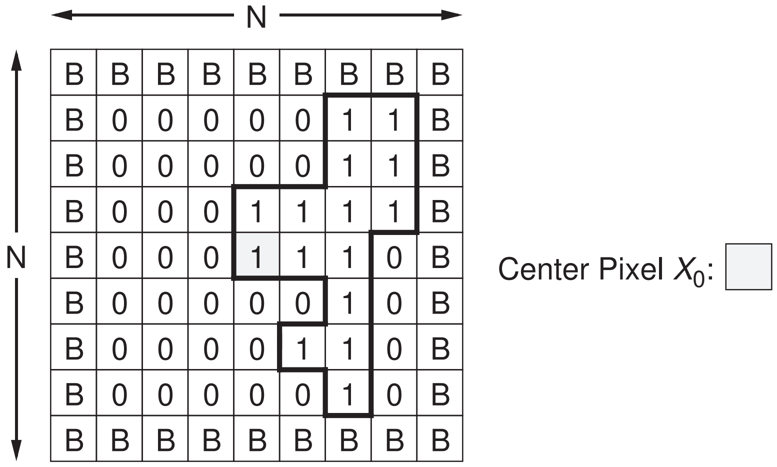
where β is a model parameter.

- Energy function is given by

$$\sum_{c\in\mathcal{C}} V_c(x_c) = \beta(\text{Boundary length})$$

- Longer boundaries $\Rightarrow$ less probable

Critical Temperature Behavior[127, 126, 100]



- $\frac{1}{\beta}$ is analogous to temperature.
- Peierls showed that for $\beta > \beta_c$

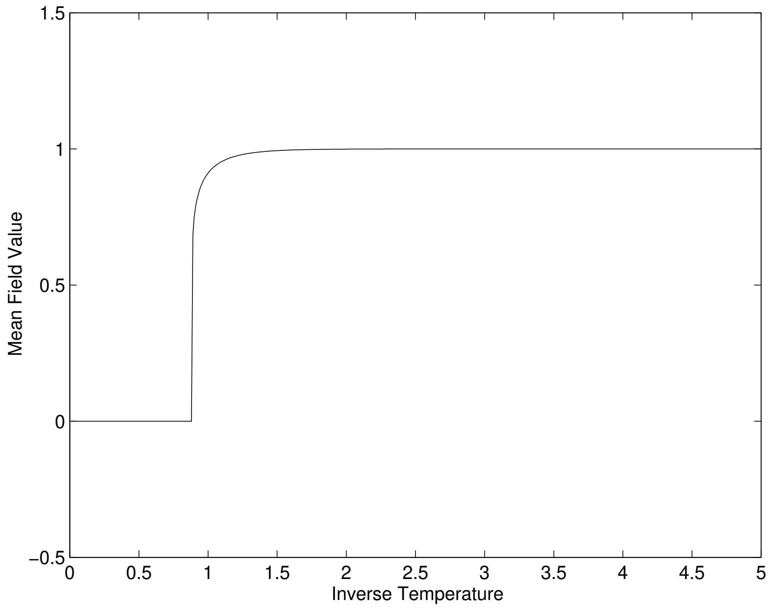
$$\lim_{N\to\infty} P(X_0 = 0|B = 0) \neq \lim_{N\to\infty} P(X_0 = 0|B = 1)$$

- The effect of the boundary does not diminish as $N \to \infty$!
- $\beta_c \approx .88$ is known as the critical temperature.

Critical Temperature Analysis[122]

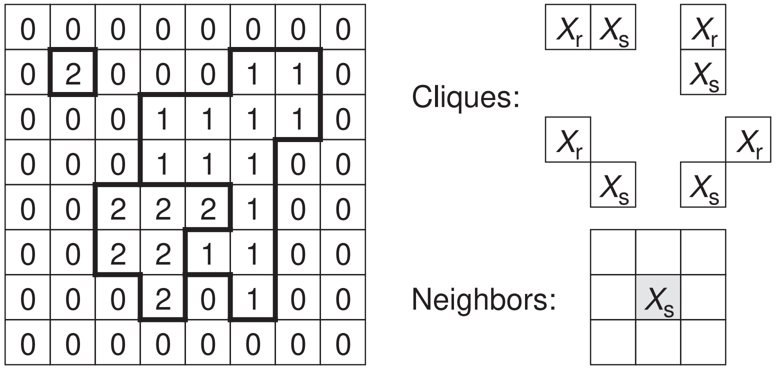
- Amazingly, Onsager was able to compute

$$E[X_0|B=1] = \begin{cases} \left(1 - \frac{1}{(\sinh(\beta))^4}\right)^{1/8} & \text{if } \beta > \beta_c \\ 0 & \text{if } \beta < \beta_c \end{cases}$$



- Onsager also computed an analytic expression for $Z(T)$!

M-Level MRF[16]



- Define $\mathcal{C}_1 \triangleq$ (hor./vert. cliques) and $\mathcal{C}_2 \triangleq$ (diag. cliques)
- Then

$$V(x_r, x_s) = \begin{cases} \beta_1 \delta(x_r \neq x_s) & \text{for } \{x_r, x_s\} \in \mathcal{C}_1 \\ \beta_2 \delta(x_r \neq x_s) & \text{for } \{x_r, x_s\} \in \mathcal{C}_2 \end{cases}$$

- Define

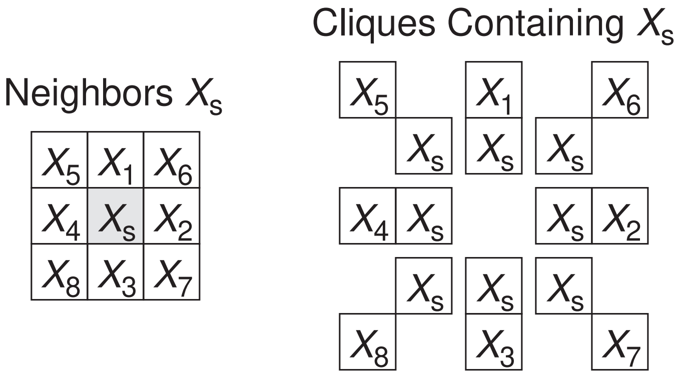
$$t_1(x) \triangleq \sum_{\{s,r\}\in\mathcal{C}_1} \delta(x_r \neq x_s)$$

$$t_2(x) \triangleq \sum_{\{s,r\}\in\mathcal{C}_2} \delta(x_r \neq x_s)$$

- Then the probability is given by

$$p(x) = \frac{1}{Z} \exp\left\{-(\beta_1 t_1(x) + \beta_2 t_2(x))\right\}$$

Conditional Probability of a Pixel



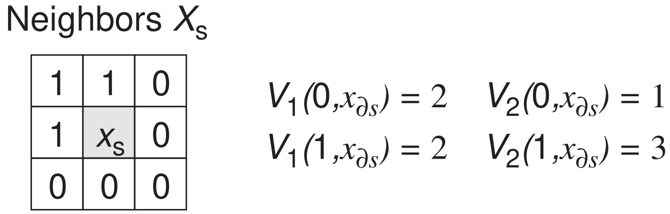
- The probability of a pixel given all other pixels is

$$p(x_s|x_{i\neq s}) = \frac{\frac{1}{Z}\exp\left\{-\sum_{c\in\mathcal{C}} V_c(x_c)\right\}}{\sum_{x_s=0}^{M-1}\frac{1}{Z}\exp\left\{-\sum_{c\in\mathcal{C}} V_c(x_c)\right\}}$$

- Notice: Any term $V_c(x_c)$ which does not include x_s cancels.

$$p(x_s|x_{i\neq s}) = \frac{\exp\left\{-\beta_1\sum_{i=1}^{4}\delta(x_s\neq x_i) - \beta_2\sum_{i=5}^{8}\delta(x_s\neq x_i)\right\}}{\sum_{x_s=0}^{M-1}\exp\left\{-\beta_1\sum_{i=1}^{4}\delta(x_s\neq x_i) - \beta_2\sum_{i=5}^{8}\delta(x_s\neq x_i)\right\}}$$

Conditional Probability of a Pixel (Continued)



- Define

$$\begin{aligned} v_1(x_s, \partial x_s) &\triangleq \text{\# of horz./vert. neighbors} \neq x_s \\ v_2(x_s, \partial x_s) &\triangleq \text{\# of diag. neighbors} \neq x_s \end{aligned}$$

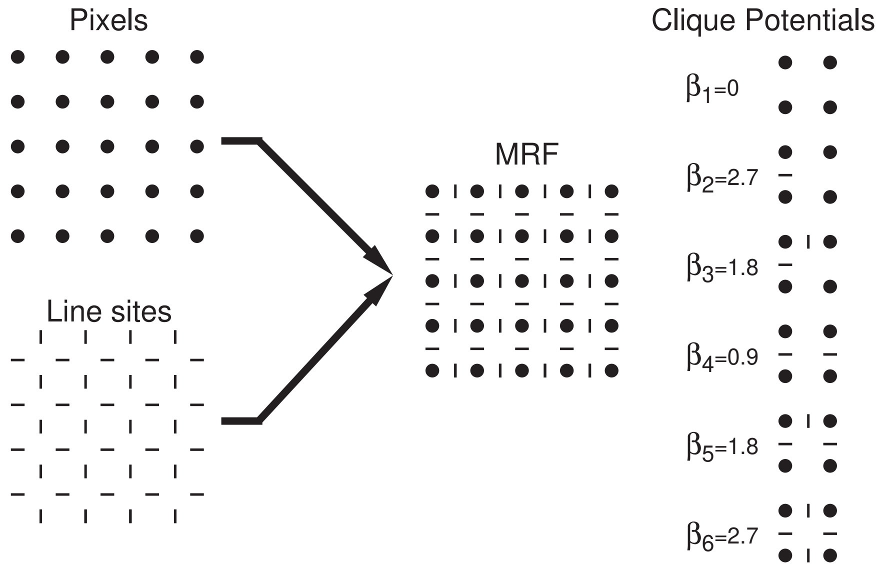
- Then

$$p(x_s|x_{i\neq s}) = \frac{1}{Z'} \exp\left\{-\beta_1 v_1(x_s, \partial x_s) - \beta_2 v_2(x_s, \partial x_s)\right\}$$

where Z' is an easily computed normalizing constant

- When $\beta_1, \beta_2 > 0$, X_s is most likely to be the majority neighboring class.

Line Process MRF [68]



- Line sites fall between pixels
- The values $\beta_1, \cdots, \beta_2$ determine the potential of line sites
- The potential of pixel values is

$$V(x_s, x_r, l_{r,s}) = \begin{cases} (x_s - x_r)^2 & \text{if } l_{r,s} = 0 \\ 0 & \text{if } l_{r,s} = 1 \end{cases}$$

- The field is
  - Smooth between line sites
  - Discontinuous at line sites

Simulation

- Topics to be covered:
  - Metropolis sampler
  - Gibbs sampler
  - Generalized Metropolis sampler

Generating Samples from a Gibbs Distribution

- How do we generate a random variable X with a Gibbs distribution?

$$p(x) = \frac{1}{Z} \exp\{-U(x)\}$$

- Generally, this problem is difficult.
- Markov Chains can be generated sequentially
- Non-causal structure of MRF's makes simulation difficult.

The Metropolis Sampler[118, 100]

- How do we generate a sample from a Gibbs distribution?

$$p(x) = \frac{1}{Z} \exp\{-U(x)\}$$

- Start with the sample x^k, and generate a new sample W with probability $q(w|x^k)$.

  Note: $q(w|x^k)$ must be symmetric.

$$q(w|x^k) = q(x^k|w)$$

- Compute $\Delta E(W) = U(W) - U(x^k)$, then do the following:

  If $\Delta E(W) < 0$

  – Accept: $X^{k+1} = W$

  If $\Delta E(W) \geq 0$

  – Accept: $X^{k+1} = W$ with probability $\exp\{-\Delta E(W)\}$
  – Reject: $X^{k+1} = x^k$ with probability $1 - \exp\{-\Delta E(W)\}$

Ergodic Behavior of Metropolis Sampler

- The sequence of random fields, X^k, form a Markov chain.
- Let $p(x^{k+1}|x^k)$ be the transition probabilities of the Markov chain.
- Then X^k is reversible

$$p(x^{k+1}|x^k)\exp\{-U(x^k)\} = \exp\{-U(x^{k+1})\}p(x^k|x^{k+1})$$

- Therefore, if the Markov chain is irreducible, then

$$\lim_{k\to\infty} P\{X^k = x\} = \frac{1}{Z}\exp\{-U(x)\}$$

- If every state can be reached, then as $k \to \infty$, X^k will be a sample from the Gibbs distribution.

Example Metropolis Sampler for Ising Model

	0	
1	x_s	0
	0	

- Assume $x_s^k = 0$.
- Generate a binary R.V., W, such that $P\{W = 0\} = 0.5$.

$$\begin{aligned} \Delta E(W) &= U(W) - U(x_s^k) \\ &= \begin{cases} 0 & \text{if } W = 0 \\ 2\beta & \text{if } W = 1 \end{cases} \end{aligned}$$

If $\Delta E(W) < 0$

- Accept $X_s^{k+1} = W$

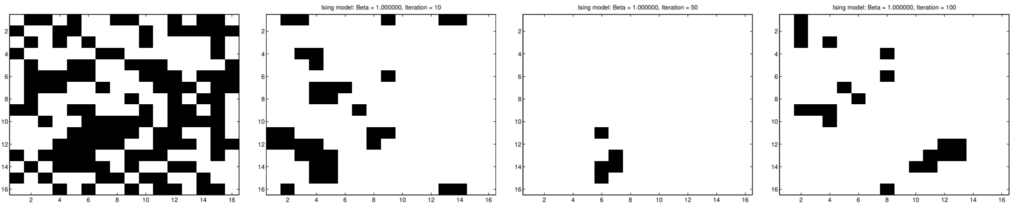
If $\Delta E(W) \geq 0$

- Accept: $X_s^{k+1} = W$ with probability $\exp\{-\Delta E(W)\}$
- Reject: $X_s^{k+1} = x_s^k$ with probability $1 - \exp\{-\Delta E(W)\}$

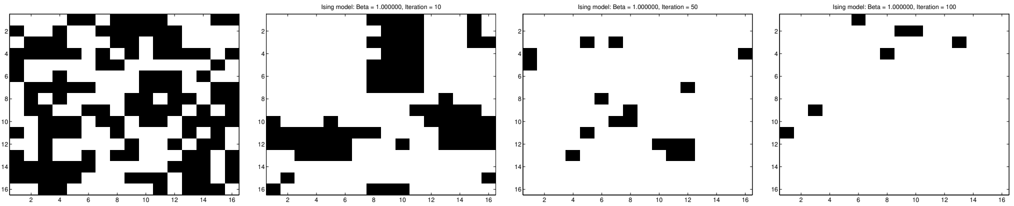
- Repeat this procedure for each pixel.
- **Warning:** for $\beta > \beta_c$ convergence can be extremely slow!

Example Simulation for Ising Model($\beta = 1.0$)

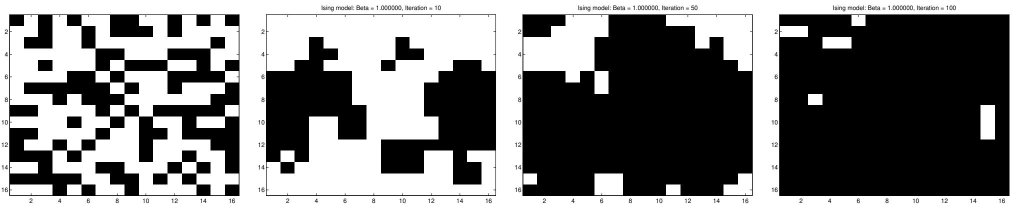
- Test 1



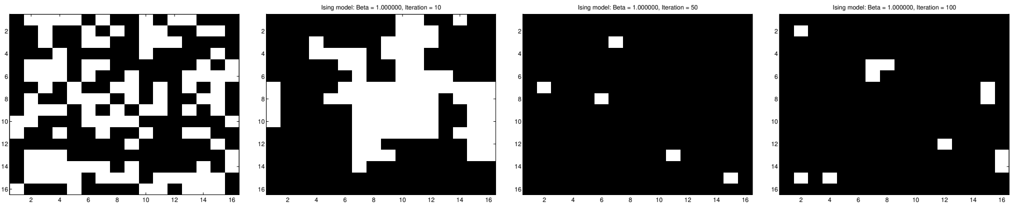
- Test 2



- Test 3



- Test 3



Advantages and Disadvantages of Metropolis Sampler

- Advantages
  - Can be implemented whenever ΔE is easy to compute.
  - Has guaranteed geometric convergence.
- Disadvantages
  - Can be slow if there are many rejections.
  - Is constrained to use a symmetric transition function $q(x^{k+1}|x^k)$.

Gibbs Sampler[68]

- Replace each point with a sample from its conditional distribution

$$p(x_s|x_i^k \; i \neq s) = p(x_s|x_{\partial s})$$

- Scan through all the points in the image.
- Advantage
  - Eliminates need for rejections $\Rightarrow$ faster convergence
- Disadvantage
  - Generating samples from $p(x_s|x_{\partial s})$ can be difficult.

Generalized Metropolis Sampler[80, 129]

- Hastings and Peskun generalized the Metropolis sampler for transition functions $q(w|x^k)$ which are not symmetric.
- The acceptance probability is then

$$\alpha(x_s^k, w) = \min\left\{1, \frac{q(x^k|w)}{q(w|x^k)} \exp\{-\Delta E(w)\}\right\}$$

- Special cases

$$q(w|x^k) = q(x^k|z) \Rightarrow \text{conventional Metropolis}$$
$$q(w_s|x^k) = p(x_s^k|x_{\partial s}^k)\big|_{x_s^k=w_s} \Rightarrow \text{Gibbs sampler}$$

- Advantage
  - Transition function may be chosen to minimize rejections[76]

Parameter Estimation for Discrete State MRF's

- Topics to be covered:
  - Why is it difficult?
  - Coding/maximum pseudolikehood
  - Least squares

Why is Parameter Estimation Difficult?

- Consider the ML estimate of β for an Ising model.
- Remember that

$$t_1(x) = (\#\text{ horz. and vert. neighbors of different value.})$$

- Then the ML estimate of β is

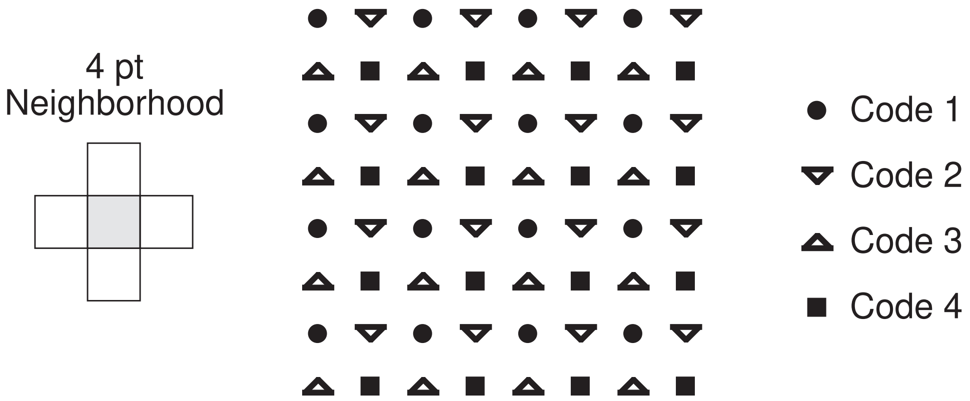
$$\begin{aligned}\hat{\beta} &= \arg\max_{\beta} \left\{ \frac{1}{Z(\beta)} \exp\left\{-\beta t_1(x)\right\} \right\} \\ &= \arg\max_{\beta} \left\{-\beta t_1(x) - \log Z(\beta)\right\}\end{aligned}$$

- However, $\log Z(\beta)$ has an intractable form

$$\log Z(\beta) = \log \sum_x \exp\left\{-\beta t_1(x)\right\}$$

- Partition function can not be computed.

Coding Method/Maximum Pseudolikelihood[15, 16]



- Assume a 4 point neighborhood
- Separate points into four groups or codes.
- Group (code) contains points which are conditionally independent given the other groups (codes).

$$\hat{\beta} = \arg\max_{\beta} \prod_{s \in \text{Code}_k} p(x_s | x_{\partial s})$$

- This is tractable (but not necessarily easy) to compute

Least Squares Parameter Estimation[49]

- It can be shown that for an Ising model

$$\log \frac{P\{X_s = 1 | x_{\partial s}\}}{P\{X_s = 0 | x_{\partial s}\}} = -\beta \left(V_1(1|x_{\partial s}) - V_1(0|x_{\partial s}) \right)$$

- For each unique set of neighboring pixel values, $x_{\partial s}$, we may compute
  - The observed rate of $\log \frac{P\{X_s=1|x_{\partial s}\}}{P\{X_s=0|x_{\partial s}\}}$
  - The value of $(V_1(1|x_{\partial s}) - V_1(0|x_{\partial s}))$
  - This produces a set of over-determined linear equations which can be solved for β.
- This least squares method is easily implemented.

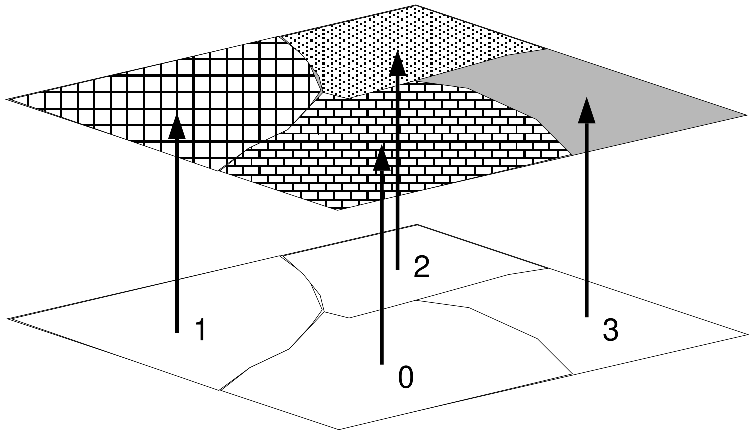
Theoretical Results in Parameter Estimation for MRF's

- Inconsistency of ML estimate for Ising model[130, 131]
  - Caused by critical temperature behavior.
  - Single sample of Ising model cannot distinguish between high β with mean 1/2, and low β with large mean.
  - Not identifiable
- Consistency of maximum pseudolikelihood estimate[69]
  - Requires an identifiable parameterization.

Application of MRF's to Segmentation

- Topics to be covered:
  - The Model
  - Bayesian Estimation
  - MAP Optimization
  - Parameter Estimation
  - Other Approaches

Bayesian Segmentation Model



- Discrete MRF is used to model the segmentation field.
- Each class is represented by a value $X_s \in \{0, \cdots, M-1\}$
- The joint probability of the data and segmentation is

$$P\{Y \in dy, X = x\} = p(y|x)p(x)$$

  where

  - $p(y|x)$ is the data model
  - $p(x)$ is the segmentation model

Bayes Estimation

- $C(x, X)$ is the cost of guessing x when X is the correct answer.
- $\hat{X}$ is the estimated value of X.
- $E[C(\hat{X}, X)]$ is the expected cost (risk).
- Objective: Choose the estimator $\hat{X}$ which minimizes $E[C(\hat{X}, X)]$.

Maximum *A Posteriori* (MAP) Estimation

- Let $C(x, X) = \delta(x \neq X)$
- Then the optimum estimator is given by

$$
\begin{aligned}
\hat{X}_{MAP} &= \arg\max_x p_{x|y}(x|Y) \\
&= \arg\max_x \log \frac{p_{y,x}(Y, x)}{p_y(Y)} \\
&= \arg\max_x \{\log p(Y|x) + \log p(x)\}
\end{aligned}
$$

- Advantage:
  - Can be computed through direct optimization
- Disadvantage:
  - Cost function is unreasonable for many applications

Maximizer of the Posterior Marginals (MPM) Estimation[116]

- Let $C(x, X) = \sum_{s \in S} \delta(x_s \neq X_s)$
- Then the optimum estimator is given by

$$\hat{X}_{MPM} = \arg\max_{x_s} p_{x_s|Y}(x_s|Y)$$

- Compute the most likely class for each pixel
- Method:
  - Use simulation method to generate samples from $p_{x|y}(x|y)$.
  - For each pixel, choose the most frequent class.
- Advantage:
  - Minimizes number of misclassified pixels
- Disadvantage:
  - Difficult to compute

MAP Optimization for Segmentation

- Assume the data model

$$p_{y|x}(y|x) = \prod_{s \in S} p(y_s|x_s)$$

- And the prior model (Ising model)

$$p_x(x) = \frac{1}{Z'} \exp\{-\beta t_1(x)\}$$

- Then the MAP estimate has the form

$$\hat{x} = \arg\min_x \left\{ -\log p_{y|x}(y|x) + \beta t_1(x) \right\}$$

- This optimization problem is very difficult

Iterated Conditional Modes [16]

- The problem:

$$\hat{x}_{MAP} = \arg\min_x \left\{ - \sum_{s \in S} \log p_{y_s|x_s}(y_s|x_s) + \beta t_1(x) \right\}$$

- Iteratively minimize the function with respect to each pixel, x_s.

$$\hat{x}_s = \arg\min_{x_s} \left\{ - \log p_{y_s|x_s}(y_s|x_s) + \beta v_1(x_s|x_{\partial s}) \right\}$$

- This converges to a local minimum in the cost function

Simulated Annealing [68]

- Consider the Gibbs distribution

$$\frac{1}{Z}\exp\left\{-\frac{1}{T}U(x)\right\}$$

where

$$U(x) = \sum_{s\in S} \log p_{y_s|x_s}(y_s|x_s) + \beta t_1(x)$$

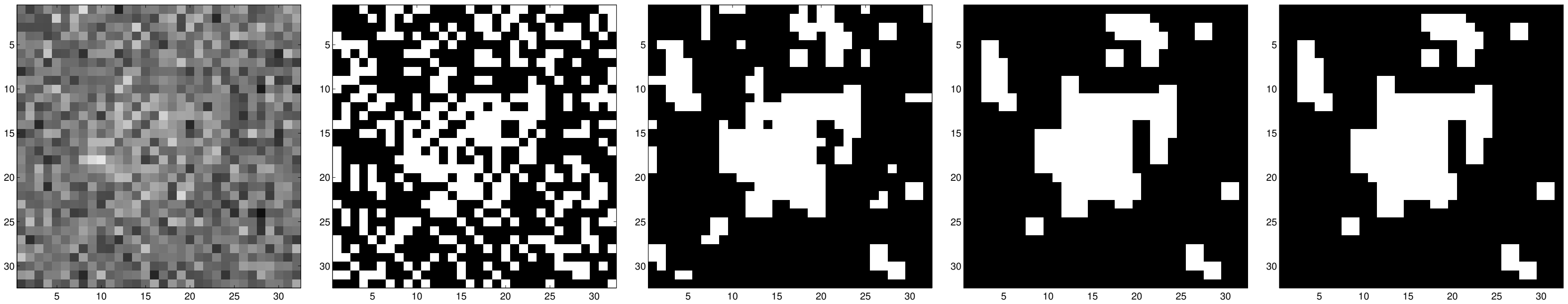
- As $T \rightarrow 0$, the distribution becomes clustered about $\hat{x}_{MAP}$.
- Use simulation method to generate samples from distribution.
- Slowly let $T \rightarrow 0$.
- If $T_k = \frac{T_1}{1+\log k}$ for iteration k, the the simulation converges to $\hat{x}_{MAP}$ almost surely.
- Problem: This is very slow!

Multiscale MAP Segmentation

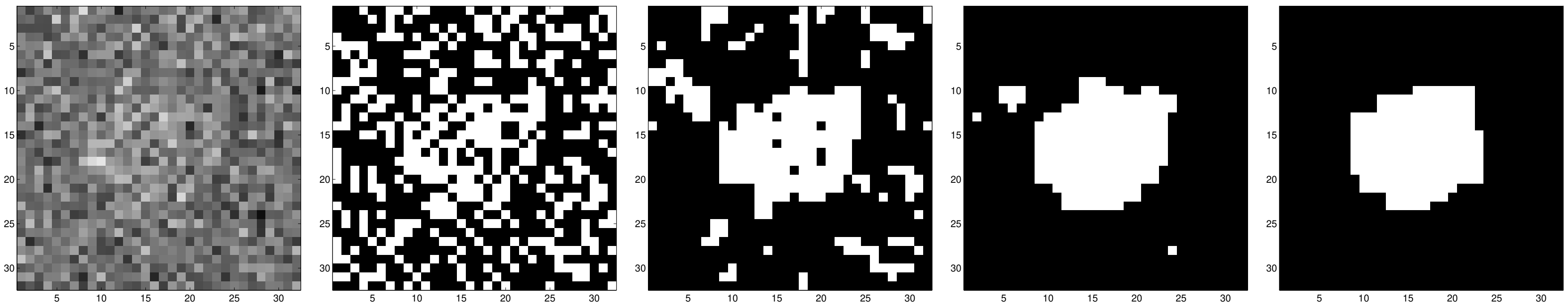
- Renormalization theory[72]
  - Theoretically results in the exact MAP segmentation
  - Requires the computation of intractable functions
  - Can be implemented with approximation
- Multiscale resolution segmentation[23]
  - Performs ICM segmentation in a coarse-to-fine sequence
  - Each MAP optimization is initialized with the solution from the previous coarser resolution
  - Used the fact that a discrete MRF constrained to be block constant is still a MRF.
- Multiscale Markov random fields[97]
  - Extended MRF to the third dimension of scale
  - Formulated a parallel computational approach

Segmentation Example

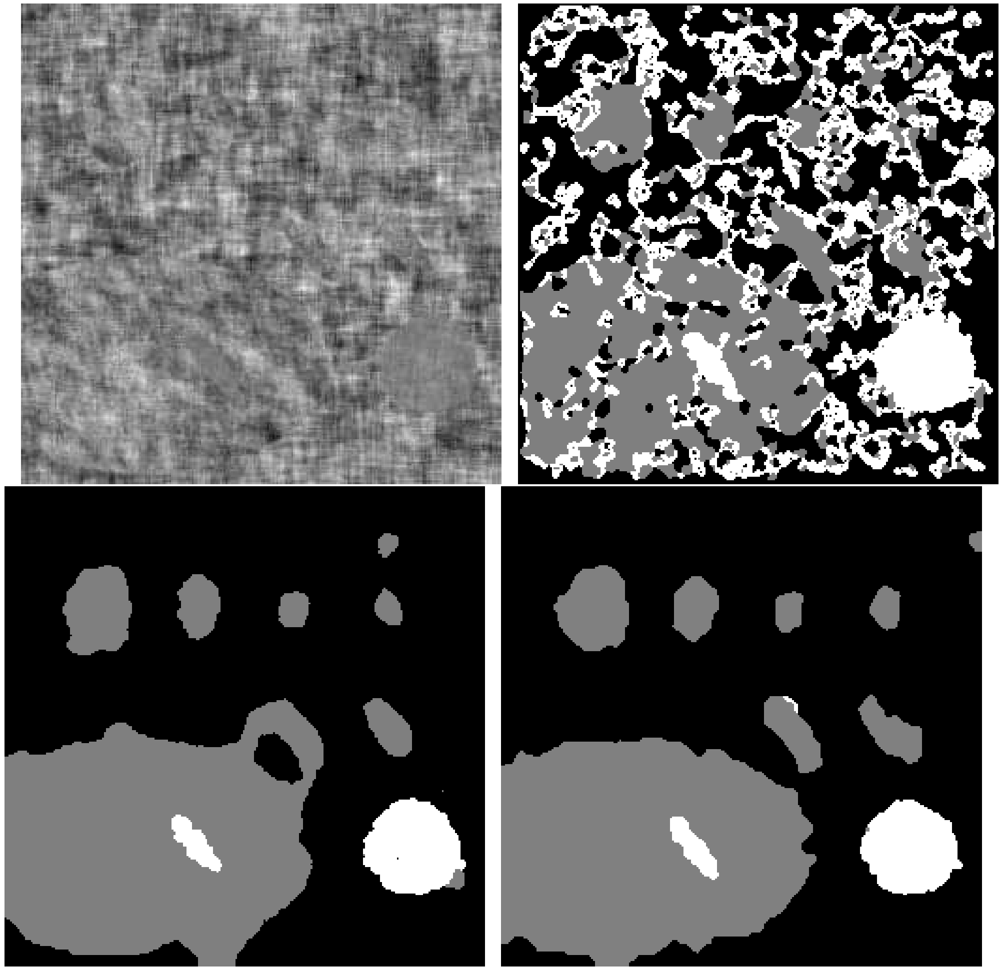
- Iterated Conditional Modes (ICM): ML ; ICM 1; ICM 5; ICM 10



- Simulated Annealing (SA): ML ; SA 1; SA 5; SA 10



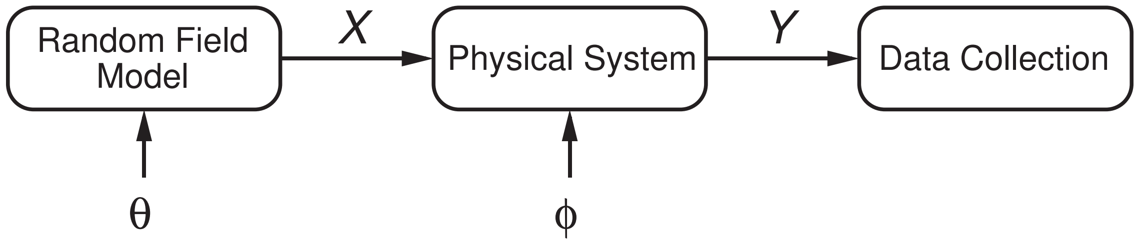
Texture Segmentation Example



a	b
c	d

a) Synthetic image with 3 textures b) ICM - 29 iterations c) Simulated Annealing - 100 iterations d) Multiresolution - 7.8 iterations

Parameter Estimation



- Question: How do we estimate θ from Y?
- Problem: We don't know X!
- Solution 1: Joint MAP estimation [104]

$$(\hat{\theta}, \hat{x}) = \arg\max_{\theta, x} p(y, x|\theta)$$

  – Problem: The solution is biased.

- Solution 2: Expectation maximization algorithm [9, 70]

$$\hat{\theta}^{k+1} = \arg\max_{\theta} E[\log p(Y, X|\theta)|Y = y, \theta^k]$$

  – Expectation may be computed using simulation techniques or mean field theory.

Other Approaches to using Discrete MRFs

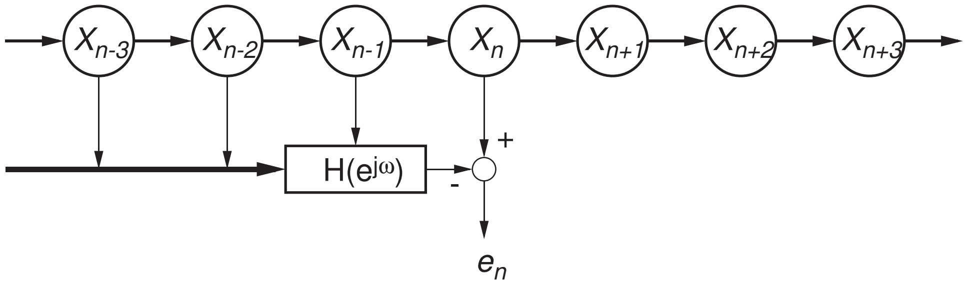
- Dynamic programming does not work in 2-D, but a number of researchers have formulated approximate recursive solutions to MAP estimation[48, 169].
- Mean field theory has also been studied as a method for computing the MPM estimate[176].

Gaussian Random Process Models

- Topics to be covered:
  - Autoregressive (AR) models
  - Simultaneous Autoregressive (SAR) models
  - Gaussian MRF's
  - Generalization to 2-D

Autoregressive (AR) Models

$$e_n = x_n - \sum_{k=1}^{\infty} x_{n-k} h_k$$



- $H(e^{j\omega})$ is an optimal predictor $\Rightarrow e(n)$ is white noise.
- The density for the N point vector X is given by

$$p_x(x) = \frac{1}{Z} \exp\left\{-\frac{1}{2} x^t \mathbf{A}^t \mathbf{A} x\right\}$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & & -h_{m-n} \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$$

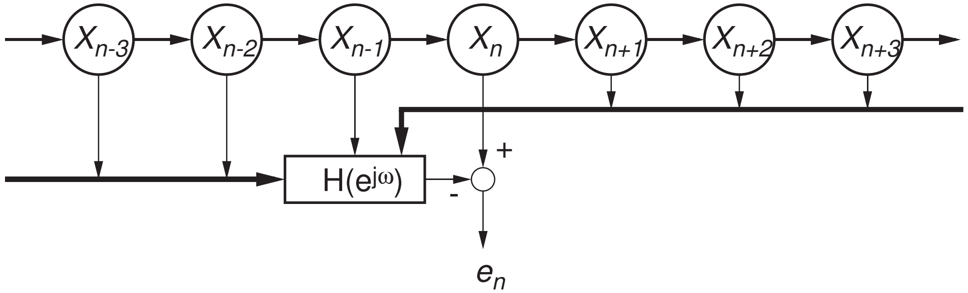
$$Z = (2\pi)^{N/2} |\mathbf{A}|^{-1} = (2\pi)^{N/2}$$

- The power spectrum of X is

$$S_x(e^{j\omega}) = \frac{\sigma_e^2}{|1 - H(e^{j\omega})|^2}$$

Simultaneous Autoregressive (SAR) Models[95, 94]

$$e_n = x_n - \sum_{k=1}^{\infty}(x_{n-k} - x_{n+k})h_k$$



- $e(n)$ is white noise $\Rightarrow H(e^{j\omega})$ is **not** an optimal non-causal predictor.
- The density for the N point vector X is given by

$$p_x(x) = \frac{1}{Z}\exp\left\{-\frac{1}{2}x^t\mathbf{A}^t\mathbf{A}x\right\}$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & & -h_{m-n} \\ & \ddots & \\ -h_{n-m} & & 1 \end{bmatrix}$$

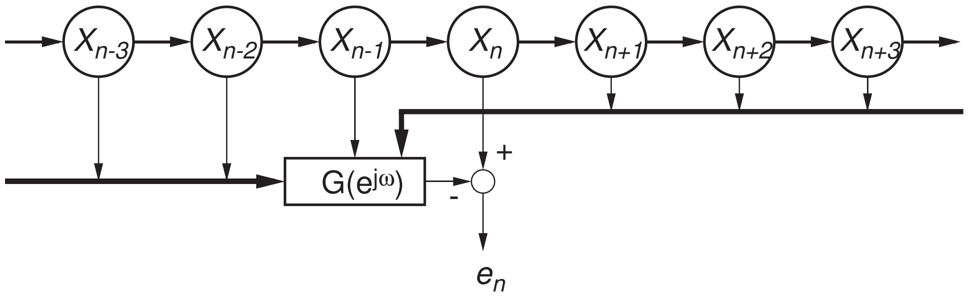
$$Z = (2\pi)^{N/2}|\mathbf{A}|^{-1} \approx (2\pi)^{N/2}\exp\left\{-\frac{N}{2\pi}\int_{-\pi}^{\pi}\log|1 - H(e^{j\omega})|d\omega\right\}$$

- The power spectrum of X is

$$S_x(e^{j\omega}) = \frac{\sigma_e^2}{|1 - H(e^{j\omega})|^2}$$

Conditional Markov (CM) Models (i.e. MRF's)[95, 94]

$$e_n = x_n - \sum_{k=1}^{\infty} (x_{n-k} - x_{n+k}) g_k$$



- $G(e^{j\omega})$ is an optimal non-causal predictor $\Rightarrow$ $e(n)$ is **not** white noise.
- The density for the N point vector X is given by

$$p_x(x) = \frac{1}{Z} \exp\left\{ -\frac{1}{2} x^t \mathbf{B} x \right\}$$

where

$$\mathbf{B} = \begin{bmatrix} 1 & & -g_{m-n} \\ & \ddots & \\ -g_{n-m} & & 1 \end{bmatrix}$$

$$Z = (2\pi)^{N/2} |\mathbf{B}|^{-1/2} \approx (2\pi)^{N/2} \exp\left\{ -\frac{N}{4\pi} \int_{-\pi}^{\pi} \log(1 - G(e^{j\omega})) d\omega \right\}$$

- The power spectrum of X is

$$S_x(e^{j\omega}) = \frac{\sigma_e^2}{1 - G(e^{j\omega})}$$

Generalization to 2-D

- Same basic properties hold.
- Circulant matrices become circulant block circulant.
- Toeplitz matrices become Toeplitz block Toeplitz.
- SAR and MRF models are more important in 2-D.

Non-Gaussian Continuous State MRF's

- Topics to be covered:
  - Quadratic functions
  - Non-Convex functions
  - Continuous MAP estimation
  - Convex functions

Why use Non-Gaussian MRF's?

- Gaussian MRF's do not model edges well.
- In applications such as image restoration and tomography, Gaussian MRF's either
  - Blur edges
  - Leave excessive amounts of noise

Gaussian MRF's

- Gaussian MRF's have density functions with the form

$$p(x) = \frac{1}{Z} \exp\left\{ -\sum_{s \in S} a_s x_s^2 - \sum_{\{s,r\} \in C} b_{sr} |x_s - x_r|^2 \right\}$$

- We will assume $a_s = 0$.
- The terms $|x_s - x_r|^2$ penalize rapid changes in gray level.
- MAP estimate has the form

$$\hat{x} = \arg\min_x \left\{ -\log p(y|x) + \sum_{\{s,r\} \in C} b_{sr} |x_s - x_r|^2 \right\}$$

- **Problem:** Quadratic function, $|\cdot|^2$, excessively penalizes image edges.

Non-Gaussian MRF's Based on Pair-Wise Cliques

- We will consider MRF's with pair-wise cliques

$$p(x) = \frac{1}{Z} \exp\left\{ - \sum_{\{s,r\} \in C} b_{sr} \rho\left(\frac{x_s - x_r}{\sigma} \right) \right\}$$

$|x_s - x_r|$ - is the change in gray level.

σ - controls the gray level variation or scale.

$\rho(\Delta)$:

– Known as the potential function.

– Determines the cost of abrupt changes in gray level.

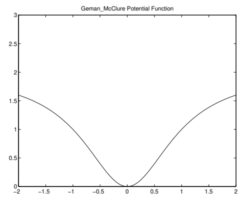
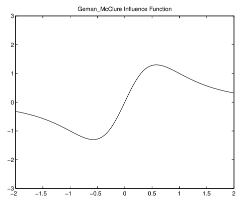
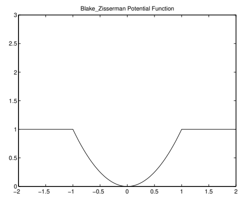
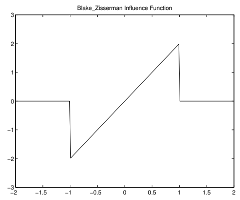
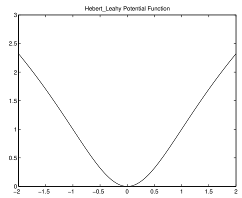
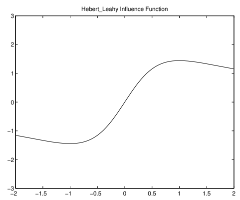
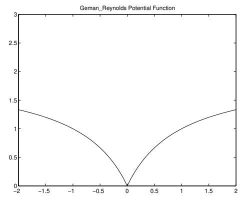
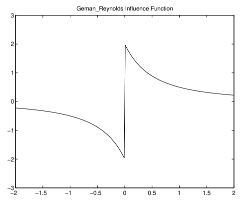
– $\rho(\Delta) = |\Delta|^2$ is the Gaussian model.

$\rho'(\Delta) = \frac{d\rho(\Delta)}{d\Delta}$:

– Known as the influence function from "M-estimation"[139, 85].

– Determines the attraction of a pixel to neighboring gray levels.

Non-Convex Potential Functions

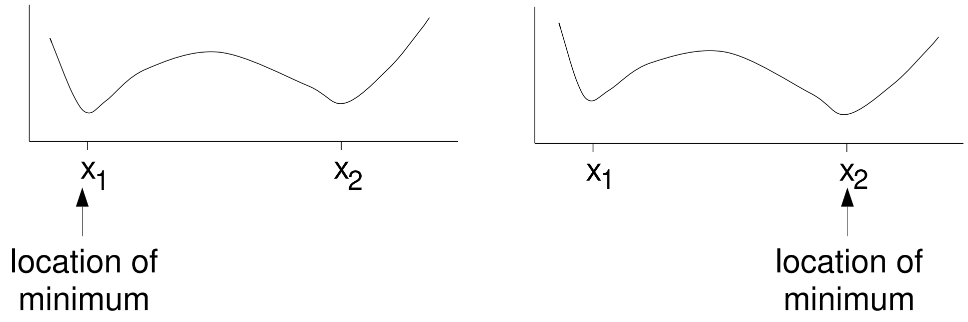
Authors	$\rho(\Delta)$	Ref.	Potential func.	Influence func.
Geman and McClure	$\frac{\Delta^2}{1+\Delta^2}$	[70, 71]		
Blake and Zisserman	$\min\left\{\Delta^2, 1\right\}$	[20, 19]		
Hebert and Leahy	$\log\left(1+\Delta^2\right)$	[81]		
Geman and Reynolds	$\frac{\|\Delta\|}{1+\|\Delta\|}$	[66]		

Properties of Non-Convex Potential Functions

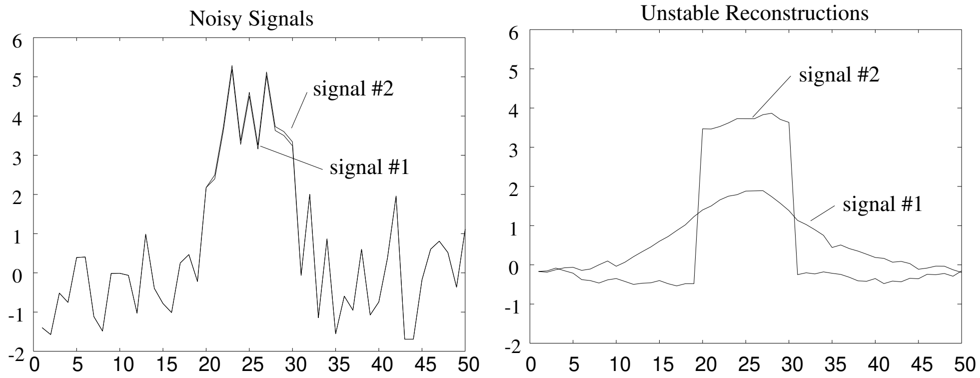
- Advantages
  - Very sharp edges
  - Very general class of potential functions
- Disadvantages
  - Difficult (impossible) to compute MAP estimate
  - Usually requires the choice of an edge threshold
  - **MAP estimate is a discontinuous function of the data**

Continuous (Stable) MAP Estimation[25]

- Minimum of non-convex function can change abruptly.

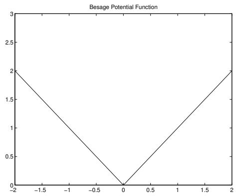
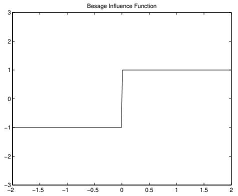
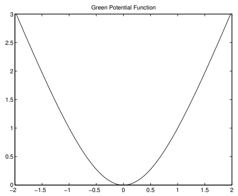
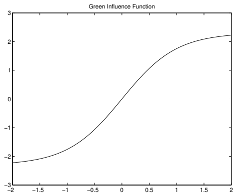
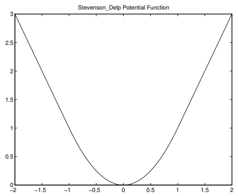
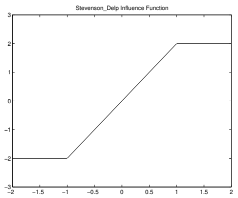
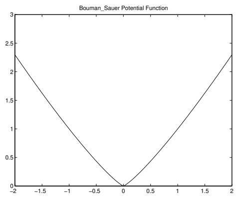
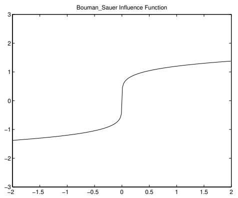


- Discontinuous MAP estimate for Blake and Zisserman potential.



- Theorem:[25] - If the log of the posterior density is **strictly convex**, then the MAP estimate is a continuous function of the data.

Convex Potential Functions

<table>
<tr><th>Authors(Name)</th><th>$\rho(\Delta)$</th><th>Ref.</th><th>Potential func.</th><th>Influence func.</th></tr>
<tr><td>Besag</td><td>$|\Delta|$</td><td>[17]</td><td>
</td><td>
</td></tr>
<tr><td>Green</td><td>$\log\cosh\Delta$</td><td>[75]</td><td>
</td><td>
</td></tr>
<tr><td>Stevenson and Delp (Huber function)</td><td>$\min\{|\Delta|^2, 2|\Delta|-1\}$</td><td>[155]</td><td>
</td><td>
</td></tr>
<tr><td>Bouman and Sauer (Generalized Gaussian MRF)</td><td>$|\Delta|^p$</td><td>[25]</td><td>
</td><td>
</td></tr>
</table>

Properties of Convex Potential Functions

- Both $\log\cosh(\Delta)$ and Huber functions
  - Quadratic for $|\Delta| << 1$
  - Linear for $|\Delta| >> 1$
  - Transition from quadratic to linear determines edge threshold.
- Generalized Gaussian MRF (GGMRF) functions
  - Include $|\Delta|$ function
  - Do not require an edge threshold parameter.
  - Convex and differentable for $p > 1$.

Parameter Estimation for Continuous MRF's

- Topics to be covered:
  - Estimation of scale parameter, σ
  - Estimation of temperature, T, and shape, p

ML Estimation of Scale Parameter, σ, for Continuous MRF's [26]

- For any continuous state Gibbs distribution

$$p(x) = \frac{1}{Z(\sigma)} \exp\left\{-U(x/\sigma)\right\}$$

  the partition function has the form

$$Z(\sigma) = \sigma^N Z(1)$$

- Using this result the ML estimate of σ is given by

$$\left.\frac{\sigma}{N}\frac{d}{d\sigma}U(x/\sigma)\right|_{\sigma=\hat{\sigma}} - 1 = 0$$

- This equation can be solved numerically using any root finding method.

ML Estimation of σ for GGMRF's [108, 26]

- For a Generalized Gaussian MRF (GGMRF)

$$p(x) = \frac{1}{\sigma^N Z(1)} \exp\left\{ -\frac{1}{p\sigma^p} U(x) \right\}$$

where the energy function has the property that for all $\alpha > 0$

$$U(\alpha x) = \alpha^p U(x)$$

- Then the ML estimate of σ is

$$\hat{\sigma} = \left(\frac{1}{N} U(x) \right)^{(1/p)}$$

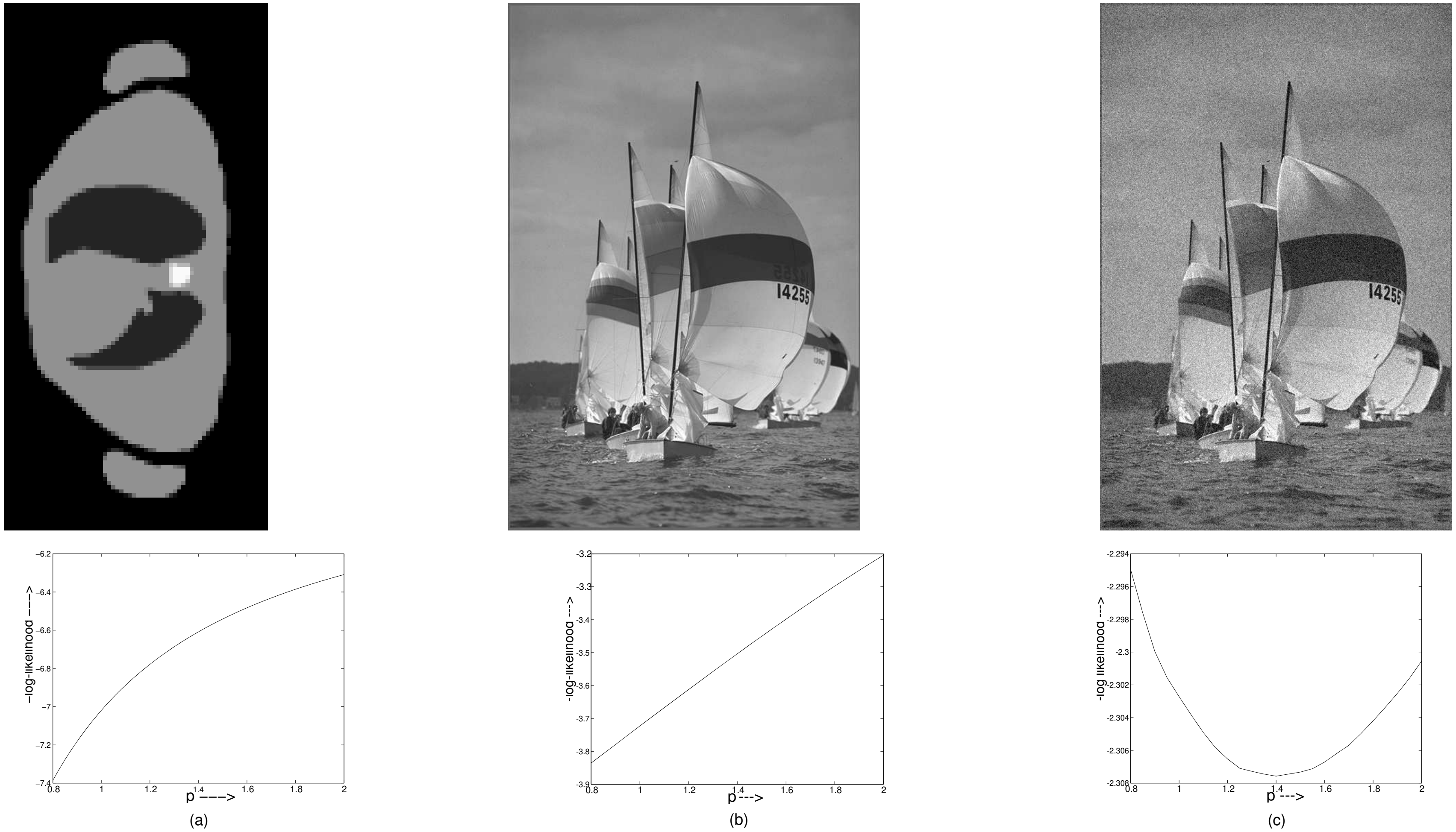
- Notice for that for the i.i.d. Gaussian case, this is

$$\hat{\sigma} = \sqrt{\frac{1}{N} \sum_s |x_s|^2}$$

Estimation of Temperature, T, and Shape, p, Parameters

- ML estimation of T[71]
  - Used to estimate T for any distribution.
  - Based on "off line" computation of log partition function.
- Adaptive method [133]
  - Used to estimate p parameter of GGMRF.
  - Based on measurement of kurtosis.
- ML estimation of p[145, 144]
  - Used to estimate p parameter of GGMRF.
  - Based on "off line" computation of log partition function.

Example Estimation of p Parameter



- ML estimation of p for (a) transmission phantom (b) natural image (c) image corrupted with Gaussian noise. The plot below each image shows the corresponding negative log-likelihood as a function of p. The ML estimate is the value of p that minimizes the plotted function.

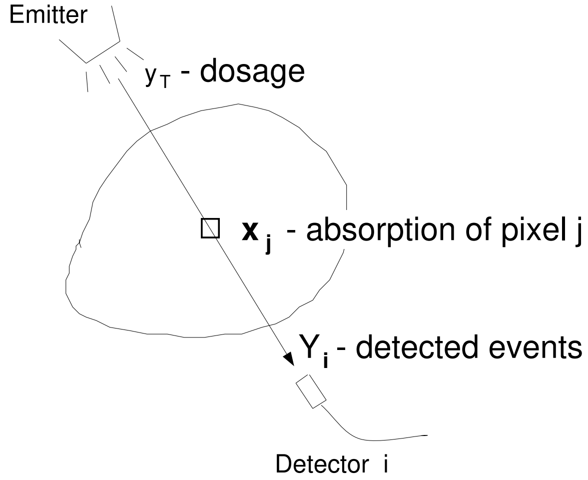
Application to Tomography

- Topics to be covered:
  - Tomographic system and data models
  - MAP Optimization
  - Parameter estimation

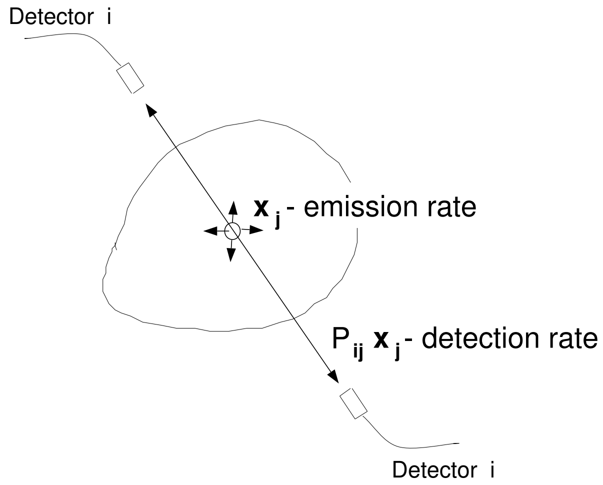
The Tomography Problem

- Recover image cross-section from integral projections

- Transmission problem



- Emission problem



Statistical Data Model[27]

- Notation
  - y - vector of photon counts
  - x - vector of image pixels
  - P - projection matrix
  - $P_{j,*}$ - j^{th} row of projection matrix
- Emission formulation

$$\log p(y|x) = \sum_{i=1}^{M} \left(-P_{i*}x \ + y_i \log\{P_{i*}x\} - \log(y_i!)\right)$$

- Transmission formulation

$$\log p(y|x) = \sum_{i=1}^{M} \left(-y_T e^{-P_{i*}x} + y_i(\log y_T - P_{i*}x) - \log(y_i!)\right)$$

- Common form

$$\boxed{\log p(y|x) = -\Sigma_{i=1}^{M} f_i(P_{i*}x)}$$

  - $f_i(\cdot)$ is a convex function
  - Not a hard problem!

Maximum A Posteriori Estimation (MAP)

- MAP estimate incorporates prior knowledge about image

$$
\begin{aligned}
\hat{x} &= \arg\max_{x} p(x|y) \\
&= \arg\max_{x>0} \left\{ -\sum_{i=1}^{M} f_i(P_{i*}x) - \sum_{k<j} b_{k,j}\, \rho(x_k - x_j) \right\}
\end{aligned}
$$

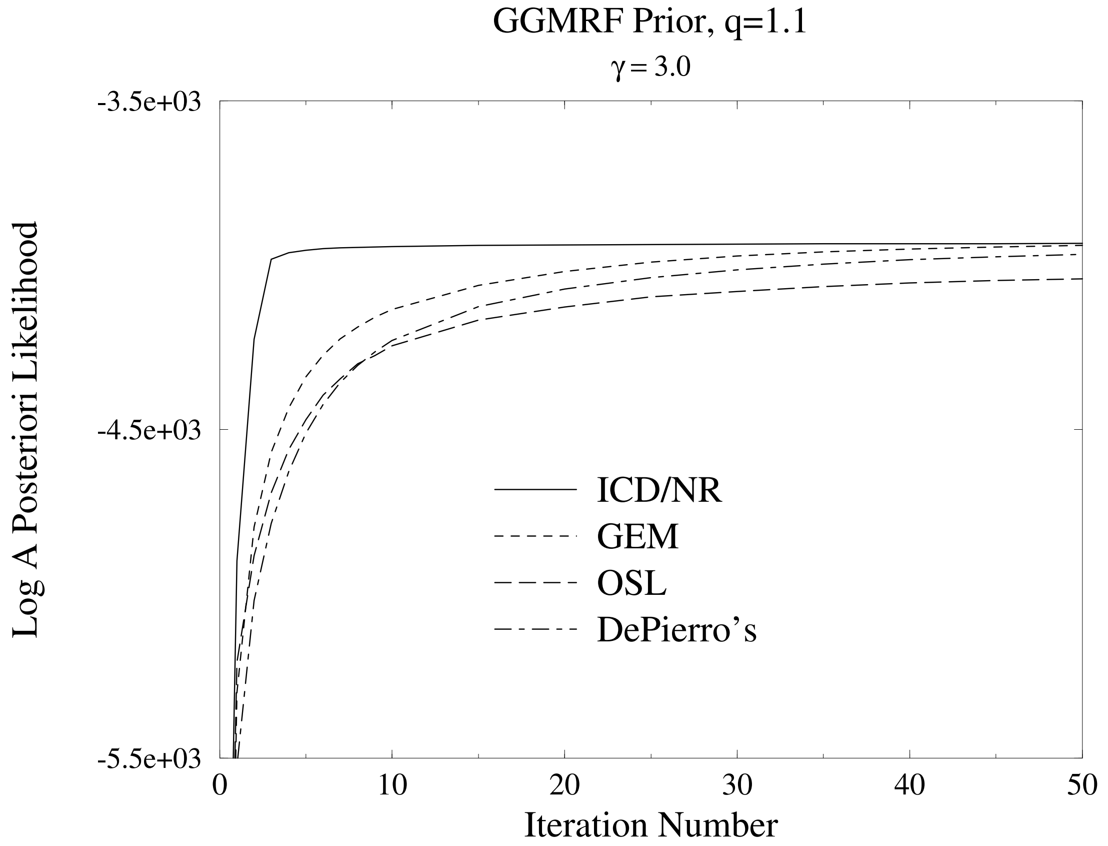
- Can be solved using direct optimization
- Incorporates positivity constraint

MAP Optimization Strategies

- Expectation maximization (EM) based optimization strategies
  - ML reconstruction[151, 107]
  - MAP reconstruction[81, 75, 84]
  - Slow convergence; Similar to gradient search.
  - Accelerated EM approach[59]
- Direct optimization
  - Preconditioned gradient descent with soft positivity constraint[45]
  - ICM iterations (also known as ICD and Gauss-Seidel)[27]

Convergence of ICM Iterations: MAP with Generalized Gaussian Prior $q = 1.1$

- ICM also known as iterative coordinate descent (ICD) and Gauss-Seidel



- Convergence of MAP estimates using ICD/Newton-Raphson updates, Green's (OSL), and Hebert/Leahy's GEM, and De Pierro's method, and a generalized Gaussian prior model with $q = 1.1$ and $\gamma = 3.0$.

Estimation of σ from Tomographic Data

- Assume a GGMRF prior distribution of the form

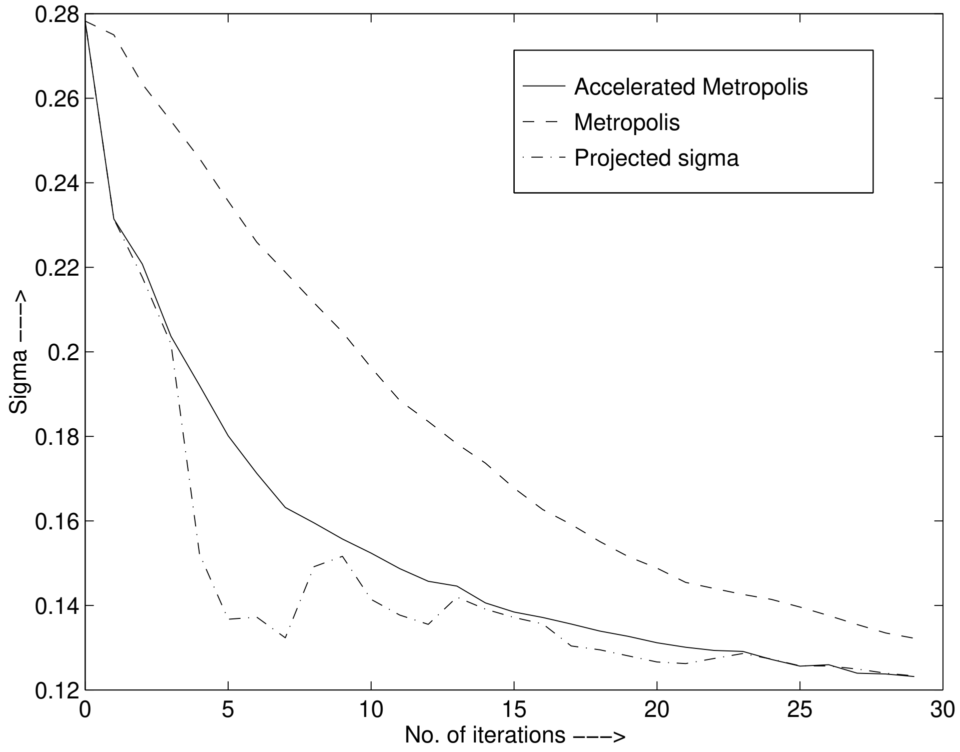
$$p(x) = \frac{1}{\sigma^N Z(1)} \exp\left\{ \frac{1}{p\sigma^p} U(x) \right\}$$

- Problem: We don't know X!
- EM formulation for incomplete data problem

$$\begin{aligned} \sigma^{(k+1)} &= \arg\max_{\sigma} E\left\{ \log p(X|\sigma) | Y = y, \sigma^{(k)} \right\} \\ &= \left(E\left\{ \frac{1}{N} U(X) | Y = y, \sigma^{(k)} \right\} \right)^{1/p} \end{aligned}$$

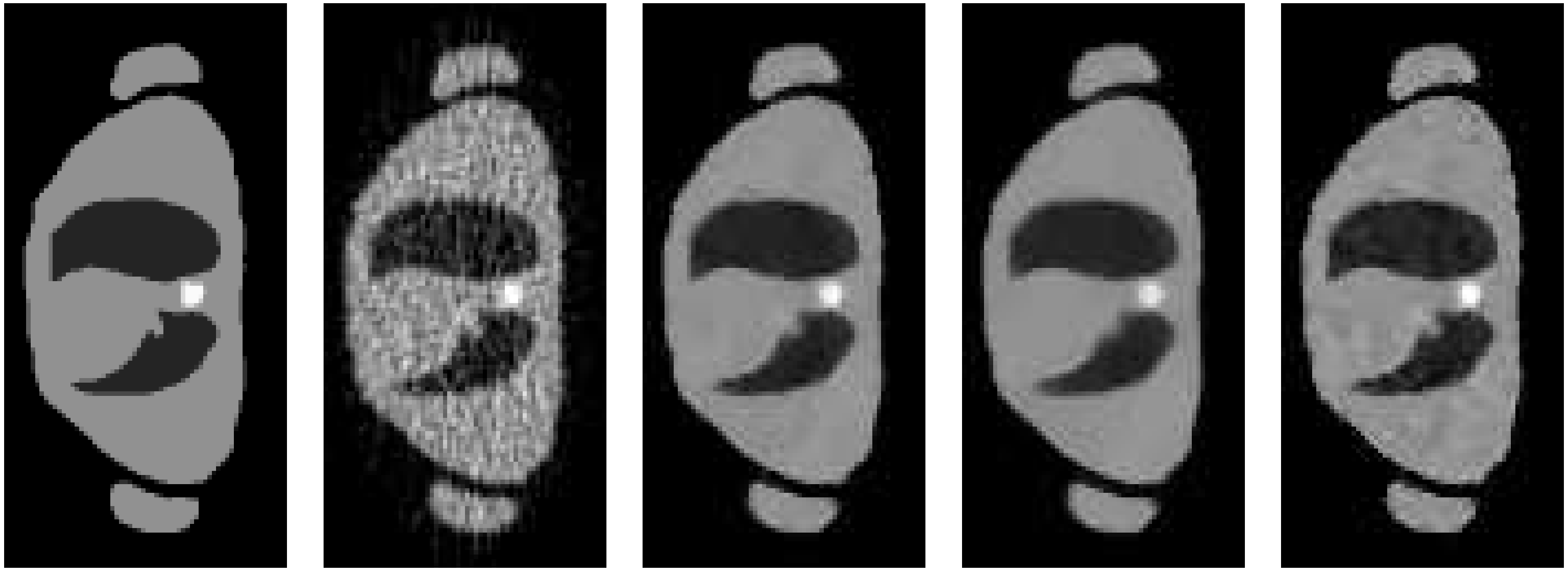
- Iterations converge toward the ML estimate.
- Expectations may be computed using stochastic simulation.

Example of Estimation of σ from Tomographic Data



- The above plot shows the EM updates for σ for the emission phantom modeled by a GGMRF prior ($p = 1.1$) using conventional Metropolis (CM) method, accelerated Metropolis (AM) and the extrapolation method. The parameter s denotes the standard deviation of the symmetric transition distribution for the CM method.

Example of Tomographic Reconstructions

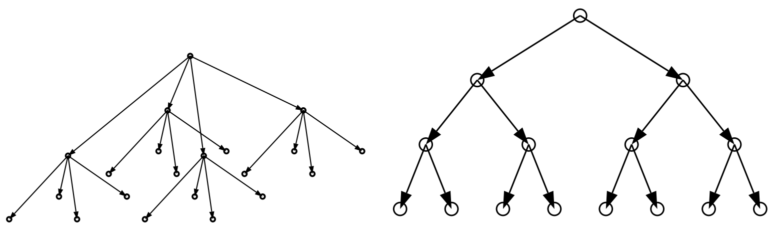


a	b	c	d	e

- (a) Original transmission phantom and (b) CBP reconstruction. Reconstructed transmission phantom using GGMRF prior with $p = 1.1$ The scale parameter σ is (c) $\hat{\sigma}_{ML} \approx \hat{\sigma}_{CBP}$, (d) $\frac{1}{2}\hat{\sigma}_{ML}$, and (e) $2\hat{\sigma}_{ML}$
- Phantom courtesy of J. Fessler, University of Michigan

Multiscale Stochastic Models

- Generate a Markov chain in scale



- Some references
  - Continuous models[12, 5, 111]
  - Discrete models[29, 111]
- Advantages:
  - Does not require a causal ordering of image pixels
  - Computational advantages of Markov chain versus MRF
  - Allows joint and marginal probabilities to be computed using forward/backward algorithm of HMM's.

Multiscale Stochastic Models for Continuous State Estimation

- Theory of 1-D systems can be extended to multiscale trees[6, 7].
- Can be used to efficiently estimate optical flow[111].
- These models can approximate MRF's[112].
- The structure of the model allows exact calculation of log likelihoods for texture segmentation[113].

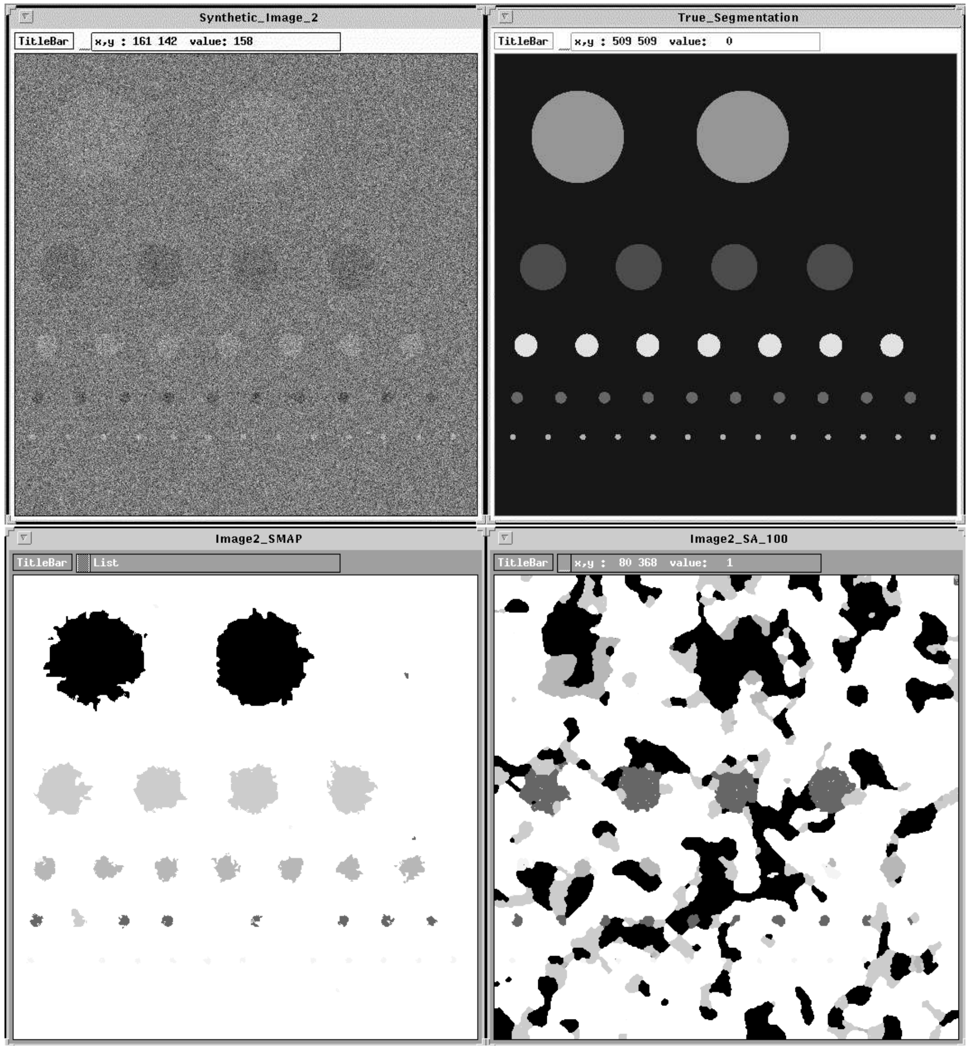
Multiscale Stochastic Models for Segmentation[29]

- Multiscale model results in non-iterative segmentation
- Sequential MAP (SMAP) criteria minimizes size of largest misclassification.
- Computational comparison

	Replacements per pixel				
	SMAP	SMAP + par. est.	SA 500	SA 100	ICM
image1	1.33	3.13	504	105	28
image2	1.33	3.55	506	108	28
image3	1.33	3.14	505	104	10

Segmentation of Synthetic Test Image

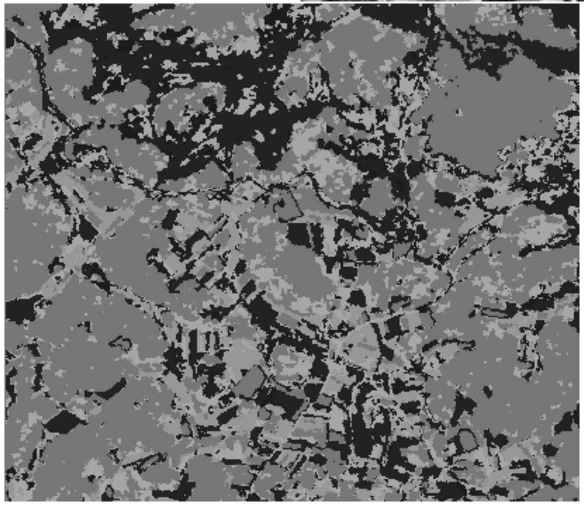
Synthetic Image | Correct Segmentation

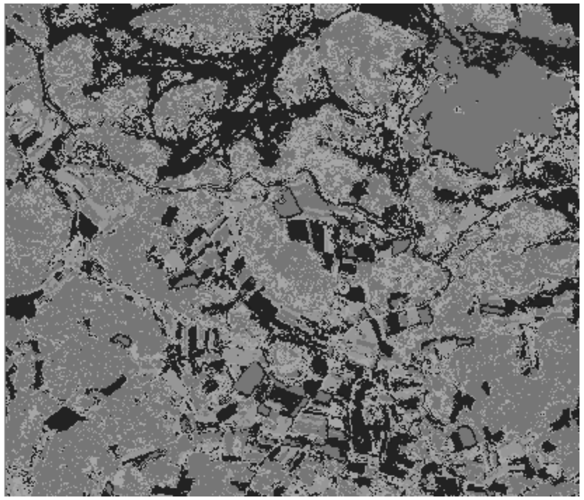


SMAP | 100 Iterations of SA

Multispectral Spot Image Segmentation

SPOT image





SMAP

Maximum Likelihood

High Level Image Models

- MRF's have been used to
  - model the relative location of objects in a scene[119].
  - model relational constraints for object matching problems[109].
- Multiscale stochastic models
  - have been used to model complex assemblies for automated inspection[166].
  - have been used to model 2-D patterns for application in image search[154].

References

[1] M. Aitkin and D. B. Rubin. Estimation and hypothesis testing in finite mixture models. *Journal of the Royal Statistical Society B*, 47(1):67–75, 1985.

[2] H. Akaike. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.*, AC-19(6):716–723, December 1974.

[3] S. Alliney and S. A. Ruzinsky. An algorithm for the minimization of mixed l_1 and l_2 norms with application to Bayesian estimation. *IEEE Trans. on Signal Processing*, 42(3):618–627, March 1994.

[4] P. Barone, A. Frigessi, and M. Piccioni, editors. *Stochastic models, statistical methods, and algorithms in image analysis.* Springer-Verlag, Berlin, 1992.

[5] M. Basseville, A. Benveniste, K. Chou, S. Golden, and R. Nikoukhah. Modeling and estimation of multiresolution stochastic processes. *IEEE Trans. on Information Theory*, 38(2):766–784, 1992.

[6] M. Basseville, A. Benveniste, and A. Willsky. Multiscale autoregressive processes, part i: Schur-Levinson parametrizations. *IEEE Trans. on Signal Processing*, 40(8):1915–1934, 1992.

[7] M. Basseville, A. Benveniste, and A. Willsky. Multiscale autoregressive processes, part ii: Lattice structures for whitening and modeling. *IEEE Trans. on Signal Processing*, 40(8):1915–1934, 1992.

[8] L. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Statistics*, 41(1):164–171, 1970.

[9] L. E. Baum and T. Petrie. Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Statistics*, 37:1554–1563, 1966.

[10] F. Beckman. The solution of linear equations by the conjugate gradient method. In A. Ralston, H. Wilf, and K. Enslein, editors, *Mathematical Methods for Digital Computers*. Wiley, 1960.

[11] M. G. Bello. A combined Markov random field and wave-packet transform-based approach for image segmentation. *IEEE Trans. on Image Processing*, 3(6):834–846, November 1994.

[12] A. Benveniste, R. Nikoukhah, and A. Willsky. Multiscale system theory. In *Proceedings of the 29^{th} Conference on Decision and Control*, volume 4, pages 2484–2489, Honolulu, Hawaii, December 5-7 1990.

[13] J. Besag. Nearest-neighbour systems and the auto-logistic model for binary data. *Journal of the Royal Statistical Society B*, 34(1):75–83, 1972.

[14] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society B*, 36(2):192–236, 1974.

[15] J. Besag. Efficiency of pseudolikelihood estimation for simple Gaussian fields. *Biometrica*, 64(3):616–618, 1977.

[16] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society B*, 48(3):259–302, 1986.

[17] J. Besag. Towards Bayesian image analysis. *Journal of Applied Statistics*, 16(3):395–407, 1989.

[18] J. E. Besag and P. A. P. Moran. On the estimation and testing of spatial interaction in Gaussian lattice processes. *Biometrika*, 62(3):555–562, 1975.

[19] A. Blake. Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(1):2–30, January 1989.

[20] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, Cambridge, Massachusetts, 1987.

[21] C. A. Bouman and B. Liu. A multiple resolution approach to regularization. In *Proc. of SPIE Conf. on Visual Communications and Image Processing*, pages 512–520, Cambridge, MA, Nov. 9-11 1988.

[22] C. A. Bouman and B. Liu. Segmentation of textured images using a multiple resolution approach. In *Proc. of IEEE Int'l Conf. on Acoust., Speech and Sig. Proc.*, pages 1124–1127, New York, NY, April 11-14 1988.

[23] C. A. Bouman and B. Liu. Multiple resolution segmentation of textured images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13:99–113, Feb. 1991.

[24] C. A. Bouman and K. Sauer. An edge-preserving method for image reconstruction from integral projections. In *Proc. of the Conference on Information Sciences and Systems*, pages 382–387, The Johns Hopkins University, Baltimore, MD, March 20-22 1991.

[25] C. A. Bouman and K. Sauer. A generalized Gaussian image model for edge-preserving map estimation. *IEEE Trans. on Image Processing*, 2:296–310, July 1993.

[26] C. A. Bouman and K. Sauer. Maximum likelihood scale estimation for a class of Markov random fields. In *Proc. of IEEE Int'l Conf. on Acoust., Speech and Sig. Proc.*, volume 5, pages 537–540, Adelaide, South Australia, April 19-22 1994.

[27] C. A. Bouman and K. Sauer. A unified approach to statistical tomography using coordinate descent optimization. *IEEE Trans. on Image Processing*, 5(3):480–492, March 1996.

[28] C. A. Bouman and M. Shapiro. Multispectral image segmentation using a multiscale image model. In *Proc. of IEEE Int'l Conf. on Acoust., Speech and Sig. Proc.*, pages III–565 – III–568, San Francisco, California, March 23-26 1992.

[29] C. A. Bouman and M. Shapiro. A multiscale random field model for Bayesian image segmentation. *IEEE Trans. on Image Processing*, 3(2):162–177, March 1994.

[30] A. Brandt. *Multigrid Techniques: 1984 Guide with Applications to Fluid Dynamics*, volume Nr. 85. GMD-Studien, 1984. ISBN:3-88457-081-1.

[31] W. Briggs. *A Multigrid Tutorial.* Society for Industrial and Applied Mathematics, Philadelphia, 1987.

[32] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud. Two deterministic half-quadratic reqularization algorithms for computed imaging. In *Proc. of IEEE Int'l Conf. on Image Proc.*, volume 2, pages 168–176, Austin, TX, November, 13-16 1994.

[33] R. Chellappa and S. Chatterjee. Classification of textures using Gaussian Markov random fields. *IEEE Trans. on Acoust. Speech and Signal Proc.*, ASSP-33(4):959–963, August 1985.

[34] K. Chou, S. Golden, and A. Willsky. Modeling and estimation of multiscale stochastic processes. In *Proc. of IEEE Int'l Conf. on Acoust., Speech and Sig. Proc.*, pages 1709–1712, Toronto, Canada, May 14-17 1991.

[35] B. Claus and G. Chartier. Multiscale signal processing: Isotropic random fields on homogeneous trees. *IEEE Trans. on Circ. and Sys.:Analog and Dig. Signal Proc.*, 41(8):506–517, August 1994.

[36] F. Cohen and D. Cooper. Simple parallel hierarchical and relaxation algorithms for segmenting noncausal Markovian random fields. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-9(2):195–219, March 1987.

[37] F. S. Cohen and Z. Fan. Maximum likelihood unsupervised texture image segmentation. *CVGIP:Graphical Models and Image Proc.*, 54(3):239–251, May 1992.

[38] F. S. Cohen, Z. Fan, and M. A. Patel. Classification of rotated and scaled textured images using Gaussian Markov random field models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13(2):192–202, February 1991.

[39] M. L. Comer and E. J. Delp. Parameter estimation and segmentation of noisy or textured images using the EM algorithm and MPM estimation. In *Proc. of IEEE Int'l Conf. on Image Proc.*, volume II, pages 650–654, Austin, Texas, November 1994.

[40] M. L. Comer and E. J. Delp. Multiresolution image segmentation. In *Proc. of IEEE Int'l Conf. on Acoust., Speech and Sig. Proc.*, pages 2415–2418, Detroit, Michigan, May 1995.

[41] D. B. Cooper. Maximum likelihood estimation of Markov process blob boundaries in noisy images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-1:372–384, 1979.

[42] D. B. Cooper, H. Elliott, F. Cohen, L. Reiss, and P. Symosek. Stochastic boundary estimation and object recognition. *Comput. Vision Graphics and Image Process.*, 12:326–356, April 1980.

[43] R. Cristi. Markov and recursive least squares methods for the estimation of data with discontinuities. *IEEE Trans. on Acoust. Speech and Signal Proc.*, 38(11):1972–1980, March 1990.

[44] G. R. Cross and A. K. Jain. Markov random field texture models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-5(1):25–39, January 1983.

[45] E. *Ü*. Mumcuoğlu, R. Leahy, S. R. Cherry, and Z. Zhou. Fast gradient-based methods for Bayesian reconstruction of transmission and emission pet images. *IEEE Trans. on Medical Imaging*, 13(4):687–701, December 1994.

[46] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, 39(1):1–38, 1977.

[47] H. Derin and W. Cole. Segmentation of textured images using Gibbs random fields. *Comput. Vision Graphics and Image Process.*, 35:72–98, 1986.

[48] H. Derin, H. Elliot, R. Cristi, and D. Geman. Bayes smoothing algorithms for segmentation of binary images modeled by Markov random fields. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-6(6):707–719, November 1984.

[49] H. Derin and H. Elliott. Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-9(1):39–55, January 1987.

[50] H. Derin and P. A. Kelly. Discrete-index Markov-type random processes. *Proc. of the IEEE*, 77(10):1485–1510, October 1989.

[51] H. Derin and C. Won. A parallel image segmentation algorithm using relaxation with varying neighborhoods and its mapping to array processors. *Comput. Vision Graphics and Image Process.*, 40:54–78, October 1987.

[52] R. W. Dijkerman and R. R. Mazumdar. Wavelet representations of stochastic processes and multiresolution stochastic models. *IEEE Trans. on Signal Processing*, 42(7):1640–1652, July 1994.

[53] P. C. Doerschuk. Bayesian signal reconstruction, Markov random fields, and X-ray crystallography. *J. Opt. Soc. Am. A*, 8(8):1207–1221, 1991.

[54] R. Dubes and A. Jain. Random field models in image analysis. *Journal of Applied Statistics*, 16(2):131–164, 1989.

[55] R. Dubes, A. Jain, S. Nadabar, and C. Chen. MRF model-based algorithms for image segmentation. In *Proc. of the 10^th Internat. Conf. on Pattern Recognition*, pages 808–814, Atlantic City, NJ, June 1990.

[56] I. M. Elfadel and R. W. Picard. Gibbs random fields, cooccurrences and texture modeling. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16(1):24–37, January 1994.

[57] H. Elliott, D. B. Cooper, F. S. Cohen, and P. F. Symosek. Implementation, interpretation, and analysis of a suboptimal boundary finding algorithm. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-4(2):167–182, March 1982.

[58] W. Enkelmann. Investigations of multigrid algorithms. *Comput. Vision Graphics and Image Process.*, 43:150–177, 1988.

[59] J. Fessler and A. Hero. Space-alternating generalized expectation-maximization algorithms. *IEEE Trans. on Acoust. Speech and Signal Proc.*, 42(10):2664–2677, October 1994.

[60] N. P. Galatsanos and A. K. Katsaggelos. Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation. *IEEE Trans. on Image Processing*, 1(3):322–336, July 1992.

[61] S. B. Gelfand and S. K. Mitter. Recursive stochastic algorithms for global optimization in r^d. *SIAM Journal on Control and Optimization*, 29(5):999–1018, September 1991.

[62] S. B. Gelfand and S. K. Mitter. Metropolis-type annealing algorithms for global optimization in r^d. *SIAM Journal on Control and Optimization*, 31(1):111–131, January 1993.

[63] S. B. Gelfand and S. K. Mitter. On sampling methods and annealing algorithms. In R. Chellappa and A. Jain, editors, *Markov Random Fields: Theory and Applications*, pages 499–515. Academic Press, Inc., Boston, 1993.

[64] D. Geman. Bayesian image analysis by adaptive annealing. In *Digest 1985 Int. Geoscience and remote sensing symp.*, Amherst, MA, October 7-9 1985.

[65] D. Geman, S. Geman, C. Graffigne, and P. Dong. Boundary detection by constrained optimization. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(7):609–628, July 1990.

[66] D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(3):367–383, 1992.

[67] D. Geman, G. Reynolds, and C. Yang. Stochastic algorithms for restricted image spaces and experiments in deblurring. In R. Chellappa and A. Jain, editors, *Markov Random Fields: Theory and Applications*, pages 39–68. Academic Press, Inc., Boston, 1993.

[68] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-6:721–741, Nov. 1984.

[69] S. Geman and C. Graffigne. Markov random field image models and their applications to computer vision. In *Proc. of the Intl Congress of Mathematicians*, pages 1496–1517, Berkeley, California, 1986.

[70] S. Geman and D. McClure. Bayesian images analysis: An application to single photon emission tomography. In *Proc. Statist. Comput. sect. Amer. Stat. Assoc.*, pages 12–18, Washington, DC, 1985.

[71] S. Geman and D. McClure. Statistical methods for tomographic image reconstruction. *Bull. Int. Stat. Inst.*, LII-4:5–21, 1987.

[72] B. Gidas. A renormalization group approach to image processing problems. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(2):164–180, February 89.

[73] J. Goutsias. Unilateral approximation of Gibbs random field images. *CVGIP:Graphical Models and Image Proc.*, 53(3):240–257, May 1991.

[74] A. Gray, J. Kay, and D. Titterington. An empirical study of the simulation of various models used for images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16(5):507–513, May 1994.

[75] P. J. Green. Bayesian reconstruction from emission tomography data using a modified EM algorithm. *IEEE Trans. on Medical Imaging*, 9(1):84–93, March 1990.

[76] P. J. Green and X. liang Han. Metropolis methods, Gaussian proposals and antithetic variables. In P. Barone, A. Frigessi, and M. Piccioni, editors, *Stochastic Models, Statistical methods, and Algorithms in Image Analysis*, pages 142–164. Springer-Verlag, Berlin, 1992.

[77] M. I. Gurelli and L. Onural. On a parameter estimation method for Gibbs-Markov random fields. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16(4):424–430, April 1994.

[78] W. Hackbusch. *Multi-Grid Methods and Applications.* Sprinter-Verlag, Berlin, 1980.

[79] K. Hanson and G. Wecksung. Bayesian approach to limited-angle reconstruction computed tomography. *J. Opt. Soc. Am.*, 73(11):1501–1509, November 1983.

[80] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[81] T. Hebert and R. Leahy. A generalized EM algorithm for 3-d Bayesian reconstruction from Poisson data using Gibbs priors. *IEEE Trans. on Medical Imaging*, 8(2):194–202, June 1989.

[82] T. J. Hebert and K. Lu. Expectation-maximization algorithms, null spaces, and MAP image restoration. *IEEE Trans. on Image Processing*, 4(8):1084–1095, August 1995.

[83] F. Heitz and P. Bouthemy. Multimodal estimation of discontinuous optical flow using Markov random fields. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15(12):1217–1232, December 1993.

[84] G. T. Herman, A. R. De Pierro, and N. Gai. On methods for maximum a posteriori image reconstruction with normal prior. *J. Visual Comm. Image Rep.*, 3(4):316–324, December 1992.

[85] P. Huber. *Robust Statistics.* John Wiley & Sons, New York, 1981.

[86] B. Hunt. The application of constrained least squares estimation to image restoration by digital computer. *IEEE Trans. on Comput.*, c-22(9):805–812, September 1973.

[87] B. Hunt. Bayesian methods in nonlinear digital image restoration. *IEEE Trans. on Comput.*, c-26(3):219–229, 1977.

[88] J. Hutchinson, C. Koch, J. Luo, and C. Mead. Computing motion using analog and binary resistive networks. *Computer*, 21:53–63, March 1988.

[89] A. Jain. Advances in mathematical models for image processing. *Proc. of the IEEE*, 69:502–528, May 1981.

[90] B. D. Jeffs and M. Gunsay. Restoration of blurred star field images by maximally sparse optimization. *IEEE Trans. on Image Processing*, 2(2):202–211, April 1993.

[91] F.-C. Jeng and J. W. Woods. Compound Gauss-Markov random fields for image estimation. *IEEE Trans. on Signal Processing*, 39(3):683–697, March 1991.

[92] B. Jeon and D. Landgrebe. Classification with spatio-temporal interpixel class dependency contexts. *IEEE Trans. on Geoscience and Remote Sensing*, 30(4), JULY 1992.

[93] S. R. Kadaba, S. B. Gelfand, and R. L. Kashyap. Bayesian decision feedback for segmentation of binary images. In *Proc. of IEEE Int'l Conf. on Acoust., Speech and Sig. Proc.*, pages 2543–2546, Detroit, Michigan, May 8-12 1995.

[94] R. Kashyap and R. Chellappa. Estimation and choice of neighbors in spatial-interaction models of images. *IEEE Trans. on Information Theory*, IT-29(1):60–72, January 1983.

[95] R. Kashyap, R. Chellappa, and A. Khotanzad. Texture classification using features derived from random field models. *Pattern Recogn. Let.*, 1(1):43–50, October 1982.

[96] R. L. Kashyap and K.-B. Eom. Robust image modeling techniques with an image restoration application. *IEEE Trans. on Acoust. Speech and Signal Proc.*, 36(8):1313–1325, August 1988.

[97] Z. Kato, M. Berthod, and J. Zerubia. Parallel image classification using multiscale Markov random fields. In *Proc. of IEEE Int'l Conf. on Acoust., Speech and Sig. Proc.*, volume 5, pages 137–140, Minneapolis, MN, April 27-30 1993.

[98] A. Katsaggelos. Image identification and restoration based on the expectation-maximization algorithm. *Optical Engineering*, 29(5):436–445, May 1990.

[99] P. Kelly, H. Derin, and K. Hartt. Adaptive segmentation of speckled images using a hierarchical random field model. *IEEE Trans. on Acoust. Speech and Signal Proc.*, 36(10):1628–1641, October 1988.

[100] R. Kindermann and J. Snell. *Markov Random Fields and their Applications.* American Mathematical Society, Providence, 1980.

[101] J. Konrad and E. Dubois. Bayesian estimation of motion vector fields. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(9):910–927, September 1992.

[102] R. Lagendijk, A. Tekalp, and J. Biemond. Maximum likelihood image and blur identification: A unifying approach. *Optical Engineering*, 29(5):422–435, May 1990.

[103] R. L. Lagendijk, J. Biemond, and D. E. Boekee. Identification and restoration of noisy blurred images using the expectation-maximization algorithm. *IEEE Trans. on Acoust. Speech and Signal Proc.*, 38(7):1180–1191, July 1990.

[104] S. Lakshmanan and H. Derin. Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(8):799–813, August 1989.

[105] S. Lakshmanan and H. Derin. Gaussian Markov random fields at multiple resolutions. In R. Chellappa and A. Jain, editors, *Markov Random Fields: Theory and Applications*, pages 131–157. Academic Press, Inc., Boston, 1993.

[106] S. Lakshmanan and H. Derin. Valid parameter space for 2-d Gaussian Markov random fields. *IEEE Trans. on Information Theory*, 39(2):703–709, March 1993.

[107] K. Lange. Convergence of EM image reconstruction algorithms with Gibbs smoothing. *IEEE Trans. on Medical Imaging*, 9(4):439–446, December 1990.

[108] K. Lange. An overview of Bayesian methods in image reconstruction. In *Proc. of the SPIE Conference on Digital Image Synthesis and Inverse Optics*, volume SPIE-1351, pages 270–287, San Diego, CA, 1990.

[109] S. Z. Li. *Markov Random Field Modeling in Computer Vision.* to be published, 1996.

[110] J. Liu and Y.-H. Yang. Multiresolution color image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16(7):689–700, July 1994.

[111] M. R. Luettgen, W. C. Karl, and A. S. Willsky. Efficient multiscale regularization with applications to the computation of optical flow. *IEEE Trans. on Image Processing*, 3(1), January 1994.

[112] M. R. Luettgen, W. C. Karl, A. S. Willsky, and R. R. Tenney. Multiscale representations of Markov random fields. *IEEE Trans. on Signal Processing*, 41(12), December 1993. special issue on Wavelets and Signal Processing.

[113] M. R. Luettgen and A. S. Willsky. Likelihood calculation for a class of multiscale stochastic models, with application to texture discrimination. *IEEE Trans. on Image Processing*, 4(2):194–207, February 1995.

[114] B. Manjunath, T. Simchony, and R. Chellappa. Stochastic and deterministic networks for texture segmentation. *IEEE Trans. on Acoust. Speech and Signal Proc.*, 38(6):1039–1049, June 1990.

[115] J. Mao and A. K. Jain. Texture classification and segmentation using multiresolution simultaneous autoregressive models. *Pattern Recognition*, 25(2):173–188, 1992.

[116] J. Marroquin, S. Mitter, and T. Poggio. Probabilistic solution of ill-posed problems in computational vision. *Journal of the American Statistical Association*, 82:76–89, March 1987.

[117] S. McCormick, editor. *Multigrid Methods.* Society for Industrial and Applied Mathematics, Philadelphia, 1987.

[118] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equations of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087–1091, 1953.

[119] J. W. Modestino and J. Zhang. A Markov random field model-based approach to image interpretation. In R. Chellappa and A. Jain, editors, *Markov Random Fields: Theory and Applications*, pages 369–408. Academic Press, Inc., Boston, 1993.

[120] H. H. Nguyen and P. Cohen. Gibbs random fields, fuzzy clustering, and the unsupervised segmentation of textured images. *CVGIP:Graphical Models and Image Proc.*, 55(1):1–19, January 1993.

[121] Y. Ogata. A Monte Carlo method for an objective Bayesian procedure. *Ann. Inst. of Statist. Math.*, 42(3):403–433, 1990.

[122] L. Onsager. Crystal statistics i. a two-dimensional model. *Physical Review*, 65:117–149, 1944.

[123] L. Onural. Generating connected textured fractal patterns using Markov random fields. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13(8):819–825, August 1991.

[124] L. Onural, M. B. Alp, and M. I. Gurelli. Gibbs random field model based weight selection for the 2-d adaptive weighted median filter. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16(8):831–837, August 1994.

[125] T. Pappas. An adaptive clustering algorithm for image segmentation. *IEEE Trans. on Signal Processing*, 4:901–914, April 1992.

[126] R. E. Peierls. On Ising's model of ferromagnetism. *Proc. Camb. Phil. Soc.*, 32:477–481, 1936.

[127] R. E. Peierls. Statistical theory of adsorption with interaction between the adsorbed atoms. *Proc. Camb. Phil. Soc.*, 32:471–476, 1936.

[128] P. Perez and F. Heitz. Multiscale Markov random fields and constrained relaxation in low level image analysis. In *Proc. of IEEE Int'l Conf. on Acoust., Speech and Sig. Proc.*, volume 3, pages 61–64, San Francisco, CA, March 23-26 1992.

[129] P. H. Peskun. Optimum Monte-Carlo sampling using Markov chains. *Biometrika*, 60(3):607–612, 1973.

[130] D. Pickard. Asymptotic inference for an ising lattice iii. non-zero field and ferromagnetic states. *J. Appl. Prob.*, 16:12–24, 1979.

[131] D. Pickard. Inference for discrete Markov fields: The simplest nontrivial case. *Journal of the American Statistical Association*, 82:90–96, March 1987.

[132] D. N. Politis. Markov chains in many dimensions. *Adv. Appl. Prob.*, 26:756–774, 1994.

[133] W. Pun and B. Jeffs. Shape parameter estimation for generalized Gaussian Markov random field models used in MAP image restoration. In *29th Asilomar Conference on Signals, Systems, and Computers*, Oct. 29 - Nov. 1 1995.

[134] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. of the IEEE*, 77(2):257–286, February 1989.

[135] L. R. Rabiner and B. H. Juang. An introduction to hidden Markov models. *IEEE Signal Proc. Magazine*, 3(1):4–16, January 1986.

[136] E. Redner and H. Walker. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, 26(2), April 1984.

[137] S. Reeves and R. Mersereau. Optimal estimation of the regularization parameter and stabilizing functional for regularized image restoration. *Optical Engineering*, 29(5):446–454, May 1990.

[138] S. J. Reeves and R. M. Mersereau. Blur identification by the method of generalized cross-validation. *IEEE Trans. on Image Processing*, 1(3):301–321, July 1992.

[139] W. Rey. *Introduction to Robust and Quasi-Robust Statistical Methods*. Springer-Verlag, Berlin, 1980.

[140] E. Rignot and R. Chellappa. Segmentation of polarimetric synthetic aperature radar data. *IEEE Trans. on Image Processing*, 1(3):281–300, July 1992.

[141] B. D. Ripley. *Stochastic Simulation*. John Wiley & Sons, New York, 1987.

[142] J. Rissanen. A universal prior for integers and estimation by minimum description length. *Annals of Statistics*, 11(2):417–431, 1983.

[143] B. Rouchouze, P. Mathieu, T. Gaidon, and M. Barlaud. Motion estimation based on Markov random fields. In *Proc. of IEEE Int'l Conf. on Image Proc.*, volume 3, pages 270–274, Austin, TX, November, 13-16 1994.

[144] S. S. Saquib, C. A. Bouman, and K. Sauer. ML parameter estimation for Markov random fields, with applications to Bayesian tomography. Technical Report TR-ECE 95-24, School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, October 1995.

[145] S. S. Saquib, C. A. Bouman, and K. Sauer. Efficient ML estimation of the shape parameter for generalized Gaussian MRF. In *Proc. of IEEE Int'l Conf. on Acoust., Speech and Sig. Proc.*, volume 4, pages 2229–2232, Atlanta, GA, May 7-10 1996.

[146] K. Sauer and C. Bouman. Bayesian estimation of transmission tomograms using segmentation based optimization. *IEEE Trans. on Nuclear Science*, 39:1144–1152, 1992.

[147] K. Sauer and C. A. Bouman. A local update strategy for iterative reconstruction from projections. *IEEE Trans. on Signal Processing*, 41(2), February 1993.

[148] R. Schultz, R. Stevenson, and A. Lumsdaine. Maximum likelihood parameter estimation for non-Gaussian prior signal models. In *Proc. of IEEE Int'l Conf. on Image Proc.*, volume 2, pages 700–704, Austin, TX, November 1994.

[149] R. R. Schultz and R. L. Stevenson. A Bayesian approach to image expansion for improved definition. *IEEE Trans. on Image Processing*, 3(3):233–242, May 1994.

[150] R. R. Schultz and R. L. Stevenson. Stochastic modeling and estimation of multispectral image data. *IEEE Trans. on Image Processing*, 4(8):1109–1119, August 1995.

[151] L. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Trans. on Medical Imaging*, MI-1(2):113–122, October 1982.

[152] J. Silverman and D. Cooper. Bayesian clustering for unsupervised estimation of surface and texture models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 10(4):482–495, July 1988.

[153] T. Simchony, R. Chellappa, and Z. Lichtenstein. Relaxation algorithms for MAP estimation of gray-level images with multiplicative noise. *IEEE Trans. on Information Theory*, 36(3):608–613, May 1990.

[154] S. Sista, C. A. Bouman, and J. P. Allebach. Fast image search using a multiscale stochastic model. In *Proc. of IEEE Int'l Conf. on Image Proc.*, pages 225–228, Washington, DC, October 23-26 1995.

[155] R. Stevenson and E. Delp. Fitting curves with discontinuities. *Proc. of the first international workshop on robust computer vision*, pages 127–136, October 1-3 1990.

[156] H. Tan, S. Gelfand, and E. Delp. A comparative cost function approach to edge detection. *IEEE Trans. on Systems Man and Cybernetics*, 19(6):1337–1349, December 1989.

[157] T. Taxt, P. Flynn, and A. Jain. Segmentation of document images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(12):1322–1329, December 1989.

[158] D. Terzopoulos. Image analysis using multigrid relaxation methods. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-8(2):129–139, March 1986.

[159] D. Terzopoulos. The computation of visible-surface representations. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 10(4):417–438, July 1988.

[160] C. Therrien. An estimation-theoretic approach to terrain image segmentation. *Comput. Vision Graphics and Image Process.*, 22:313–326, 1983.

[161] C. Therrien, T. Quatieri, and D. Dudgeon. Statistical model-based algorithm for image analysis. *Proc. of the IEEE*, 74(4):532–551, April 1986.

[162] C. W. Therrien, editor. *Decision Estimation and Classification: An Introduction to Pattern Recognition and Related Topics*. John Wiley & Sons, New York, 1989.

[163] A. Thompson, J. Kay, and D. Titterington. A cautionary note about the cross validatory choice. *J. statist. Comput. Simul.*, 33:199–216, 1989.

[164] A. M. Thompson, J. C. Brown, J. W. Kay, and D. M. Titterington. A study of methods of choosing the smoothing parameter in image restoration. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13(4):326–339, 1991.

[165] A. Tikhonov and V. Arsenin. *Solutions of Ill-Posed Problems*. Winston and Sons, New York, 1977.

[166] D. Tretter, C. A. Bouman, K. Khawaja, and A. Maciejewski. A multiscale stochastic image model for automated inspection. *IEEE Trans. on Image Processing*, 4(12):507–517, December 1995.

[167] C. Won and H. Derin. Segmentation of noisy textured images using simulated annealing. In *Proc. of IEEE Int'l Conf. on Acoust., Speech and Sig. Proc.*, pages 14.4.1–14.4.4, Dallas, TX, 1987.

[168] C. S. Won and H. Derin. Unsupervised segmentation of noisy and textured images using Markov random fields. *CVGIP:Graphical Models and Image Proc.*, 54(4):308–328, July 1992.

[169] J. Woods, S. Dravida, and R. Mediavilla. Image estimation using double stochastic Gaussian random field models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-9(2):245–253, March 1987.

[170] C. Wu. On the convergence properties of the EM algorithm. *Annals of Statistics*, 11(1):95–103, 1983.

[171] C.-h. Wu and P. C. Doerschuk. Cluster expansions for the deterministic computation of Bayesian estimators based on Markov random fields. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(3):275–293, March 1995.

[172] C.-h. Wu and P. C. Doerschuk. Texture-based segmentation using Markov random field models and approximate Bayesian estimators based on trees. *J. Math. Imaging and Vision*, 5(4), December 1995.

[173] C.-h. Wu and P. C. Doerschuk. Tree approximations to Markov random fields. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(4):391–402, April 1995.

[174] Z. Wu and R. Leahy. An approximate method of evaluating the joint likelihood for first-order GMRFs. *IEEE Trans. on Image Processing*, 2(4):520–523, October 1993.

[175] J. Zerubia and R. Chellappa. Mean field annealing using compound Gauss-Markov random fields for edge detection and image estimation. *IEEE Trans. on Neural Networks*, 4(4):703–709, July 1993.

[176] J. Zhang. The mean field theory in EM procedures for Markov random fields. *IEEE Trans. on Signal Processing*, 40(10):2570–2583, October 1992.

[177] J. Zhang. The mean field theory in EM procedures for blind Markov random field image restoration. *IEEE Trans. on Image Processing*, 2(1):27–40, January 1993.

[178] J. Zhang and G. G. Hanauer. The application of mean field theory to image motion estimation. *IEEE Trans. on Image Processing*, 4(1):19–33, January 1995.

[179] J. Zhang and J. W. Modestino. A model-fitting approach to cluster validation with application to stochastic model-based image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(10):1009–1017, October 1990.

[180] J. Zhang, J. W. Modestino, and D. A. Langan. Maximum-likelihood parameter estimation for unsupervised stochastic model-based image segmentation. *IEEE Trans. on Image Processing*, 3(4):404–420, July 1994.

[181] M. Zhang, R. Haralick, and J. Campbell. Multispectral image context classification using stochastic relaxation. *IEEE Trans. on Systems Man and Cybernetics*, vol. 20(1):128–140, February 1990.