Lecture 20

Application : Segmentation

X - Class labels
Y - Image Pixels

$$X_s = \begin{cases} 1 \rightarrow \text{object} \\ 0 \rightarrow \text{background} \end{cases}$$

$$p(Y|X) = \prod_{s \in S} p(Y_s | X_s)$$

$p(Y_s | 0) \leftarrow$ distribution of background
$p(Y_s | 1) \leftarrow$ distribution of object

$X \leftarrow$ Ising model MRF

$$\log p(Y, X) = \log p(Y|X) + \log p(X)$$

$$\log p(Y|X) = \sum_{s \in S} \log p(Y_s | X_s)$$

$$\log p(X) = \log \left\{ \frac{1}{Z} \exp \left\{ - \sum_{\{i,j\} \in C} \beta\, t(X_i, X_j) \right\} \right\}$$

$$= - \sum_{\{i,j\} \in C} \beta\, t(X_i, X_j) + C$$

Define

$$l(y_s \mid x_s) \triangleq -\log p(y_s \mid x_s)$$

$$\log p(x, y) = -\sum_{s \in S} l(y_s \mid x_s) - \sum_{\{i,j\} \in C} \beta t(x_i, x_j)$$

$$\hat{X}_{MAP} = \arg\min_{x} \left\{ \sum_{s \in S} l(y_s \mid x_s) + \sum_{\{i,j\} \in C} \beta t(x_i, x_j) \right\}$$

Find the segmentation $\hat{X}_{MAP}$ which

1) Fits data
2) Minimizes boundary length

Optimization

Let $U(x) = \sum_{s \in S} l(y_s | x_s) + \sum_{\{i,j\} \in C} \beta t(x_i, x_j)$

By Bayes rule

$$p(x|y) = \frac{1}{2} \exp\{-U(x)\}$$

$$p(x_s | x_i\ i \neq s, y) = \frac{\exp\{-U(x_s, x_i\ i \neq s)\}}{\sum_{k=1}^{N} \exp\{-U(x_s = k, x_i\ i \neq s)\}}$$

The conditional "mode" for x_s is

$$\hat{x}_s = \arg\max_{x_s} p(x_s | x_i\ i \neq s, y)$$

$$= \arg\max_{x_s} \exp\{-U(x_s, x_i\ i \neq s)\}$$

$$= \arg\max_{x_s} -U(x_s, x_i\ i \neq s)$$

This is coordinate ascent / Gauss-Seidel at location s.

Hence the name:
Iterated Conditional Modes (ICM)

ICM: Iteratively apply replacement rule.

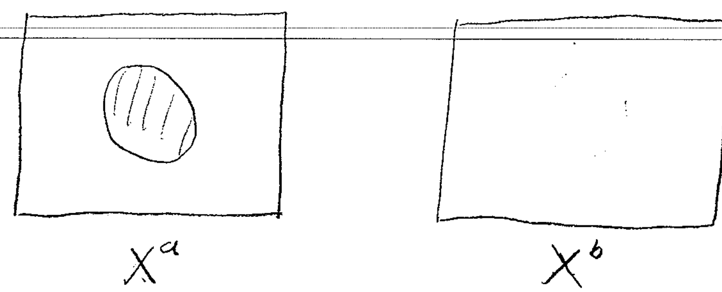
$$X_s^{(k+1)} = \arg\min_{x_s} \left\{ U(x_s, x_i \; i \neq s) \right\}$$

$$X_s^{(k+1)} = \arg\min_{K} \left\{ l(y_s | K) + \beta \sum_{i \in \partial s} t(K, X_i^{(k)}) \right\}$$

$$= \arg\min_{K} \left\{ l(y_s | K) + \beta V(K, x_{\partial s}) \right\}$$

Properties:

1) Convergence $\rightarrow$ yes
2) Global Convergence $\rightarrow$ no



$$U(X^b) < U(X^a)$$

but locall operation will not get you from X^a to X^b

Lecture 21

Simmulated Annealing

Consider the MRF $\tilde{X}$ with distribution

$$p(\tilde{x}) = \frac{1}{Z} e^{-\frac{1}{T} U(\tilde{x})}$$

If 1) $T = 1 \Rightarrow$ posterior distribution of X given Y

2) $T = \infty \Rightarrow$ uniform on $[1, 2]^N$

3) $T \to 0 \Rightarrow$ Distribution concentrated around

$$\hat{X}_{MAP} = \arg\min_{X} U(x)$$

First Idea

1) pick $T = .0001$
2) Generate Markov chain of $\tilde{X}$'s using Gibbs sampler
3) Wait for stationary behavior
4) pick any sample segmentation

Bad idea: as $T \to 0$ 3) takes longer

Second Idea (Simmulated annealing)

1) Choose an annealing schedule
T_n for iteration n
$T_{n+1} < T_n$
2) Use Gibbs Sampler algorithm with temperature T_n at iteration n
3) Wait until T_n get small and pick a sample segmentation

Replacement distribution

$$W_s \sim \frac{\exp\left\{-\frac{1}{T} U(x_s, x_i \; i \neq s)\right\}}{\sum_{k=1}^{M} \exp\left\{-\frac{1}{T} U(x_s = k, x_i \; i \neq s)\right\}}$$

$$= \frac{\exp\left\{-\frac{1}{T}\left(l(Y_s | X_s) + \beta \; V(X_s, X_{\partial s})\right)\right\}}{\sum_{k=1}^{M} \exp\left\{-\frac{1}{T}\left(l(Y_s | k) + \beta \; V(k, X_{\partial s})\right)\right\}}$$

$$= p(\tilde{x}_s \,/\, \tilde{x}_i \; i \neq s)$$

If $T \to 0 \Rightarrow$ ICM!

Optimal
Annealing Schedule

N - # of pixels

$\Delta = \underset{x}{\arg\max}\, U(x) - \underset{x}{\arg\min}\, U(x)$

K - # of pixels replaced + 1

$$T(K) = \frac{N\Delta}{\log K}$$

Example

If $\Delta = 1$ $N = 250{,}000$

$$\frac{N\Delta}{\log K} = \frac{1}{2}$$

$$K = e^{500{,}000}$$

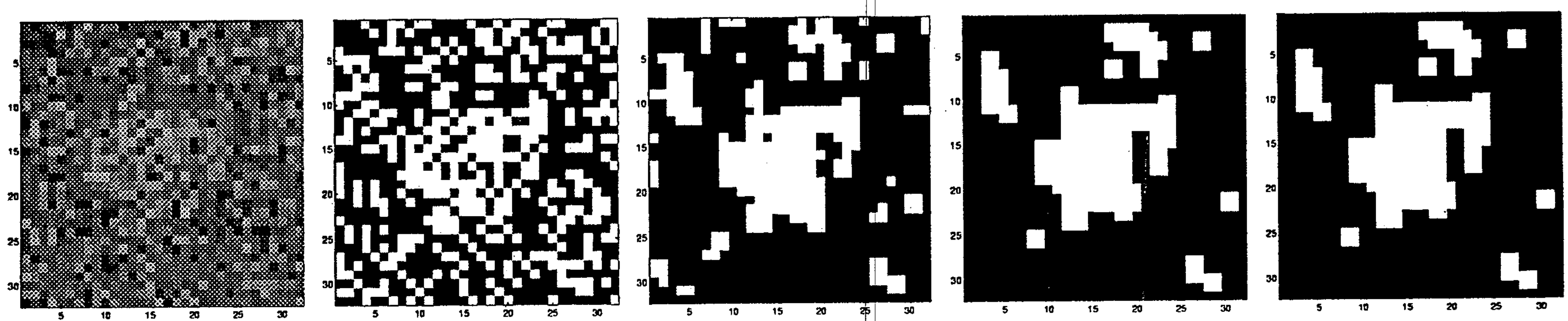
"a long time"

Typical annealing schedule

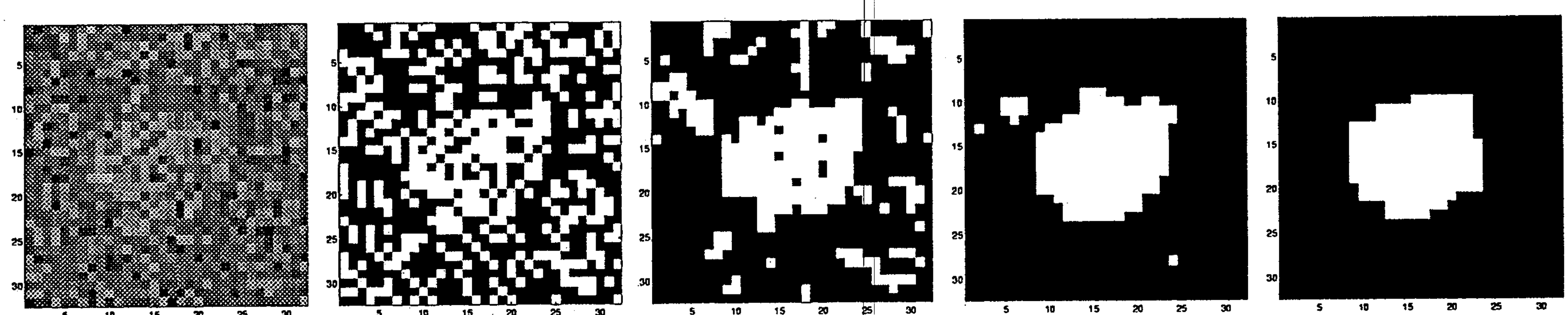
$$\frac{1}{T(K+1)} = \frac{1}{T(K)} + \Delta$$

Segmentation Example

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



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



Other Bayesian Estimates

Y - observed image
X - segmentation
$\hat{X}$ - estimate of segmentation

$C(X, \hat{x})$ = cost of choosing $\hat{x}$ if X is the correct answer.

We would like to minimize risk in our choice.

$$E[C(X,x) \mid Y=y] = R(y, x)$$

$$\hat{X} = \underset{x}{\operatorname{argmin}}\ R(y,x)$$

$$= \underset{x}{\operatorname{argmin}}\ E[C(X,x) \mid Y=y]$$

Choose

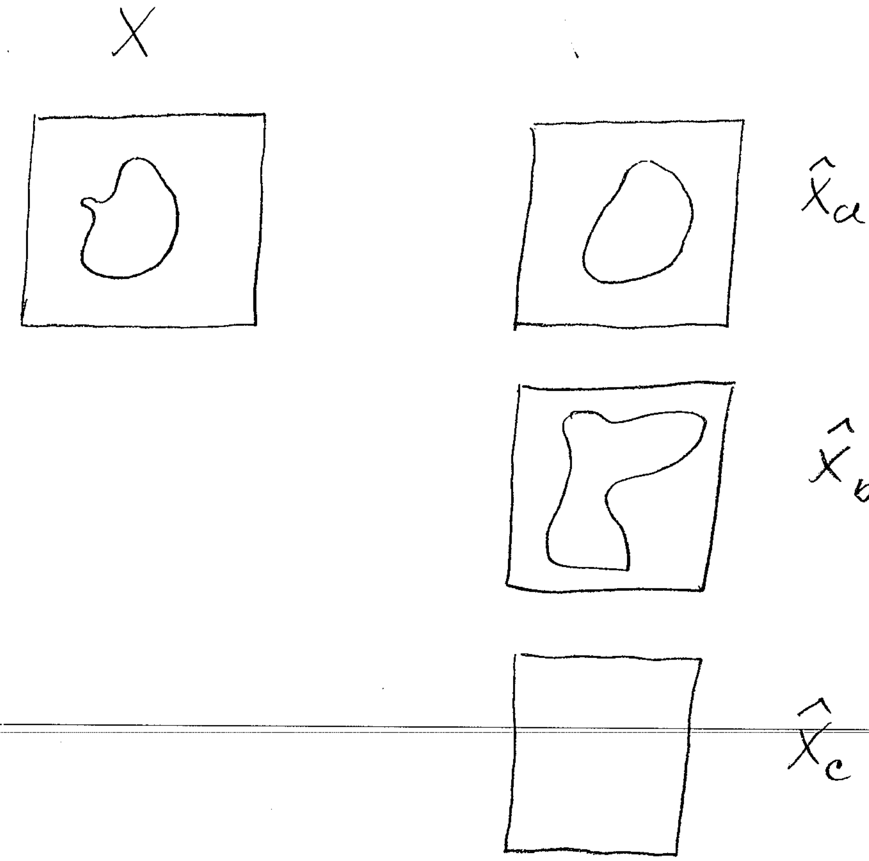
$$C_{MAP}(X, \hat{x}) = \begin{cases} 1 & X \neq \hat{x} \\ 0 & X = \hat{x} \end{cases}$$

$$E[C_{MAP}(X, x) \mid Y=y] = R(x, y)$$

$$= 1 - P(X=x \mid Y=y)$$

$$\hat{X} = \operatorname{argmin} \{1 - P(X=x \mid Y=y)\}$$
$$= \operatorname{argmax} P(X=x \mid Y=y) = \hat{X}_{MAP}$$

Problem with MAP estimate



All these have the same "cost"
But clearly $\hat{X}_a$ is the best.

Lecture 22

Choose

$$C_{MPM}(X, \hat{x}) = \sum_{s \in S} t(X_s - \hat{x}_s)$$

$$R(\hat{x}, y) = E[C_{MPM}(X, \hat{x}) \mid Y = y]$$

$$= \sum_{s \in S} E[t(X_s - \hat{x}_s) \mid Y = y]$$

$$= \sum_{s \in S} \{1 - P(X_s = \hat{x}_s \mid Y = y)\}$$

$$\hat{X} = \arg\max_{x} \sum_{s \in S} \{1 - P(X_s = x \mid Y = y)\}$$

$$= \arg\min_{x} \sum_{s} P(X_s = x_s \mid Y = y)$$

$$\hat{x}_s = \arg\min_{x_s} P(X_s = x_s \mid Y = y)$$

$\hat{x}_s$ is the maximizer of the posterior marginal distribution (MPM)

How do we compute X_{MPM}?

$$U(x) = \sum_{s \in S} l(y|x_s) + \sum_{\{i,j\} \in C} \beta\, t(x_i, x_j)$$

$$p(x|y) = \frac{1}{z} e^{-U(x)}$$

1) $$p(x_s | y) = \underbrace{\sum_{x_i\ i \neq s} p(x_s, x_i\ i \neq s, y)}_{\text{impossible to compute}}$$

2) Generate $X^{(1)}, X^{(2)} \ldots$ from $p(x|y)$ using Gibbs sampler

Then by ergodic property of Markov Chain

$$p(x_s | y) = \lim_{N \to \infty} \frac{1}{N} \underbrace{\sum_{k=0}^{N} \delta(X_s^{(k)} - x_s)}_{\#\text{ of times } X_s^{(k)} = x_s}$$

Pick
$$\hat{x}_s = \arg\max_{x_s} p(x_s | y)$$

Problems with MPM

1) If β is large then N must be VERY large to achieve ergodic behavior. Computationally expensive

2) CMPM requires no "consistancy" in the solutions. Each pixel is choosen individually

Multi Scale Techniques

Idea

Make large scale "global" moves in search algorithms

1) Much faster computation
2) More sensitive to large scale patterns.

Define

$$t_1(X) = \sum_{\{i,j\} \in C} t(x_i, x_j)$$

= # of horizontal/vertical pairs of diferent value.

$$U(x) = \beta t_1(X)$$

1) Restrict X to have fixed value on 2×2 blocks

$$\begin{array}{cccc} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{array} \qquad \begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array}$$

$$X \qquad\qquad X^{(2)}$$

2) Denote the new "decimated" field by $X^{(2)}$

$$u(x) = \beta t_1(x) = 2\beta t_1(X^{(2)})$$

$$\beta^{(2)} \triangleq 2\beta$$
$$t_1 \triangleq t_1(X)$$
$$t_1^{(2)} \triangleq t_1(X^{(2)})$$

$$U(X) = \beta^{(2)} t_1^{(2)} = U^{(2)}(X^{(2)})$$

MRS algorithm

recursive $\longrightarrow$ 1. Use MRS algorithm to solve

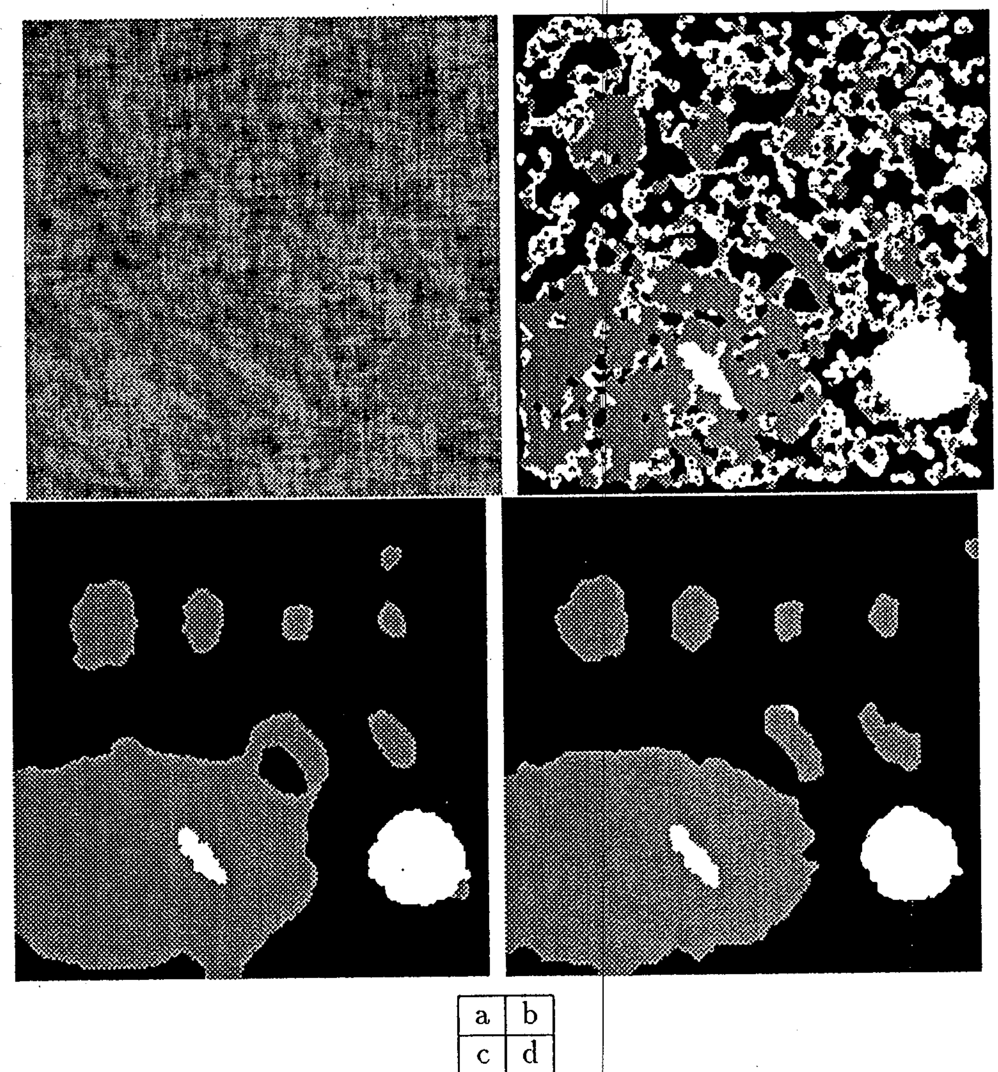
$$\hat{X}^{(2)} = \arg\min_{X^{(2)}} U^{(2)}(X^{(2)})$$

2. Replicate pixels to form $X^{(1)}$ from $X^{(2)}$

3. Use ICM with the starting point $X^{(1)}$ to approximately solve

$$\hat{X}^{(1)} = \arg\min_{X^{(1)}} U(X^{(1)})$$

Texture Segmentation Example



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