(5) Weak Membership

Bayesian Model - Humford-Shah

\[ P(D|E, F) = P(F|E) \]

- log \[ P(D|E, F) = 2 \int \left\{ 2(x, y) - p(x, y) \right\}^2 dx \, dy \]
- log \[ P(F|E) = \alpha \int 1 \mid F(x, y) \mid^2 dx \, dy + \int ds \]

Bayes Posteriori

\[ P(F, E|D) = \frac{P(D|E, F) P(F|E)}{P(D)} \]

Maximizing \[ P(F, E|D) \]

with respect to \( E \) and \( F \)

(Maximum A Posteriori Estimation)

is equivalent to minimizing \[ E_{F, E} \mid F | D | \] with \( F, E \).

Some technical problems here - how to define probability distributions for functionals?

Prior assumption is that images are piecewise smooth - often wrong for natural images.

Also imaging model assumes additive Gaussian noise.

Also, MLE interpretation.
Weak Membrane

Ceman & Ceman

\[ D = \{ d_i \}, \quad F = \{ f_i \}, \quad L = \{ l_i \} \] depend on a lattice.

Markov Random Field.

The line processes \( x \) lie between the image pixels \( \{ d_i \} \) at \( \theta \).

\[
P(D | F) = \prod_{i=1}^{N} P(d_i | f_i) = \prod_{i=1}^{N} \left( \frac{1}{\sqrt{2\pi\theta}} \right) e^{-\frac{(d_i - f_i)^2}{2\theta}}
\]

\[
P(F, \lambda) = e^{-E[F, \lambda]}
\]

\[
E[F, \lambda] = A \sum_{i=1}^{N-1} (f_{i+1} - f_i)^2 (1 - l_i) + B \sum_{i=1}^{N-1} l_i
\]

Bayes

\[
P(F, l | D) = \frac{P(D | F) P(F, l)}{P(D)}
\]

\[
P(F, l | D) = \frac{1}{Z} e^{-E[F, \lambda]} \frac{1}{P(D)}
\]

\[
E[F, \lambda] = \frac{1}{Z} \sum_{i=1}^{N} (d_i - f_i)^2 + A \sum_{i=1}^{N-1} (f_{i+1} - f_i)^2 (1 - l_i)
\]

Like a discrete form of Mumford-Shah.

Can we get Mumford-Shah by taking the limit as lattice points get close together?

No, not without a lot of hard work.
Weak Membrane

Without line processes, we would smooth over the boundary.

1) Idi - di small, then smooth the image.
2) Idi - di large, then break the smoothing.

Extend to 2D Horizontal & Vertical line processes.

\[ E(F, V, h) = \frac{1}{2} \sum_{ij} (f_{ij} - d_{ij})^2 + \sum_{ij} (f_{ij+1} - f_{ij})^2 (1 - h_{ij}) \]

\[ + \sum_{ij} (f_{ij+1} - f_{ij})^2 (1 - V_{ij}) + \sum_{ij} h_{ij} (1 - V_{ij}) + \sum_{ij} V_{ij} \]

To encourage continued edges e.g. 

\[ \text{or } \]

\[ \text{or } \]
(3c) Weak Membrane

Algorithm to estimate $F, e$ from $[P[F, e|D]^T]$

(1) Simulated Annealing


Start at large $T$

Gradually decrease $T$.

In fact, Ceman & Ceman were the first to prove that simulated annealing works if $T$ is lowered sufficiently slowly.

But impractically slow.

Continuation Methods & Expectation Maximization

This is analogous to methods used for Humpre-Sheh.

$EM \quad p[F|D] = \frac{1}{2} p[F, e|D)$

Treat the $e$'s as hidden variables to be summed out.

Minimize $-\log p[F|D] + \frac{1}{2} \frac{Q(e) \log Q(e)}{P(e|F, D)}$

$= \frac{1}{2} Q(e) \log Q(e) - \frac{1}{2} \frac{Q(e) \log P(e|F, D)}{Q(e)}$

Minimizing alternately w.r.t. $Q(e)$ & $F$ gives the EM algorithm.
(9) **Weak Membrane**

\[ \ln(1-D) \text{ denote } q(u) \text{ by } q_i = q(u_i = 1), \quad 1 - q_i = q(u_i = 0). \]

Then \[ \sum_i \mathbb{E}[\mathbf{Q},\Theta] = \frac{1}{c} \sum_i q_i \log q_i + (1 - q_i) \log (1 - q_i) \]
\[ + \frac{1}{2} \sum_{i=1}^n (d - f_i)^2 + A \sum_i (f_i - f_i^*)^2 (1 - q_i) + B \sum_i q_i \]

**E-STEP:** Fix \( F = \langle f_i \rangle \)

Solve for \( q_i \).

\[ \Rightarrow \quad \hat{q}_i = \frac{1}{1 + e^{-A(f_i + f_i^*) + B}} \]

**M-STEP:** Fix \( \mathbf{Q} = \langle q_i \rangle \) solve for \( f_i \).

\[ \frac{d}{df_i} \left( \sum_i \mathbb{E}[\mathbf{Q},\Theta] \right) \]

Linear equation for \( f_i \):

\[ \Rightarrow \quad \frac{1}{d} (\hat{f}_i - d) + 2A \sum_i (\hat{f}_i - f_i) (\hat{f}_i - f_i^*) (1 - q_i) = 0 \]

Iterate to solution.

To get a continuation method, work with \( \langle P[\mathbf{F}, l_{ij}] \rangle^4 = \frac{1}{Z} e^{-E[\mathbf{F}, \mathbf{Q}] / T} \)

gives deterministic annealing.

Same as EM, but with:

\[ \hat{q}_i = \frac{1}{1 + e^{-A(f_i + f_i^*) + B}} \]

Slowly reduce \( T \). Large \( T \), soft edges; Small \( T \), hard.
Week Mendeleev.

Can extend \( \mathcal{F} \) to 2D.

If coupling between the line processes then exact EM does not work.

Instead do a mean field approximation to approximate \( \mathcal{F}_{\text{MF}} \).

Beyond scope of this lecture.

Alternatively, first sum out \( \ell \).

\[
P[F | \ell] = \frac{1}{Z_c} P[F | \ell, D]
\]

\[
= \frac{1}{Z_c} \prod_{\ell=1}^{\ell_{\max}} e^{-\frac{1}{2 \Omega^2} (\ell - \ell_0)^2} e^{-A_0(H_0 + \ell^2 (1 - \ell_c)) - B_0}
\]

\[
= \frac{1}{Z_c} \prod_{\ell=1}^{\ell_{\max}} e^{-\frac{1}{2 \Omega^2} (\ell - \ell_0)^2} e^{-A_0(H_0 + \ell^2) + e^{-B_0}}
\]

\[
= \frac{1}{Z_c} \prod_{\ell=1}^{\ell_{\max}} e^{-\frac{1}{2 \Omega^2} (\ell - \ell_0)^2} \log (e^{-A_0(H_0 + \ell^2) + e^{-B_0}})
\]

Effective Interaction.

Compare to quadratic potential for no line processes.

For large \( \Delta t \), cut off quadratic for small \( \Delta t = \text{non} - \ell_c \).

\[\text{Effective Interaction} \]

\[\text{Compare to quadratic potential for no line processes.}\]

\[\text{For large } \Delta t, \text{ cut off quadratic for small } \Delta t = \text{non} - \ell_c.\]
Maximum Entropy Learning.

\[ \text{Statistics } \phi(x) \text{, observed statistics } \Psi \]

Learn distribution from observed statistics by using the maximum entropy principle.

\[ \text{maximize: } -\frac{1}{x} \sum p(x) \log p(x) + \lambda \left( \frac{1}{x} \sum p(x) \phi(x) - \Psi \right) \]

\[ \text{with } p(x) \]

\[ \Rightarrow p(x) = \frac{e^{x' \phi(x)}}{Z} \frac{2}{e^{\mu \phi(x)}} \text{ exponential distribution} \]

(E.g. In 1-Dimension, if \( \phi(x) = (x, x^2) \) then \( \max \)-ent gives the Gaussian distribution \( p(x) = \frac{1}{\sqrt{2\pi \sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \)).

Solve for \( \mu \) by the condition \( \sum p(x) \phi(x) = \Psi \).

Equivalent to minimizing the convex function

\[ \Psi[\mu] = \log Z[\mu] - \mu \cdot \Psi \]

\[ \Psi[\mu] = \sum \frac{1}{Z} e^{x' \phi(x)} \mu^t \phi(x) \]

Almost all distributions are exponential or can be approximated by them.

Algorithm like Generalized Iterative Scaling (GIS)

\[ \mu^{t+1} = \mu^t + \log p - \log \Psi^t \]

where \( \Psi^t = \sum \frac{1}{Z} e^{M^t \phi(x)} \phi(x) \)

\( \mu^t \) hard to evaluate, need stochastic sampling or approximation (e.g., belief propagation).
**Maximum Entropy Learning**

For vision, the histogram of filter \( f_i(x) = x_i + c - x_c \)
image \( x = (x_1, \ldots, x_n) \).

\[
H_x(z) = \frac{1}{N} \sum_{i=1}^{N} f_i(x), \quad z \text{ takes } N \text{ discrete values.}
\]

Max Ent. \( \Rightarrow \) \( p(x) = \frac{1}{Z[H]} e^{\frac{1}{2} \sum_{i=1}^{N} \mu_i(z) h_i(x)} = \frac{1}{Z[H]} e^{\frac{1}{2} \sum_{i=1}^{N} \mu_i(f_i(x))} \)

Solve \( \sum_x p(x) H_x(z) = \psi(z) \) to estimate the potential \( \mu \).

Final: (Zhu & Huang)
similar to Clemen & Clemen.

**Multiple Filters/Statistics**

External class of filters \( \Phi_1(x), \Phi_2(x) \)
parameters \( \mu_1, \mu_2, \ldots, \mu_M \).

\( p(x) = \frac{1}{Z[\mu_1, \mu_2]} e^{\mu_1 \Phi_1(x) + \cdots + \mu_M \Phi_M(x)} \)

**Model Comparison:**

1. Do you do better with more filters? \( p(x|M_1) \) versus \( p(x|M_2) \), Bayes Model Selection.

2. Or penalize no. of filter. E.g., AIC (Akaike Information Criterion) or Schwarz's criterion (based on asymptotic analysis).