Image Modeling & Segmentation

Aly Farag and Asem Ali

Lecture #7
Fitting a MRF model to an image requires that the parameters of the model be estimated from a sample of the image.

The literature is rich with works that propose different MGRF models which are suitable for a specific system behavior.

Usually, these works identify their models parameters using an optimization technique. This technique tries to maximize either the likelihood or the entropy of the proposed probability distributions.

**Maximum Likelihood Estimation (MLE)**

For the Gibbs probability distribution (GPD):

\[
P(f) = \frac{1}{Z} \exp \left( \sum_{(p,q) \in N} V(f_p, f_q, \beta) \right)
\]

The log-likelihood function is defined by

\[
L(f | \beta) = \log P(f) = \sum_{(p,q) \in N} V(f_p, f_q, \beta) - \log(Z(\beta))
\]

The maximum log-likelihood estimator is defined by

\[
\beta^* = \arg \max_{\beta} \left( \sum_{(p,q) \in N} V(f_p, f_q, \beta) - \log(Z(\beta)) \right)
\]
MGRF-based Image Analysis

Coding Method (Besage’74):

- Maximizes the log-likelihood in coding $j$

$$L_j(\beta) = \sum_{p \in P_j} \log \left( \frac{\exp(-U(f, \beta))}{\sum_{l \in \mathcal{L}} \exp(-U(l, \beta))} \right)$$

$$\beta = \frac{1}{4} \sum_{j=1}^{4} \beta_j .$$

Colors of pixels belong to the same coding are conditionally independent.

Least Square Error method (LSQR) (Derin and Elliot PAMI’87)

$$\sum_{q \in N_p} (V(l_1, f_q) - V(l_2, f_q)) = \log \frac{P(l_2 | f_{N_p}) + \epsilon}{P(l_1 | f_{N_p}) + \epsilon},$$

- The ratio is estimated by counting the number of blocks of type 1 and dividing by the number of blocks of type 2.
- Solving overdetermined system of linear equations using the most frequently occurring blocks types.

MGRF-based Image Analysis

Anisotropic Potts Models

\[ V = \beta_q \ast \delta(f_p \neq f_q) \]

\[ \beta_q \propto \exp \left( -\frac{(I_p-I_q)^2}{2\sigma^2} \right) \cdot \frac{1}{\|p-q\|_2} \]

\[ V_{pq} = \begin{cases} \frac{-\|I_p-I_q\|_2^2}{\sigma^2} \frac{-\|p-q\|_2^2}{\sigma^2} & \text{if } \|p-q\|_2 < r; \\ 0 & \text{otherwise} \end{cases} \]

- Different Types of these potential functions

\[ V = \beta |f_p - f_q| \quad \text{or} \quad V = \beta \ast \delta(f_p \neq f_q) \quad \text{or} \quad V = \min(\beta, |f_p - f_q|) \]

Analytical Estimation (Farag et al) for Potts Model

Approximate the log likelihood is obtained by truncating the Taylor’s series expansion

\[ \beta = \frac{K^2}{K-1} \left( \frac{K-1}{K} - \bar{\delta}_{\text{neq}}(f) \right) \]

\[ \bar{\delta}_{\text{neq}}(f) = \frac{1}{|T|} \sum_{\{p,q\} \in T} \delta(f_p \neq f_q) \]
References


“Random field models in image analysis” Journal of Applied statistics, by Dube and Jain

“Markov Random Fields and Images”, by Patrick Perez CWI Quarterly

(NOTE: you don’t need to read the whole paper in each case, pick and choose the related sections)
Put All Together
Image Modeling

Image Labeling

Intensity

Spatial

Interaction

Image

Shape
Labeling Problem

In labeling Problem we have a set of sites \( \mathcal{P} \) and a set of labels \( \mathcal{L} \).

\( \mathcal{P} \) : represents image features \{e.g. pixels, edges, image segments, … etc.\}. Features may have some natural structure as pixels are arranged in 2D array.

\( \mathcal{L} \) : represents intensities, disparities, … etc.

Labeling problem is a mapping \( \mathcal{P} \rightarrow \mathcal{L} \). We denote the labeling by \( f \).

\( \mathcal{P} = \{1, 2, \ldots, n\} \) \hspace{1cm} \( \mathcal{L} = \{l_1, l_2, \ldots, l_k\} \) \hspace{1cm} \( f = \{f_1, f_2, \ldots, f_n\} \)

Set of all labeling \( \mathcal{L}^n \) is denoted by \( \mathcal{F} \)

Simple Example:

\[
\begin{array}{cccc}
 a & b & c & d \\
\end{array}
\]

\( \mathcal{P} = \{1, 2, 3, 4\} \)

\( \mathcal{L} = \{50, 100, 150\} \)

\( f = \{100, 50, 50, 150\} \) \hspace{1cm} \( f = \{50, 50, 50, 150\} \) \hspace{1cm} \( f = \{100, 50, 100, 150\} \)

The set of all labeling \( \mathcal{F} = \mathcal{L}^4 \) consists of \( 3^4 = 81 \) labeling sets.
Labeling problem concept gives a common notation for diverse vision problem, such as:

**Image Segmentation**

\[ P = \{1, 2, \ldots, R \times C\} \]

\[ \mathcal{L} = \{0, 255\} \]

**Image Restoration**

\[ P = \{1, 2, \ldots, R \times C\} \]

\[ \mathcal{L} = \{(0, 0, 0), \ldots, (255, 255, 255)\} \]
Labeling problem concept gives a common notation for diverse vision problem, such as:

**Stereo Matching**

\[ P = \{1, 2, \ldots, R \times C\} \]

\[ L = \{d_{\text{min}} : d_{\text{max}}\} \]

Disparity range

**Depth/disparity map**
Labeling problem concept gives a common notation for diverse vision problem, such as:

**Image Matching**

Shekhovtsov, et al CVPR’07

\[ \mathcal{P} = \{1, 2, \ldots, R \times C\} \]

\[ \mathcal{L} = \{(\delta x_{\text{min}}, \delta y_{\text{min}}) : (\delta x_{\text{max}}, \delta y_{\text{max}})\} \] Displacement range

**Digital Tapestry (Rother et al CVPR’05)**

\[ \mathcal{P} = \{1, 2, \ldots, n_{\text{Blocks}}\} \]

\[ \mathcal{L} = \mathcal{I} \times \mathcal{S} \]
The input image \( I \) and the desired segmented image \( \mathbf{f} \) are described by a joint Markov-Gibbs random field (MGRF).

MGRF model is fitted within the Bayesian framework of Maximum-A-Posteriori (MAP) estimation to estimate \( \mathbf{f} \):

\[
\mathbf{f}^* = \arg \max_{\mathbf{f} \in \mathcal{F}} P(I|\mathbf{f})P(\mathbf{f}).
\]

In the pairwise interaction models, Gibbs energy is defined in terms of clique of size 2. The image \( \mathbf{f} \) is represented by a MGRF with joint distribution:

\[
P(f) = Z^{-1} \exp(- \sum_{\{p,q\} \in \mathcal{N}} V(f_p, f_q)) \quad \text{(A)}
\]

The distribution \( P(I|f) \) is a MRF by assuming the noise at each pixel is independent (Dube and Jain’89)

\[
P(I|f) = \prod_{p \in \mathcal{P}} P(I_p|f_p) \quad \text{(B)}
\]
From (A) & (B) the MAP estimator

$$f^* = \arg \max_{f \in \mathcal{F}} \exp\left(\sum_{p \in \mathcal{P}} \log(P(I_p | f_p)) - \sum_{\{p, q\} \in \mathcal{N}} V(f_p, f_q)\right).$$

Equivalent to minimize the energy

$$E(f) = \sum_{\{p, q\} \in \mathcal{N}} V(f_p, f_q) - \sum_{p \in \mathcal{P}} \log(P(I_p | f_p)).$$

First term expresses smoothing constraints on labeling. Labels varies smoothly everywhere except at the object’s boundaries “discontinuity”.

Second term measures how much assigning label $f_p$ to pixel $p$ disagrees with the observation $I_p$. 
Problem Solvers

- Modern energy minimization methods such as:
  - Graph cuts (Zabih PAMI’01)
  - Belief Propagation (BP) (Felzenszwalb CVPR’04)
  - Tree-ReWeighted message passing (TRW) (Wainwright Info Theory’05)
  - Extended Roof duality (Kolmogorov CVPR’07)

- Classical methods such as:
  - Iterated Conditional Modes (ICM) (Besag’74)
  - Simulated Annealing (Geman & Geman’84)
<table>
<thead>
<tr>
<th>Problem Solvers</th>
<th>Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ICM</strong>: (Szeliski, ECCV06)</td>
<td></td>
</tr>
<tr>
<td>• Fast technique</td>
<td>• Local energy optimization technique</td>
</tr>
<tr>
<td>• Very sensitive to the initial labeling</td>
<td></td>
</tr>
<tr>
<td><strong>SA</strong>: (Szeliski, ECCV06)</td>
<td></td>
</tr>
<tr>
<td>• Finds the global solution with certain temperature schedules</td>
<td></td>
</tr>
<tr>
<td>• The schedules that lead to the global are very slow in practice.</td>
<td></td>
</tr>
<tr>
<td><strong>BP</strong>: (Szeliski, ECCV06)</td>
<td></td>
</tr>
<tr>
<td>• It gives exact minimization if the graph of the energy is a tree,</td>
<td></td>
</tr>
<tr>
<td>• It diverges in the case of graphs that have loops</td>
<td></td>
</tr>
<tr>
<td>• It gives solutions with higher energy than graph cuts</td>
<td></td>
</tr>
<tr>
<td><strong>TRW-S</strong>: (Kolmogorov CVPR’07)</td>
<td></td>
</tr>
<tr>
<td>• Similar to the BP algorithm.</td>
<td></td>
</tr>
<tr>
<td>• Guarantees the convergence; the lower bound estimate is not to decrease</td>
<td></td>
</tr>
<tr>
<td>• Same performance of the roof duality, but it is much slower.</td>
<td></td>
</tr>
<tr>
<td><strong>Graph Cuts</strong>: (Szeliski, ECCV06)</td>
<td></td>
</tr>
<tr>
<td>• Outperforms the other competitive methods (accuracy and time efficiency).</td>
<td></td>
</tr>
<tr>
<td>• Applied to submodular functions.</td>
<td></td>
</tr>
<tr>
<td><strong>Roof duality</strong>: (Kolmogorov CVPR’07)</td>
<td></td>
</tr>
<tr>
<td>• A generalization of the standard graph cut algorithm.</td>
<td></td>
</tr>
<tr>
<td>• For submodular functions, same performance (accuracy and time).</td>
<td></td>
</tr>
<tr>
<td>• Non-submodular functions, roof duality produces part of an optimal solution.</td>
<td></td>
</tr>
</tbody>
</table>
Iterative research of MAP estimate stochastic (e.g., simulated annealing) or deterministic (e.g., iterated conditional modes)

- Simulates a process in metallurgy which determines the low energy states of a material by gradually lowering the energy
- Finds MAP estimators for all pixels simultaneously
- Finds the global solution with certain temperature schedules
- Computationally expensive; the schedules that lead to the global are very slow in practice.
Algorithm 7 Simulated annealing [23]

1: Choose an initial temperature $T$.

2: Select labeling $f$ that maximizes $P(I|f)$

3: repeat

4: while $i < N_{iter}$ do

5: Perturb $f$ into $\tilde{f}$. Let $\Delta = U(\tilde{f}) - U(f)$

6: if ($\Delta < 0$) then $f \leftarrow \tilde{f}$ else $f \leftarrow \tilde{f}$ with probability $e^{\frac{\Delta}{T}}$

7: increase $i$.

8: end while

9: Update $T$ using monotonically decreasing function.

10: until frozen.
Iterated Conditional Modes (ICM) (Besag’74)

- Pixels are processed sequentially, and for each pixel the algorithm selects the label that maximize $P(I_P|f_p)P(f_p|\hat{f}_{N_p})$
- Faster than simulated Annealing
- Very sensitive to the initial labeling
- Local energy optimization technique

Algorithm 8 Iterated Conditional Modes (ICM) [3]

1: Choose a MGRF model for $P(f)$.

2: Select labeling $\hat{f}$ that maximizes $P(I|f)$

3: while $i < N_{iter}$ do

4: for all $p \in \mathcal{P}$ do

5: Update $\hat{f}_p$ by the value of $f_p$ that maximizes $P(I_P|f_P)P(f_p|\hat{f}_{N_p})$

6: end for

7: increase $i$.

8: end while
Example

- One row image
  - Observed image
    - Piecewise Constant Prior “Potts’ model”
      \[ V(f_p, f_q) = \begin{cases} 
      50 & \text{if } f_p \neq f_q; \\
      0 & \text{if } f_p = f_q 
      \end{cases} \]
    - Data penalty term
      \[ P(I_p | f_p) \propto \exp(-|I_p - f_p|) \]
Example

\[ E(f) = \sum_{\{p, q\} \in \mathcal{N}} 50 \delta(f_p \neq f_q) + \sum_{p \in \mathcal{P}} |I_p - f_p|. \]

\[
\begin{array}{ccccccccc}
140 & 40 & 10 & 220 & 120 & 170 & 190 & 80 & 30 & 100 \\
\end{array}
\]

Best labeling

\[
E = 50 + 50 + 90 + 10 + 40 + 20 + 80 + 30 + 10 + 30 + 20 + 50 = 480
\]

Threshold labeling

\[
E = 50 + 50 + 50 + 50 + 50 + 60 + 10 + 40 + 20 + 70 + 30 + 10 + 30 + 20 + 50 = 590
\]
Graph Cuts

Graph Cut Basic Definition & Notation

The weighted graph \( \mathcal{G} = (\mathcal{V}, \mathcal{E}) \)

- \( \mathcal{V} \) is the set of vertices in graph correspond to pixels.
- \( \{t, s\} \) (sink & source) are two distinguished vertices called terminals.
- \( \mathcal{E} \) a subset of pairs \((p, q)\) of elements from \( \mathcal{V} \) “Edges”
- A path is a sequence of edges.
- N-link: connects pairs of neighboring vertices.
  Cost/weight: a penalty for discontinuities between vertices
- T-link: connects vertex with terminal.
  Cost/weight : a penalty for assigning the corresponding label to the vertex
Graph Cuts

Min-Cut & Max-Flow

- A cut $\mathcal{C} \subset \mathcal{E}$ is a set of edges such that terminals are separated in the induced graph $\mathcal{G}(\mathcal{C}) = \langle \mathcal{V}, \mathcal{E} - \mathcal{C} \rangle$
- No proper subset of $\mathcal{C}$ separate the terminals in $\mathcal{G}(\mathcal{C})$
- Cost of the cut $\mathcal{C}$, denoted $|\mathcal{C}|$, the sum of its edge weights
- Min-cut is to find the cut with minimum cost among all cuts.
- “Min-Cut can be solved by computing Max-Flow between terminals” Ford & Fulkerson’ 62
Graph Cuts

Min-Cut = Min Capacity = Max-Flow

Cut Cost = 7

Cut Cost = 20

Cut Cost = 30
Min-Cut & Max-Flow Example

The Graph

Max Flow = 4

Min Cut = 10

s-t Min-Cut/Max-Flow algorithm of Boykov & Kolmogorov’04
Graph Cuts

Graph cuts as a minimization technique

\[ E(f) = \sum_{\{p,q\} \in \mathcal{N}} V(f_p, f_q) - \sum_{p \in \mathcal{P}} \log(P(I_p | f_p)). \]

- Every pixel represents a vertex in the graph.
- N-link \((p, q)\) weight \(V(f_p | f_q)\)
- T-link \((s, p), (t, p)\) weight \(- \log P(I_p | f_p)\)
- Compute s-t MinCut
Graph Cuts

Graph cut as a minimization technique (Example)

Max-Flow = $50$
Graph Cuts

Graph cut as a minimization technique (Example)

Max-Flow = 340
1. Compute the empirical density of CT slice

2. Using EM fit N Gaussians to estimate the density

Estimate the marginal density of each class

Fit a MGRF on the image by selecting

1. Neighborhood system
2. Cliques order
3. Potential function
4. Compute potential parameters from initial $f$

Integrate $E(f)$

Minimize

Final output $f$

Initial output $f$
THANK YOU
1st Midterm Exam Wednesday OCT 13
- WHERE ..? In class
- HOW LONG ..? One Hour
- TOPICS ..? Lecture #1 - #7
- MATERIALS ..?
  Closed Book, BUT write what you want in a single sided sheet